

*Manual de usuario de IBM SPSS
Modeler 16*

IBM

Nota

Antes de utilizar esta información y el producto al que da soporte, lea la información incluida en el “Avisos” en la página 181.

Información del producto

Esta edición se aplica a la versión 16, release 0, modificación 0 de IBM(r) SPSS(r) Modeler y a todos los releases y las modificaciones posteriores, hasta que se indique lo contrario en nuevas ediciones.

Contenido

Capítulo 1. Acerca de IBM SPSS Modeler 1

Productos IBM SPSS Modeler	1
IBM SPSS Modeler	1
IBM SPSS Modeler Server	1
IBM SPSS Modeler Administration Console	2
IBM SPSS Modeler Batch	2
IBM SPSS Modeler Solution Publisher	2
IBM SPSS Modeler ServerAdaptadores paraIBM SPSS Collaboration and Deployment Services	2
Ediciones de IBM SPSS Modeler	2
Documentación de IBM SPSS Modeler	3
Documentación de SPSS Modeler Professional	3
Documentación de SPSS Modeler Premium	4
Ejemplos de aplicaciones	5
Carpeta Demos	5

Capítulo 2. Nuevas características en IBM SPSS Modeler 16. 7

Capítulo 3. IBM SPSS Modeler Visión general. 11

Primeros pasos	11
Inicio de IBM SPSS Modeler	11
Ejecución desde la línea de comandos	11
Conexión con IBM SPSS Modeler Server.	12
Modificación del directorio temporal	14
Inicio de varias sesiones de IBM SPSS Modeler	14
Conceptos básicos sobre la interfaz de IBM SPSS Modeler	15
Lienzo de rutas de IBM SPSS Modeler	15
Paleta de nodos	16
Gestores de IBM SPSS Modeler	17
Proyectos de IBM SPSS Modeler	19
Barra de herramientas de IBM SPSS Modeler	19
Personalización de la barra de herramientas	20
Personalización de la ventana de IBM SPSS Modeler	21
Cambio del tamaño de icono de una ruta	22
Utilización del ratón en IBM SPSS Modeler.	22
Uso de teclas de acceso directo	22
Impresión	23
Automatización de IBM SPSS Modeler	24

Capítulo 4. Comprensión de la minería de datos 25

Conceptos básicos de la minería de datos	25
Evaluación de los datos	26
Una estrategia para la minería de datos	27
El modelo del proceso CRISP-DM	28
Tipos de modelos	29
Ejemplos de minería de datos	34

Capítulo 5. Generación de rutas 35

Conceptos básicos de la generación de rutas	35
---	----

Generación de rutas de datos	35
Cómo trabajar con nodos	36
Cómo trabajar con rutas	41
Descripciones de ruta	53
Ejecución de rutas	54
Trabajo con modelos	55
Adición de comentarios y anotaciones a nodos y rutas.	55
Almacenamiento de rutas de datos	61
Carga de archivos	62
Correlacionar rutas de datos.	63
Sugerencias y atajos	65

Capítulo 6. Cómo gestionar valores perdidos 67

Conceptos básicos de valores perdidos	67
Cómo gestionar valores perdidos	68
Gestión de registros con valores perdidos	68
Gestión de campos con valores perdidos	68
Imputación o relleno de valores perdidos	69
Funciones CLEM para valores perdidos	69

Capítulo 7. Generación de expresiones CLEM 71

Acerca de CLEM	71
CLEMEjemplos	71
Valores y tipos de datos	73
Expresiones y condiciones	74
Parámetros de ruta, sesión y Supernodo.	74
Cómo trabajar con cadenas	75
Gestión de elementos vacíos y valores perdidos	76
Cómo trabajar con números	76
Cómo trabajar con fechas y horas	76
Resumen de varios campos	77
Trabajo con datos de respuestas múltiples	78
Generador de expresiones	79
Acceso al generador de expresiones	79
Creación de expresiones	79
Selección de funciones.	79
Selección de campos, parámetros y variables globales	80
Presentación o selección de valores	80
Comprobación de expresiones CLEM.	81
Buscar y reemplazar	81

Capítulo 8. CLEM Referencia del lenguaje 85

Conceptos básicos de la referencia de CLEM	85
Tipos de datos de CLEM	85
Enteros	85
Reales	86
Caracteres	86
Cadenas	86
Listas	86

Fields	87
Fechas	87
Hora	88
Operadores de CLEM	88
Referencia de funciones	90
Convenciones en las descripciones de las funciones	91
Funciones de información	92
Funciones de conversión	92
Funciones de comparación	93
Funciones lógicas	96
Funciones numéricas	97
Funciones trigonométricas	98
Funciones de probabilidad	98
Operaciones de enteros a nivel de bit.	99
Funciones aleatorias	100
Funciones de cadena	100
Funciones SoundEx	105
Funciones de fecha y hora	105
Funciones de secuencia	109
Funciones globales	114
Funciones para gestionar los valores vacíos y nulos	115
Campos especiales.	115

Capítulo 9. Uso de IBM SPSS Modeler con un repositorio 119

Acerca de IBM SPSS Collaboration and Deployment Services Repository	119
Almacenamiento y despliegue de objetos de repositorio	120
Conexión con el repositorio.	121
Introducción de credenciales para el repositorio	121
Exploración del contenido del repositorio	121
Almacenamiento de objetos en el repositorio	122
Configuración de propiedades de objeto	122
Almacenamiento de rutas	124
Almacenamiento de proyectos.	124
Almacenamiento de nodos	125
Almacenamiento de objetos de resultado	125
Almacenamiento de modelos y paletas de modelos	126
Recuperación de objetos desde el repositorio	126
Elección de un objeto que recuperar	127
Selección de una versión de objeto	127
Búsqueda de objetos en el repositorio	127
Modificación de objetos del repositorio.	129
Creación, cambio de nombre y eliminación de carpetas	129
Bloqueo y desbloqueo de objetos y repositorio	129
Eliminación de objetos del repositorio	129
Gestión de las propiedades de objetos del repositorio	130
Visualización de propiedades de carpeta	130
Visualización y edición de propiedades de objetos	131
Gestión de etiquetas de versión de objetos.	132
Despliegue de rutas	132
Opciones de despliegue de rutas	133
La rama de puntuación	135

Capítulo 10. Exportación a aplicaciones externas. 139

Acerca de Exportación a aplicaciones externas	139
Apertura de una ruta en IBM SPSS Modeler Advantage	139
Cómo importar y exportar modelos como PMML	140
Tipos de modelos que admiten PMML	140

Capítulo 11. Proyectos e informes . . . 143

Introducción a los proyectos	143
Vista CRISP-DM	143
Vista Clases	144
Generación de un proyecto	144
Creación de un nuevo proyecto	144
Adición a un proyecto	144
Transferencia de proyectos a IBM SPSS Collaboration and Deployment Services Repository	145
Configuración de las propiedades de un proyecto	146
Anotaciones de un proyecto	146
Propiedades de objeto	147
Cierre de un proyecto	147
Generación de un informe	147
Almacenamiento y exportación de informes generados	149

Capítulo 12. Personalización de IBM SPSS Modeler 151

Personalización de opciones de IBM SPSS Modeler	151
Configuración de las opciones de IBM SPSS Modeler	151
Opciones de sistema	151
Configuración de directorios predeterminados	152
Opciones de configuración de usuario	152
Personalización de la paleta de nodos	156
Personalización del gestor de paletas	156
Modificación de la vista de la pestaña de paleta	158
Gestión de nodos CEMI	159

Capítulo 13. Consideraciones de rendimiento para rutas y nodos . . . 161

Orden de los nodos	161
Almacenamiento en caché de los nodos	162
Rendimiento: nodos de proceso	163
Rendimiento: nodos de modelado	164
Rendimiento: expresiones CLEM	165

Capítulo 14. Accesibilidad en IBM SPSS Modeler 167

Conceptos básicos sobre la accesibilidad de IBM SPSS Modeler	167
Tipos de soporte de accesibilidad.	167
Accesibilidad para personas con problemas de visión	167
Accesibilidad para usuarios invidentes	168
Accesibilidad desde el teclado.	168
Utilización de un lector de pantallas.	175
Sugerencias de utilización	176

Interferencias con otro software	177	Avisos	181
JAWS y Java.	177	Marcas comerciales	182
Utilización de gráficos en IBM SPSS Modeler	177	Índice.	185
Capítulo 15. Compatibilidad con			
Unicode.	179		
Compatibilidad con Unicode en IBM SPSS Modeler	179		

Capítulo 1. Acerca de IBM SPSS Modeler

IBM® SPSS Modeler es un conjunto de herramientas de minería de datos que permite desarrollar rápidamente modelos predictivos mediante técnicas empresariales y desplegarlos en operaciones empresariales para mejorar la toma de decisiones. Con un diseño que sigue el modelo CRISP-DM, estándar del sector, IBM SPSS Modeler admite el proceso completo de minería de datos, desde los propios datos hasta obtener los mejores resultados empresariales.

IBM SPSS Modeler ofrece una gran variedad de métodos de modelado procedentes del aprendizaje automático, la inteligencia artificial y el estadístico. Los métodos disponibles en la paleta de modelado permiten derivar nueva información procedente de los datos y desarrollar modelos predictivos. Cada método tiene ciertos puntos fuertes y es más adecuado para determinados tipos de problemas.

SPSS Modeler puede adquirirse como producto independiente o utilizarse como cliente junto con SPSS Modeler Server. También hay disponible cierto número de opciones adicionales que se resumen en las siguientes secciones. Si desea obtener más información, consulte <http://www.ibm.com/software/analytics/spss/products/modeler/>.

Productos IBM SPSS Modeler

La familia de productos IBM SPSS Modeler y su software asociado se componen de lo siguiente:

- IBM SPSS Modeler
- IBM SPSS Modeler Server
- IBM SPSS Modeler Administration Console
- IBM SPSS Modeler Batch
- IBM SPSS Modeler Solution Publisher
- IBM SPSS Modeler Server adaptadores para IBM SPSS Collaboration and Deployment Services

IBM SPSS Modeler

SPSS Modeler es una versión con todas las funcionalidades del producto que puede instalar y ejecutar en su ordenador personal. Puede ejecutar SPSS Modeler en modo local como un producto independiente o utilizarla en modo distribuido junto con IBM SPSS Modeler Server para mejorar el rendimiento a la hora de trabajar con grandes conjuntos de datos.

Con SPSS Modeler, puede crear modelos predictivos precisos de forma rápida e intuitiva sin necesidad de programación. Mediante su exclusiva interfaz visual, podrá visualizar fácilmente el proceso de minería de datos. Con ayuda del análisis avanzado incrustado en el producto podrá detectar patrones y tendencias en sus datos que anteriormente estaban ocultos. Podrá modelar los resultados y comprender los factores que influyen en ellos, lo que le permitirá aprovechar oportunidades comerciales y mitigar los riesgos.

SPSS Modeler está disponible en dos ediciones: SPSS Modeler Professional y SPSS Modeler Premium. Consulte el tema “Ediciones de IBM SPSS Modeler” en la página 2 para obtener más información.

IBM SPSS Modeler Server

SPSS Modeler utiliza una arquitectura de cliente/servidor para distribuir peticiones de cliente para operaciones que requieren un uso intensivo de los recursos a un software de servidor de gran potencia, lo que proporciona un rendimiento más rápido con conjuntos de datos de mayor volumen.

SPSS Modeler Server es un producto con licencia independiente que se ejecuta de manera continua en modo de análisis distribuido en un host de servidor junto con una o más instalaciones de IBM SPSS

Modeler. De este modo, SPSS Modeler Server ofrece un mejor rendimiento cuando se trabaja con grandes conjuntos de datos, ya que las operaciones que requieren un uso intensivo de memoria se pueden realizar en el servidor sin tener que descargar datos al equipo cliente. IBM SPSS Modeler Server también ofrece asistencia para las capacidades de optimización de SQL y modelado interno de bases de datos, lo que proporciona mayores ventajas en cuanto al rendimiento y la automatización.

IBM SPSS Modeler Administration Console

Modeler Administration Console es una aplicación gráfica para administrar muchas de las opciones de configuración de SPSS Modeler Server, las cuales también pueden configurarse a través de un archivo de opciones. La aplicación proporciona una interfaz de usuario de la consola para supervisar y configurar las instalaciones de SPSS Modeler Server y está disponible de forma completamente gratuita para los clientes actuales de SPSS Modeler Server. La aplicación solamente se puede instalar en los ordenadores con Windows; sin embargo, puede administrar un servidor que esté instalado en cualquier plataforma compatible.

IBM SPSS Modeler Batch

Aunque la minería de datos suele ser un proceso interactivo, también es posible ejecutar SPSS Modeler desde una línea de comandos, sin necesidad de la interfaz gráfica del usuario. Por ejemplo, puede que tenga tareas repetitivas o cuya ejecución sea de larga duración que quiera realizar sin intervención por parte del usuario. SPSS Modeler Batch es una versión especial del producto que ofrece asistencia para todas las capacidades analíticas de SPSS Modeler sin acceder a la interfaz de usuario habitual. Es necesario disponer de una licencia de SPSS Modeler Server para utilizar SPSS Modeler Batch.

IBM SPSS Modeler Solution Publisher

SPSS Modeler Solution Publisher es una herramienta que le permite crear una versión empaquetada de una ruta de SPSS Modeler que se puede ejecutar en un motor de tiempo de ejecución externo o incrustado en una aplicación externa. De este modo, podrá publicar y desplegar rutas completas de SPSS Modeler para utilizarlas en entornos que no tengan SPSS Modeler instalado. SPSS Modeler Solution Publisher se distribuye como parte del servicio IBM SPSS Collaboration and Deployment Services - Puntuación, para el que se necesita una licencia independiente. Con esta licencia, recibirá SPSS Modeler Solution Publisher Runtime, que le permite ejecutar las rutas publicadas.

IBM SPSS Modeler Server Adaptadores para IBM SPSS Collaboration and Deployment Services

Tiene a su disposición un determinado número de adaptadores para IBM SPSS Collaboration and Deployment Services que permiten que SPSS Modeler y SPSS Modeler Server interactúen con un repositorio de IBM SPSS Collaboration and Deployment Services. De este modo, varios usuarios podrán compartir una ruta de SPSS Modeler desplegada en el repositorio, o bien se podrá acceder a ella desde la aplicación cliente de baja intensidad IBM SPSS Modeler Advantage. Debe instalar el adaptador en el sistema donde se aloje el repositorio.

Ediciones de IBM SPSS Modeler

SPSS Modeler está disponible en las siguientes ediciones.

SPSS Modeler Professional

SPSS Modeler Professional proporciona todas las herramientas que necesita para trabajar con la mayoría de los tipos de datos estructurados, como los comportamientos e interacciones registrados en los sistemas de CRM, datos demográficos, comportamientos de compra y datos de ventas.

SPSS Modeler Premium

SPSS Modeler Premium es un producto con licencia independiente que amplía SPSS Modeler Professional para poder trabajar con datos especializados, como los utilizados para el análisis de entidades o las redes sociales, así como con datos de texto no estructurados. SPSS Modeler Premium está formado por los siguientes componentes:

IBM SPSS Modeler Entity Analytics incorpora una dimensión adicional al análisis predictivo de IBM SPSS Modeler predictive analytics. Mientras que el análisis predictivo trata de predecir comportamientos futuros a partir de datos del pasado, el análisis de entidades se centra en mejorar la coherencia de los datos actuales mediante la resolución de conflictos de identidades dentro de los propios registros. La identidad de un individuo, una organización, un objeto o cualquier otra entidad puede estar expuesta a ambigüedades. La resolución de identidades puede ser vital en diversos campos, entre los que se incluyen la gestión de la relación con el cliente, la detección de fraudes, la lucha contra el blanqueo de dinero y la seguridad nacional e internacional.

IBM SPSS Modeler Social Network Analysis transforma la información sobre relaciones en campos que caracterizan el comportamiento social de individuos y grupos. Mediante el uso de datos que describen las relaciones subyacentes de las redes sociales, IBM SPSS Modeler Social Network Analysis identifica a los líderes sociales que influyen en el comportamiento de otros en la red. Además, puede determinar qué personas se ven más afectadas por otros participantes de la red. Al combinar estos resultados con otras medidas, puede crear perfiles completos de individuos en los que basar sus modelos predictivos. Los modelos que incluyan esta información social tendrán un mejor rendimiento que los modelos que no la incluyan.

IBM SPSS Modeler Text Analytics utiliza tecnologías de lingüística avanzada y Procesamiento del lenguaje natural (PLN) para procesar con rapidez una gran variedad de datos de texto sin estructurar, extraer y organizar los conceptos clave y agruparlos en categorías. Las categorías y conceptos extraídos se pueden combinar con los datos estructurados existentes, como pueden ser datos demográficos, y se pueden aplicar para modelar utilizando el conjunto completo de herramientas de minería de datos de IBM SPSS Modeler para tomar decisiones mejores y más certeras.

Documentación de IBM SPSS Modeler

Tiene a su disposición documentación en formato de ayuda en línea desde el menú Ayuda de SPSS Modeler. Se incluye documentación para SPSS Modeler, SPSS Modeler Server y SPSS Modeler Solution Publisher, así como el Manual de aplicaciones y otros materiales de apoyo.

La documentación completa de cada producto (incluidas las instrucciones de instalación) en formato PDF está disponible en la carpeta *\Documentation* en cada DVD del producto. También es posible descargar los documentos de instalación en Internet en <http://www-01.ibm.com/support/docview.wss?uid=swg27038316>.

La documentación en ambos formatos también está disponible desde el centro de información de SPSS Modeler en <http://publib.boulder.ibm.com/infocenter/spssmodl/v16r0m0/>.

Documentación de SPSS Modeler Professional

El conjunto de documentación de SPSS Modeler Professional (excluidas las instrucciones de instalación) es el siguiente.

- **IBM SPSS Modeler Manual del usuario.** Introducción general sobre cómo usar SPSS Modeler, incluyendo cómo crear rutas de datos, tratar valores perdidos, crear expresiones CLEM, trabajar con proyectos e informes y empaquetar rutas para su despliegue en IBM SPSS Collaboration and Deployment Services, Predictive Applications o IBM SPSS Modeler Advantage.
- **Nodos de origen, proceso y resultado de IBM SPSS Modeler.** Descripciones de todos los nodos utilizados para leer, procesar y dar salida a datos en diferentes formatos. En la práctica, esto implica todos los nodos que no sean nodos de modelado.

- **Nodos de Modelado de IBM SPSS Modeler.** Descripciones de todos los nodos utilizados para crear modelos de minería de datos. IBM SPSS Modeler ofrece una gran variedad de métodos de modelado procedentes del aprendizaje automático, la inteligencia artificial y el estadístico.
- **Manual de algoritmos de IBM SPSS Modeler.** Descripciones de los fundamentos matemáticos de los métodos de modelado que se utilizan en IBM SPSS Modeler. Esta guía está disponible únicamente en formato PDF.
- **Manual de aplicaciones de IBM SPSS Modeler.** Los ejemplos de esta guía ofrecen introducciones breves y concisas a métodos y técnicas de modelado específicos. También tiene a su disposición una versión en línea de este manual en el menú Ayuda. Consulte el tema “Ejemplos de aplicaciones” en la página 5 para obtener más información.
- **Scripts y automatización de IBM SPSS Modeler.** Información sobre la automatización del sistema a través de scripts, incluidas las propiedades que se pueden utilizar para manipular nodos y rutas.
- **IBM SPSS Modeler Manual de despliegue.** Información sobre la ejecución de rutas y escenarios de IBM SPSS Modeler como pasos en trabajos de procesamiento en IBM SPSS Collaboration and Deployment Services Deployment Manager.
- **Guía del desarrollador de IBM SPSS Modeler CLEF.** CLEF ofrece la capacidad de integrar programas de otros fabricantes, como rutinas de procesamiento de datos o algoritmos de modelado como nodos en IBM SPSS Modeler.
- **Manual de minería interna de bases de datos de IBM SPSS Modeler.** Este manual incluye información sobre cómo utilizar la potencia de su base de datos, tanto para mejorar su rendimiento como para ampliar su oferta de capacidades analíticas a través de algoritmos de terceros.
- **Guía de administración de IBM SPSS Modeler Server y su rendimiento.** Información sobre la configuración y administración de IBM SPSS Modeler Server.
- **Manual del usuario de IBM SPSS Modeler Administration Console.** Información sobre cómo instalar y utilizar la interfaz de usuario de la consola para supervisar y configurar IBM SPSS Modeler Server. La consola se implementa como complemento de la aplicación Gestor de despliegue.
- **Manual CRISP-DM de IBM SPSS Modeler.** Manual que explica paso a paso cómo utilizar la metodología de CRISP-DM en la minería de datos con SPSS Modeler.
- **IBM SPSS Modeler Batch Manual del usuario.** Guía completa de cómo utilizar IBM SPSS Modeler en modo por lotes, incluida información detallada sobre la ejecución del modo por lotes y argumentos de línea de comandos. Esta guía está disponible únicamente en formato PDF.

Documentación de SPSS Modeler Premium

El conjunto de documentación de SPSS Modeler Premium (excluidas las instrucciones de instalación) es el siguiente.

- **IBM SPSS Modeler Entity Analytics Manual del usuario.** Información sobre cómo utilizar el análisis de entidades con SPSS Modeler, que cubre la instalación y configuración de repositorios, nodos de análisis de entidades y tareas administrativas.
- **IBM SPSS Modeler Social Network Analysis Manual del usuario.** Una guía para realizar análisis de redes sociales con SPSS Modeler, incluido el análisis de grupos y el análisis de difusión.
- **SPSS Modeler Text Analytics Manual del usuario.** Información sobre cómo utilizar el análisis de texto con SPSS Modeler, que cubre los nodos de minería de texto, programa interactivo, plantillas y otros recursos.
- **Manual del usuario de IBM SPSS Modeler Text Analytics Administration Console.** Información sobre cómo instalar y utilizar la interfaz de usuario de la consola para supervisar y configurar IBM SPSS Modeler Server para su uso con SPSS Modeler Text Analytics . La consola se implementa como complemento de la aplicación Gestor de despliegue.

Ejemplos de aplicaciones

Mientras que las herramientas de minería de datos de SPSS Modeler pueden ayudar a resolver una amplia variedad de problemas organizativos y empresariales, los ejemplos de la aplicación ofrecen introducciones breves y adaptadas de técnicas y métodos de modelado específicos. Los conjuntos de datos utilizados aquí son mucho más pequeños que los enormes almacenes de datos gestionados por algunos analizadores de datos, pero los conceptos y métodos implicados deberían ser escalables a las aplicaciones reales.

Para acceder a los ejemplos pulsando **Ejemplos de aplicación** en el menú Ayuda de SPSS Modeler. Los archivos de datos y rutas de ejemplo se instalan en la carpeta *Demos* en el directorio de instalación del producto. Consulte el tema “Carpeta Demos” para obtener más información.

Ejemplos de modelado de bases de datos. Consulte los ejemplos que figuran en el *Manual de minería interna de bases de datos de IBM SPSS Modeler*.

Ejemplos de scripts. Consulte los ejemplos que figuran en la *Guía de scripts y automatización de IBM SPSS Modeler*.

Carpeta Demos

Los archivos de datos y rutas de ejemplo utilizados con los ejemplos de la aplicación se instalan en la carpeta *Demos* en el directorio de instalación del producto. También puede acceder a esta carpeta desde el grupo de programas IBM SPSS Modeler en el menú Inicio de Windows o pulsando *Demos* de la lista de directorios recientes en el cuadro de diálogo Abrir archivo.

Capítulo 2. Nuevas características en IBM SPSS Modeler 16

IBM SPSS Modeler añade las características siguientes en este release.

Simulación Monte Carlo. Un nuevo nodo de origen de simulación proporciona una forma fácil de generar datos sintéticos de cero utilizando una amplia selección de distribuciones estadísticas. De forma alternativa, el nuevo nodo de ajuste puede crear automáticamente un nodo de origen configurado previamente que refleja las distribuciones de y las relaciones entre variables históricas. El nodo de evaluación de simulación es un nodo terminal diseñado para evaluar campos que se originan de una ruta de análisis simulado y proporciona una distribución práctica y gráficos de correlación.

Scripts Python. Los scripts en IBM SPSS Modeler, utilizados para automatizar procesos en la interfaz de usuario, ahora pueden utilizar el lenguaje Python, así como seguir soportando la modalidad de scripts de legado. Python es un lenguaje popular y bien conocido que proporciona un conjunto avanzado de características que incluye una sintaxis avanzada y concisa, manejo de errores y potentes módulos adicionales. **Nota:** La pestaña **Script** en **Herramientas > Propiedades de ruta** ahora se llama **Ejecución**.

Ejecución de bucles y condicional. Estas nuevas opciones permiten la ejecución de bucles y condicional de rutas sin tener que codificar los scripts. Encuentre estas nuevas opciones en **Herramientas > Propiedades de ruta > Ejecución** o pulsando con el botón derecho del ratón en un nodo de una ruta y eligiendo la opción de ejecución **Bucles/Condicional**.

Nodo STB (cuadros de espacio tiempo). Cree rectángulos de ubicación y datos de indicación de fecha y hora para dar soporte a análisis más sofisticados. En el modo de desconexión, este nodo también identifica los tiempos y lugares donde residen las entidades. Las funciones del creador de expresiones adicionales soportan la extracción de los centroides STB, así como los códigos geográficos.

Mejoras de análisis de entidades. Así como resolver entidades individuales, esto ahora puede identificar relaciones de n-grado entre entidades. El soporte adicional se proporciona para la actualización en tiempo real a través del nodo de ruta, aplanando entidades resueltas (nodo Distinto) y para hacer que los datos sean anónimos cuando llenan un repositorio de entidad. **Nota:** DB2 ha sustituido la base de datos SolidDB local anterior.

Nuevo tipo de gráfico de nodo de evaluación de característica operativa del receptor (ROC) y medidas de área debajo de la curva (AUC) y Gini en el nodo de análisis. Soporta los objetivos binarios.

Nueva opción de nodo distinto para crear un registro compuesto. Le permite especificar un método de agregación para cada campo que se agrupe (primer valor, último valor, valores de concatenación, etc.) en lugar de descartar grupos duplicados. Mientras que el nodo de agregación normalmente se utiliza para resumir los datos en un nivel superior, esta nueva opción se utiliza para aplanar duplicados; por ejemplo, los duplicados identificados a través de la resolución de entidad.

Nodos de exportación y origen TM1. Le permite acceder a vistas de cubo de TM1 a través del nodo de origen TM1 y volver a puntuar los datos en un cubo TM1 existente utilizando el nodo de exportación TM1.

Expresión de agregación y funciones de agregación de ventana. Puede crear expresiones de agregación personalizadas en el nodo de agregación, incorporando funciones de agregación integradas (MEAN, SUM, y etc.) y/o funciones definidas por el usuario de agregación de base de datos. En las expresiones derivadas, puede derivar campos que requieren funciones de agregación con ventanas (como, por ejemplo, medias móviles). Están disponibles funciones de agregación de ventana proporcionadas con las base de datos e incorporadas.

Mejoras de minería interna de la base de datos de IBM Netezza Analytics. Nuevo algoritmo de dos pasos de Netezza, opción adicional de aplicación de ayuda para gestionar (suprimir, renombrar, etc.) modelos de análisis de Netezza, soporte para visores de Modelos para árboles de regresión, árboles de decisiones, K-medias y TwoStep. **Nota:** Las nuevas características requieren INZA 3.0.

Nuevos nodos R y generador de cuadros de diálogo personalizados para R. Además del nodo de generador de modelos R y el nugget de modelo introducidos en SPSS Modeler 15 fixpack 2, este release añade 2 nodos nuevos: proceso R y salida R. Con el nodo de proceso R, puede tomar datos de una ruta de SPSS Modeler y aplicar transformaciones en los datos utilizando scripts R. Con el nodo de salida R, puede utilizar sus propios scripts R personalizados para realizar análisis de datos y para resumir los resultados de las puntuaciones de modelos.

Puede generar resultados gráficos y de texto de los análisis. Esta salida se puede direccionar a un archivo, o visualizar en el explorador de la salida del nodo de salida R. El generador de cuadros de diálogo personalizados proporciona la capacidad para crear tipos de nodo de creación, proceso y salida personalizados y nuggets de modelo, incluyendo un seleccionador de campo, texto, números, botones de selección y controles de subdiálogo para permitir la abstracción y parametrización de programas R. Puede elegir el tipo de nodo, la paleta de destino y el icono de nodo antes de instalar el nodo o compartirlo para que lo utilicen otros usuarios de SPSS Modeler. El generador de cuadros de diálogo personalizados se inicia desde el menú Herramientas. **Nota:** Para utilizar esta característica, debe haber instalado SPSS Modeler - Essentials para R.

R en base de datos. Soporte de retrotracción SQL para nodos R; para Netezza, SAP Hana y Oracle utilizando su soporte para R. **Nota:** Las bases de datos deben tener instaladas las ampliaciones R correctas proporcionadas por el proveedor.

Nodo de proceso de serie temporal de rutas. Cree y puntúe los modelos de serie temporal en un único paso para proporcionar despliegue en tiempo real a través de IBM InfoSphere Streams, el servicio de puntuación de IBM SPSS Collaboration and Deployment Services o IBM SPSS Modeler Solution Publisher.

Botón Presentación preliminar. Cuando se utiliza junto con un origen de base de datos, la vista previa permite que se visualice la retrotracción SQL. Cuando se utiliza junto con un origen de datos de Analytic Server, utilice el botón Vista previa para recibir información sobre posibles grandes movimientos de datos.

Nuevas opciones de Analytic Server en los nodos de clasificador automático, numérico automático y clúster automático. Al ejecutar una ruta en IBM SPSS Analytic Server, puede elegir entre realizar la ejecución con las divisiones habilitadas (si está utilizando la característica de modelo de división, utilice esta opción) o las opciones de datos muy grandes (se ignoran las divisiones, y los objetivos de creación de modelos se definen para datos grandes). Ahora la puntuación está soportada para los modelos de clasificador automático, numérico automático y clúster automático que se pueden generar en SPSS Modeler Server.

Soporte mejorado del adaptador de puntuación. Soporte para puntuar modelos de minería de texto a través de adaptadores de puntuación de base de datos. Asimismo, nuevos adaptadores de puntuación de base de datos para DB2 LUW y Oracle.

Nodo de origen de vista de datos. El nodo de vista de datos le permite leer datos de una vista de datos de análisis en una ruta de SPSS Modeler. La vista de datos de análisis es una nueva forma para crear una vista de datos unificada en IBM SPSS Collaboration and Deployment Services 6.

Miscelánea de otras mejoras

- Bandas de consultas Teradata en valores preestablecidos de conexión de base de datos.
- Opciones de nodo GLMM actualizadas.
- Soporte para archivos .sav cifrados y comprimidos.

- Soporte el inicio de sesión único (SSO) de SPSS Modeler Server sin necesitar IBM SPSS Collaboration and Deployment Services.
- Soporte de capa de base de datos para el inicio de sesión único.
- Retrotracción SQL para el nodo de muestra para zDB2.
- Los nodos R ahora se instalan como parte de la instalación base de SPSS Modeler.
- Los adaptadores para modelador de IBM SPSS Collaboration and Deployment Services se instalan mediante IBM Installation Manager.
- Actualización de modelo de división.
- Soporte la la localización en ruso.

Capítulo 3. IBM SPSS Modeler Visión general

Primeros pasos

Como aplicación de minería de datos, IBM SPSS Modeler ofrece un método estratégico para encontrar relaciones útiles entre grandes conjuntos de datos. Al contrario que los métodos estadísticos más tradicionales, no es necesario saber lo que se está buscando al comenzar. Puede explorar los datos, mediante el ajuste de diferentes modelos y la investigación de diferentes relaciones, hasta que encuentre la información que resulte útil.

Inicio de IBM SPSS Modeler

Para iniciar la aplicación, pulse en:

Inicio > [Todos los] Programas > IBM SPSS Modeler 16 > IBM SPSS Modeler 16

La ventana principal se mostrará transcurridos unos segundos.

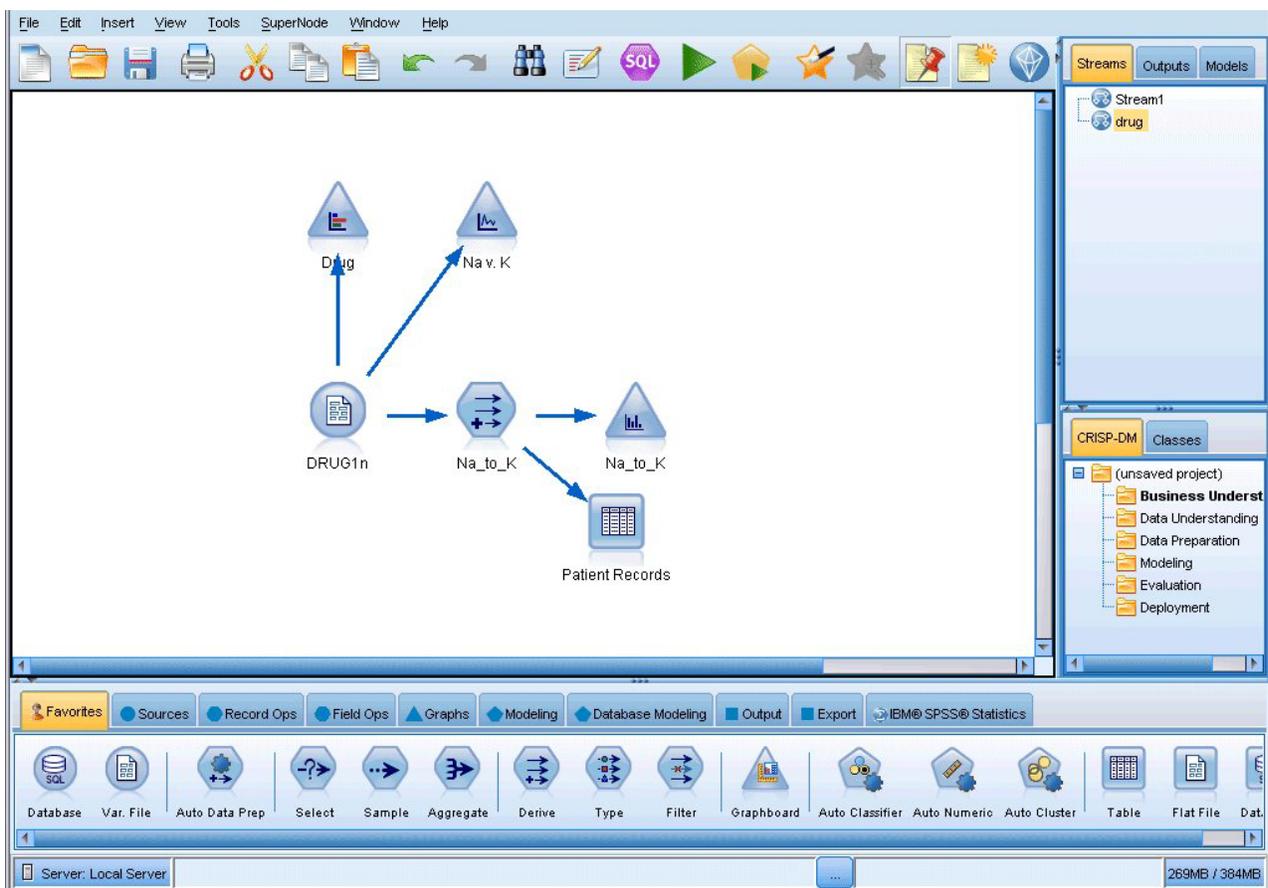


Figura 1. Ventana principal de la aplicación IBM SPSS Modeler

Ejecución desde la línea de comandos

Puede utilizar la línea de comandos del sistema operativo para iniciar IBM SPSS Modeler de la siguiente manera:

1. En un ordenador en el que se haya instalado IBM SPSS Modeler, abra una ventana de DOS o del indicador de comandos.
2. Para iniciar la interfaz de IBM SPSS Modeler en modo interactivo, escriba el comando `clementine` seguido de los argumentos necesarios; por ejemplo:

```
modelerclient -stream report.str -execute
```

Los argumentos disponibles (modificadores) permiten conectar con un servidor, cargar rutas, ejecutar scripts o especificar otros parámetros, según sea necesario.

Conexión con IBM SPSS Modeler Server

IBM SPSS Modeler puede ejecutarse como una aplicación independiente o como un cliente conectado a IBM SPSS Modeler Server directamente o a IBM SPSS Modeler Server o un clúster de servidores a través del complemento Coordinator of Processes de IBM SPSS Collaboration and Deployment Services. El estado de la conexión actual se muestra en la parte inferior izquierda de la ventana de IBM SPSS Modeler.

Siempre que desee conectarse a un servidor, puede introducir manualmente el nombre de servidor al que desee conectarse o seleccione un nombre que haya definido anteriormente. Sin embargo, si tiene IBM SPSS Collaboration and Deployment Services, puede buscar en una lista de servidores o clústeres de servidores del cuadro de diálogo Inicio de sesión del servidor. La capacidad de buscar entre los servicios de Estadísticas que se ejecutan en una red está disponible a través de Coordinator of Processes.

Para conectar con un servidor

1. En el menú Herramientas, pulse en **Inicio de sesión del servidor**. Se abre el cuadro de diálogo Inicio de sesión del servidor. Si lo prefiere, pulse dos veces con el ratón en el área de estado de la conexión de la ventana de IBM SPSS Modeler.
2. En el cuadro de diálogo, especifique las opciones para conectarse al equipo servidor local o seleccione una conexión de la tabla.
 - Pulse en **Añadir** o **Edición** para añadir o editar una conexión. Consulte el tema “Adición y edición de la conexión de IBM SPSS Modeler Server” en la página 13 para obtener más información.
 - Pulse en **Buscar** para acceder a un servidor o clúster de servidores en Coordinator of Processes. Consulte el tema “Búsqueda de servidores en IBM SPSS Collaboration and Deployment Services” en la página 13 para obtener más información.

Tabla Servidor. Esta tabla contiene el conjunto de conexiones de servidor definidas. La tabla muestra la conexión predeterminada, el nombre de servidor, la descripción y el número de puerto. Puede añadir manualmente una nueva conexión, así como seleccionar o buscar una conexión existente. Para establecer un servidor específico como la conexión predeterminada, seleccione la casilla de verificación en la columna Valor predeterminado de la tabla para la conexión.

Ruta predeterminada de acceso a los datos. Especifique la ruta utilizada para los datos del equipo servidor. Pulse en el botón de puntos suspensivos (...) para examinar la ubicación deseada.

Establecer credenciales. Deje esta casilla sin seleccionar para activar la característica de **inicio de sesión único**, que tratará de iniciar la sesión del usuario en el servidor con los detalles de nombre de usuario y contraseña del equipo local. Si no es posible el inicio de sesión único o si selecciona esta casilla para desactivar el inicio de sesión único (por ejemplo, para iniciar la sesión en una cuenta de administrador), tendrá activados los siguientes campos para que introduzca las credenciales.

ID de usuario. Introduzca el nombre de usuario con el que se inicia sesión en el servidor.

Contraseña. Introduzca la contraseña asociada al nombre de usuario especificado.

Dominio. Especifique el dominio utilizado para iniciar sesión en el servidor. El nombre de dominio es obligatorio cuando el equipo servidor está en un dominio de Windows distinto que el equipo cliente.

3. Pulse en **Aceptar** para completar la conexión.

Desconexión de un servidor

1. En el menú Herramientas, pulse en **Inicio de sesión del servidor**. Se abre el cuadro de diálogo Inicio de sesión del servidor. Si lo prefiere, pulse dos veces con el ratón en el área de estado de la conexión de la ventana de IBM SPSS Modeler.
2. En el cuadro de diálogo, seleccione el Servidor local y pulse en **Aceptar**.

Adición y edición de la conexión de IBM SPSS Modeler Server

Puede editar o añadir manualmente una conexión de servidor en el cuadro de diálogo Inicio de sesión del servidor. Si pulsa en **Añadir**, puede acceder al cuadro de diálogo Añadir/editar servidor vacío en el que puede introducir los detalles de conexión de servidor. Al seleccionar una conexión existente y pulsar en **Editar** en el cuadro de diálogo Inicio de sesión del servidor, se abre el cuadro de diálogo Añadir/editar servidor con los detalles de dicha conexión de modo que puede realizar cualquier cambio.

Note: No puede editar una conexión de servidor que se haya añadido desde IBM SPSS Collaboration and Deployment Services, ya que el nombre, puerto y otros detalles se definen en IBM SPSS Collaboration and Deployment Services.

Adición de conexiones de servidor

1. En el menú Herramientas, pulse en **Inicio de sesión del servidor**. Se abre el cuadro de diálogo Inicio de sesión del servidor.
2. En este cuadro de diálogo, pulse en **Añadir**. Se abre el cuadro de diálogo Inicio de sesión del servidor: Añadir/editar servidor.
3. Introduzca los detalles de conexión de servidor y pulse en **Aceptar** para guardar la conexión y volver al cuadro de diálogo Inicio de sesión del servidor.
 - **Servidor.** Especifique un servidor disponible o seleccione uno de la lista. El equipo servidor se puede identificar por un nombre alfanumérico (por ejemplo, *miservidor*) o por una dirección IP asignada al equipo servidor (por ejemplo, 202.123.456.78).
 - **Puerto.** Especifique el número de puerto en el que el servidor escucha. Si no funciona el número de puerto predeterminado, solicite el número de puerto correcto al administrador del sistema.
 - **Descripción.** Introduzca una descripción opcional para esta conexión de servidor.
 - **Asegurar conexión segura (utilizar SSL).** Especifica si se debe usar una conexión SSL (del inglés **Secure Sockets Layer**, capa de sockets seguros). SSL es un protocolo normalmente utilizado para asegurar el conjunto de datos que se envía a través de una red. Para utilizar esta característica, SSL debe estar activado en el servidor que aloja IBM SPSS Modeler Server. Si es preciso, póngase en contacto con el administrador local para obtener más detalles.

Edición de conexiones de servidor

1. En el menú Herramientas, pulse en **Inicio de sesión del servidor**. Se abre el cuadro de diálogo Inicio de sesión del servidor.
2. En este cuadro de diálogo, seleccione la conexión que desee editar y, a continuación, pulse en **Editar**. Se abre el cuadro de diálogo Inicio de sesión del servidor: Añadir/editar servidor.
3. Cambie los detalles de conexión de servidor y pulse en **Aceptar** para guardar los cambios y volver al cuadro de diálogo Inicio de sesión del servidor.

Búsqueda de servidores en IBM SPSS Collaboration and Deployment Services

En lugar de introducir una conexión de servidor manualmente, puede seleccionar un servidor o clúster de servidores disponible en la red a través de Coordinator of Processes, disponible en IBM SPSS Collaboration and Deployment Services. Un clúster de servidores es un grupo de servidores entre los que Coordinator of Processes determina el servidor más adecuado para responder a una solicitud de procesamiento.

Aunque puede añadir servidores manualmente al cuadro de diálogo Inicio de sesión del servidor, la búsqueda de servidores disponibles le permite conectarse a servidores sin que sea necesario que conozca

el nombre de servidor y número de puerto correctos. Esta información se proporciona automáticamente. Sin embargo, todavía necesita la información de inicio de sesión correcta, como el nombre de usuario, dominio y contraseña.

Note: Si no tiene acceso a la capacidad Coordinator of Processes, todavía puede introducir manualmente el nombre de servidor al que desee conectarse o seleccionar un nombre que haya definido anteriormente. Consulte el tema "Adición y edición de la conexión de IBM SPSS Modeler Server" en la página 13 para obtener más información.

Búsqueda de servidores y clústeres

1. En el menú Herramientas, pulse en **Inicio de sesión del servidor**. Se abre el cuadro de diálogo Inicio de sesión del servidor.
2. En este cuadro de diálogo, pulse en **Buscar** para abrir el cuadro de diálogo Buscar servidores. Si no ha iniciado sesión en IBM SPSS Collaboration and Deployment Services cuando intente buscar en Coordinator of Processes, se le pedirá que lo haga.
3. Seleccione el servidor o el clúster de servidores de la lista.
4. Pulse en **Aceptar** para cerrar el cuadro de diálogo y añadir esta conexión a la tabla en el cuadro de diálogo Inicio de sesión del servidor.

Modificación del directorio temporal

IBM SPSS Modeler Server realiza algunas operaciones que requieren la creación de archivos temporales. De forma predeterminada, IBM SPSS Modeler utiliza el directorio temporal del sistema para crear archivos temporales. Se puede modificar la ubicación del directorio temporal con los pasos siguientes.

1. Cree un nuevo directorio denominado *spss* y un subdirectorio denominado *servertemp*.
2. Edite *options.cfg*, que se encuentra en el directorio */config* del directorio de instalación de IBM SPSS Modeler. Edite el parámetro *temp_directory* de este archivo, para que indique: *temp_directory*, "C:/spss/servertemp".
3. A continuación, es necesario reiniciar el servicio IBM SPSS Modeler Server. Esta operación se puede realizar pulsando en la pestaña **Servicios** del Panel de control de Windows. Es necesario detener el servicio e iniciarlo de nuevo para activar los cambios realizados. Cuando se reinicie el equipo también se reiniciará el servicio.

Todos los archivos temporales se escribirán a partir de este momento en este directorio.

Nota: el error más habitual cuando se intenta realizar esta acción es el uso de un tipo de barras incorrecto; se utilizan las barras inclinadas.

Inicio de varias sesiones de IBM SPSS Modeler

Si necesita iniciar más de una sesión de IBM SPSS Modeler a la vez, deberá realizar algunos cambios en la configuración de IBM SPSS Modeler y Windows. Por ejemplo, puede que necesite hacerlo si tiene dos licencias de servidor independientes y desee ejecutar dos rutas frente a dos servidores diferentes del mismo equipo cliente.

Para activar varias sesiones de IBM SPSS Modeler:

1. Pulse en:
Inicio > [Todos los] Programas > IBM SPSS Modeler 16
2. En el acceso directo de IBM SPSS Modeler 16 (el que tiene un icono), pulse con el botón derecho del ratón y seleccione **Propiedades**.
3. En el cuadro de texto **Objetivo**, añada *-noshare* al final de la cadena.
4. En Windows Explorer, seleccione:
Herramientas > Opciones de carpeta...

5. En la pestaña Tipos de archivo, seleccione la opción Ruta de IBM SPSS Modeler y pulse en **Opciones avanzadas**.
6. En el cuadro de diálogo Editar tipo de archivo, seleccione Abrir con IBM SPSS Modeler y pulse en **Editar**.
7. En el cuadro de texto **Aplicación utilizada para realizar la acción**, añada -noshare delante del argumento **-stream**.

Conceptos básicos sobre la interfaz de IBM SPSS Modeler

En cada punto del proceso de minería de datos, la interfaz de IBM SPSS Modeler fácil de usar implica el uso de técnicas empresariales. Los algoritmos de modelado, tales como predicción, clasificación, segmentación y detección de asociaciones, garantizan la obtención de modelos exactos y potentes. Los resultados del modelo se pueden desplegar y leer fácilmente en bases de datos, IBM SPSS Statistics y en una amplia variedad de aplicaciones.

El trabajo con IBM SPSS Modeler es un proceso de tres pasos para trabajar con datos.

- En primer lugar, lee los datos en IBM SPSS Modeler.
- A continuación, ejecuta los datos mediante una serie de manipulaciones.
- Por último, envía los datos a un destino.

Esta secuencia de operaciones se denomina **ruta de datos** porque los datos fluyen registro por registro desde el origen pasando por cada manipulación y, finalmente, llega al destino, que puede ser un modelo o un tipo de datos de resultados.

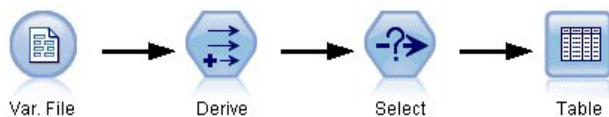


Figura 2. Una ruta simple

Lienzo de rutas de IBM SPSS Modeler

El lienzo de rutas es el área más grande de la ventana de IBM SPSS Modeler y en éste se generan y manipulan rutas de datos.

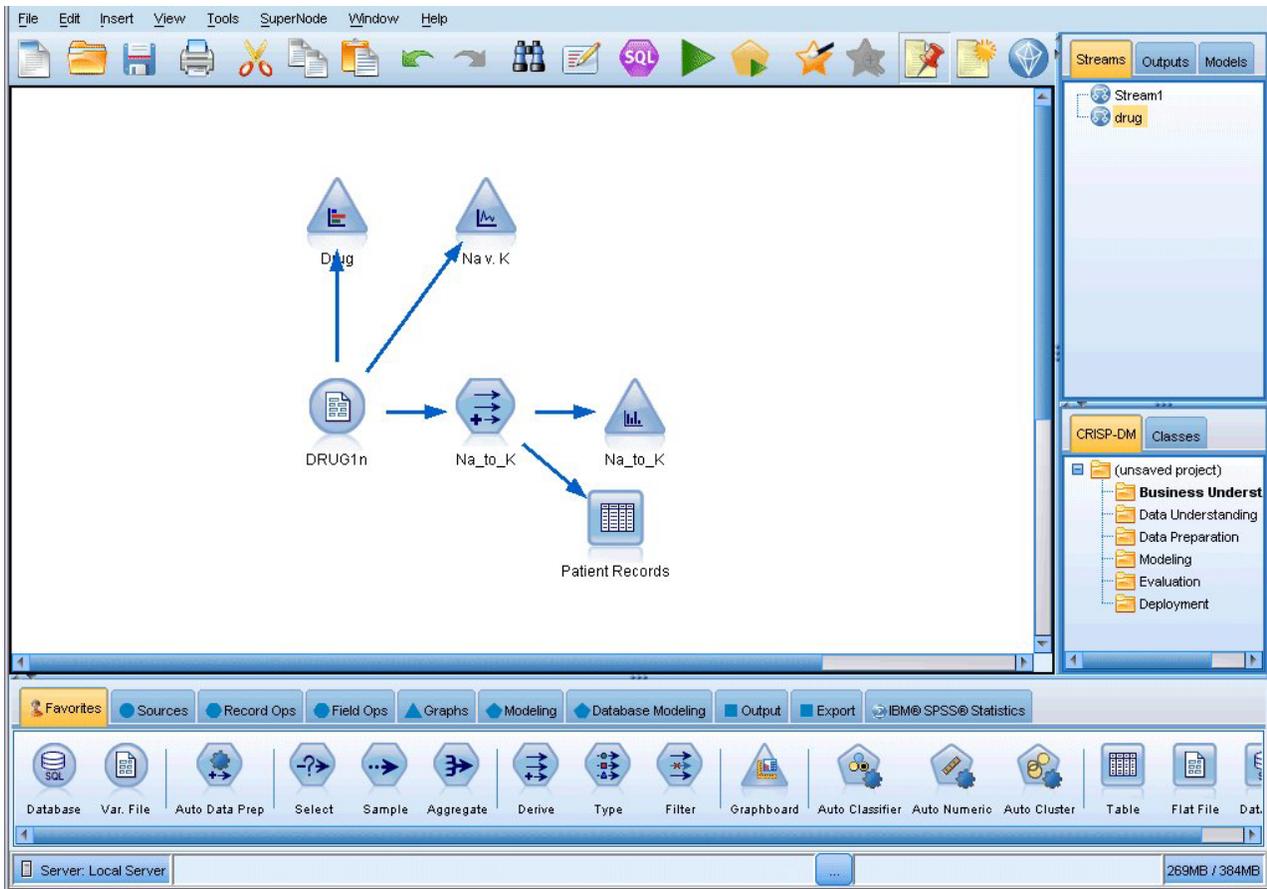


Figura 3. Espacio de trabajo de IBM SPSS Modeler (vista predeterminada)

Las rutas se crean dibujando diagramas de operaciones de datos relevantes para su negocio en el lienzo principal de la interfaz. Cada operación se representa con un icono o un **nodo** y los nodos están vinculados entre sí en una **ruta** que representa el flujo de datos en cada operación.

Se puede trabajar con varias rutas al mismo tiempo en IBM SPSS Modeler, en el mismo lienzo de rutas o abriendo uno nuevo. Durante una sesión, las rutas se almacenan en el gestor de rutas, en la parte superior derecha de la ventana de IBM SPSS Modeler.

Paleta de nodos

La mayoría de los datos y las herramientas de modelado de IBM SPSS Modeler se encuentran en la **Paleta de nodos**, situadas por la parte inferior de la ventana bajo el lienzo de rutas.

Por ejemplo, la pestaña Paleta Oper. con registros contiene nodos que puede utilizar para realizar operaciones en los **registros** de datos, como la selección, la fusión y la adición.

Para añadir nodos al lienzo, pulse dos veces en los iconos de la Paleta de nodos o arrástrelos y suéltelos en el lienzo. A continuación, conéctelos para crear una **ruta**, que represente el flujo de datos.

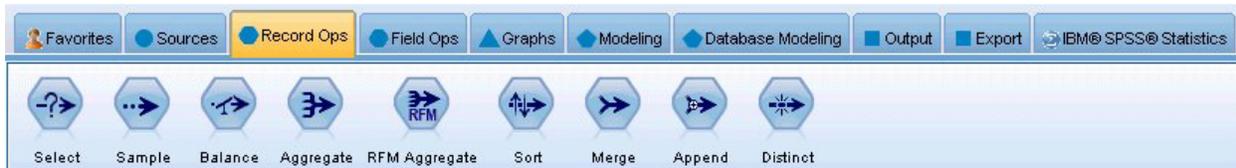


Figura 4. Pestaña Operaciones con registros de la paleta de nodos

Cada pestaña de paleta contiene una colección de nodos relacionados entre sí que se utilizan en distintas fases de las operaciones de rutas, tales como:

- **Orígenes.** Los nodos introducen datos en IBM SPSS Modeler.
- **Operaciones con registros** Los nodos realizan operaciones en los **registros** de datos como la selección, la fusión y la adición.
- **Operaciones con campos** Los nodos realizan operaciones en los **campos** de datos como el filtrado, la derivación de campos nuevos y la determinación del nivel de medición de campos dados.
- **Gráficos.** Los nodos muestran gráficamente los datos antes y después del modelado. Entre ellos se incluyen gráficos, histogramas, nodos de malla y diagramas de evaluación.
- **Modelado.** Los nodos utilizan los algoritmos de modelado disponibles en IBM SPSS Modeler, tales como las redes neuronales, los árboles de decisión, los algoritmos de agrupación en clústeres y las secuencias de datos.
- **Modelado de bases de datos.** Los nodos utilizan los algoritmos de modelado disponibles en las bases de datos Microsoft SQL Server, IBM DB2, Netezza y Oracle.
- **Resultados.** Los nodos generan una diversidad de resultados para los datos, gráficos y resultados de modelos que pueden visualizarse en IBM SPSS Modeler.
- **Exportar.** Los nodos generan una diversidad de resultados que pueden visualizarse en aplicaciones externas, como IBM SPSS Data Collection o Excel.
- **IBM SPSS Statistics.** Los nodos importan datos y exportan datos a IBM SPSS Statistics, ejecutando también procedimientos de IBM SPSS Statistics.

Una vez que se familiarice más con IBM SPSS Modeler, podrá personalizar el contenido de la paleta para su propio uso.

Debajo de la Paleta de nodos, hay un panel de informe que proporciona información sobre el progreso de distintas operaciones, como la lectura de datos en la ruta de datos. Situado también debajo de la Paleta de nodos, hay un panel de estado que proporciona información acerca de la operación que está realizando la aplicación e indica cuándo son necesarios los comentarios del usuario.

Gestores de IBM SPSS Modeler

En la parte superior derecha de la ventana se encuentra el panel de gestores. Este panel cuenta con tres pestañas que se utilizan para administrar rutas, resultados y modelos.

Se puede utilizar la pestaña Rutas para abrir, cambiar nombres, guardar o eliminar las rutas creadas en una sesión.



Figura 5. Pestaña Rutas



Figura 6. Pestaña Resultados

La pestaña Resultados contiene una serie de archivos, como gráficos y tablas, generados mediante operaciones de rutas en IBM SPSS Modeler. Puede mostrar, guardar, cambiar el nombre y cerrar las tablas, gráficos e informes que se enumeran en esta pestaña.



Figura 7. Pestaña Modelos que contiene nuggets de modelo

La pestaña Modelos es la pestaña de gestor más potente. Esta pestaña contiene todos los **nugget** de modelo, que son modelos generados en IBM SPSS Modeler, para la sesión actual. Estos modelos se pueden examinar directamente en la pestaña Modelos o añadirlos a la ruta en el lienzo.

Proyectos de IBM SPSS Modeler

En la parte inferior derecha de la ventana se encuentra el panel de proyectos, que se utiliza para crear y administrar los **proyectos** de minería de datos (grupo de archivos relacionados con una tarea de minería de datos). Existen dos formas de ver los proyectos que se crean en IBM SPSS Modeler: en la vista Clases y la vista CRISP-DM.

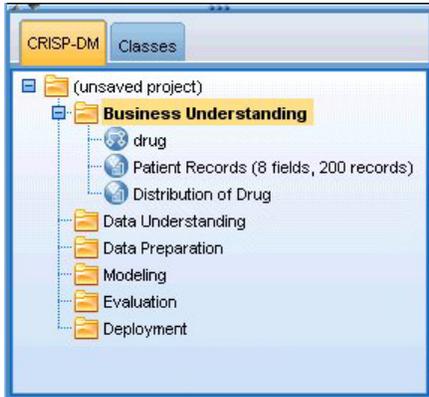


Figura 8. vista CRISP-DM

La pestaña CRISP-DM permite organizar los proyectos según el proceso CRISP-DM (Cross-Industry Standard Process for Data Mining), una metodología independiente y probada en el sector. Los analizadores de datos con o sin experiencia pueden utilizar la herramienta CRISP-DM para mejorar la organización y la comunicación de los esfuerzos.

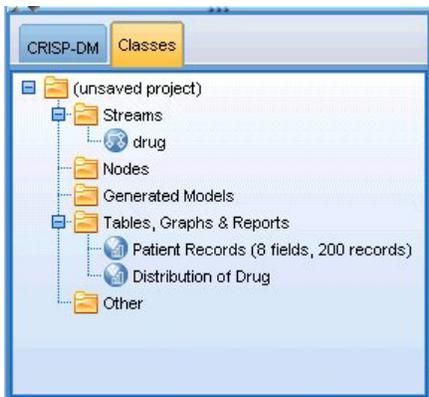


Figura 9. Vista Clases

La pestaña Clases permite organizar el trabajo en IBM SPSS Modeler de forma categórica, por los tipos de los objetos que se hayan creado. Esta vista resulta útil al realizar un inventario de datos, rutas y modelos.

Barra de herramientas de IBM SPSS Modeler

En la parte superior de la ventana de IBM SPSS Modeler hay una barra de herramientas con iconos que proporciona una serie de funciones muy útiles. A continuación se detallan los botones de la barra de herramientas y sus funciones.



Crear una nueva ruta



Abrir una ruta existente

	Guardar la ruta actual		Imprimir la ruta actual
	Cortar & mover la selección al Portapapeles		Copiar al Portapapeles
	Pegar el contenido del Portapapeles en la selección		Deshacer la última acción
	Rehacer		Buscar nodos
	Editar las propiedades de la ruta		Presentación preliminar de generación de SQL
	Ejecutar ruta actual		Ejecutar selección de ruta
	Detener ruta (sólo se activa durante la ejecución de la ruta)		Añadir Supernodo
	Acercar Supernodo (sólo con Supernodos)		Alejar Supernodo (sólo con Supernodos)
	Sin marcación en la ruta		Insertar comentario
	Ocultar marcación de ruta (si la hay)		Mostrar marcación de ruta oculta
	Abrir una ruta existente en IBM SPSS Modeler Advantage		

La marcación de ruta consta de comentarios, enlaces de modelos e indicaciones de las ramas de puntuación.

Los enlaces de modelos se describen en el manual *Nodos de modelado de IBM SPSS*.

Personalización de la barra de herramientas

Puede cambiar varios aspectos de la barra de herramientas, como:

- Si se visualiza
- Si los iconos tienen información sobre herramientas
- Si utiliza iconos grandes o pequeños

Para activar o desactivar la barra de herramientas:

1. En el menú principal, pulse en:
Ver > Barra de herramientas > Visualización

Para cambiar la información sobre herramientas o la configuración del tamaño de iconos:

1. En el menú principal, pulse en:
Ver > Barra de herramientas > Personalizar

Pulse **Mostrar información sobre herramientas** o **Botones grandes**, según sea necesario.

Personalización de la ventana de IBM SPSS Modeler

Se puede cambiar el tamaño de las herramientas o cerrarlas con los separadores de las distintas partes de la interfaz de IBM SPSS Modeler. Por ejemplo, si trabaja con una ruta larga, puede utilizar las flechas pequeñas situadas en cada separador para cerrar la paleta de nodos, el panel de gestores y el de proyectos. De esta forma se maximiza el lienzo de rutas y se proporciona espacio de trabajo suficiente para varias rutas o para rutas grandes.

También puede pulsar desde el menú Ver en **Paleta de nodos**, **Gestores** o **Proyecto** para activar o desactivar la visualización de estos elementos.

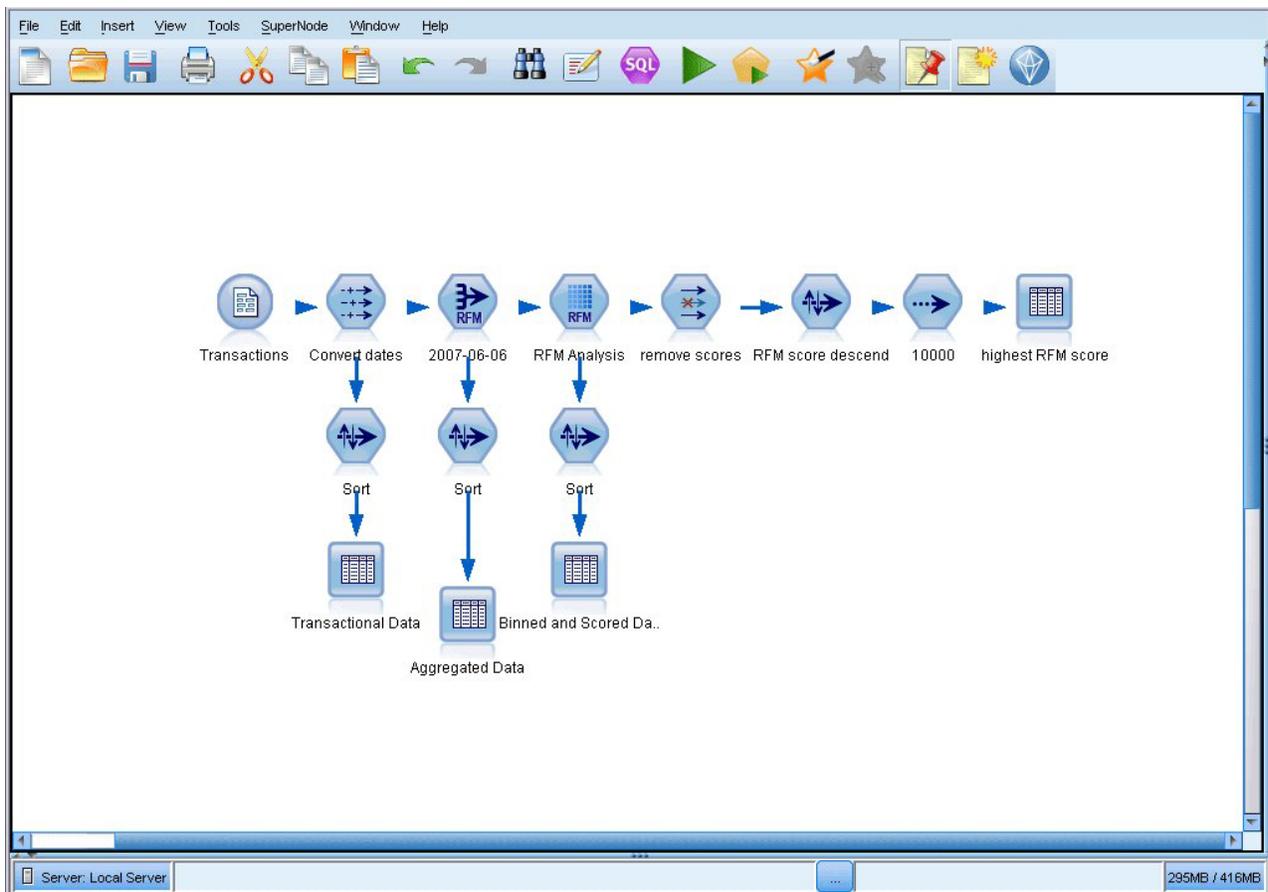


Figura 10. Lienzo de rutas maximizado

En lugar de cerrar la paleta de nodos o los paneles de gestores y de proyectos, también se puede utilizar el lienzo de rutas como una página desplazable moviéndolo vertical y horizontalmente con las barras de desplazamiento situadas en el lateral y en la parte inferior de la ventana de IBM SPSS Modeler.

También puede controlar la visualización de la marcación de pantalla, que consta de los comentarios de rutas, los enlaces de modelos y las indicaciones de las ramas de puntuación. Para activar o desactivar esta visualización, pulse:

Ver > Marcación de rutas

Cambio del tamaño de icono de una ruta

Puede cambiar el tamaño de los iconos de ruta de las maneras siguientes.

- Mediante un ajuste de propiedades de ruta
- Mediante un menú emergente en la ruta
- Mediante el teclado

Puede adaptar la totalidad de la vista de ruta a uno de los tamaños disponibles entre el 8% y el 200% del tamaño de icono estándar.

Para adaptar toda la ruta (método de propiedades de ruta)

1. En el menú principal, elija:
Herramientas > Propiedades de ruta > Opciones > Diseño.
2. Seleccione el tamaño que quiera en el menú Tamaño de icono.
3. Pulse en **Aplicar** para ver el resultado.
4. Pulse en **Aceptar** para guardar el cambio.

Para adaptar toda la ruta (método de menú)

1. Pulse dos veces en el fondo de la ruta en el lienzo.
2. Elija **Tamaño de icono** y seleccione el tamaño que quiera.

Para adaptar toda la ruta (método de teclado)

1. Pulse Ctrl + [-] en el teclado principal para alejarse hasta el siguiente tamaño más pequeño.
2. Pulse Ctrl + Mayús + [+] en el teclado principal para acercarse hasta el siguiente tamaño más grande.

Esta característica es especialmente útil para obtener una vista general de una ruta compleja. También puede utilizarla para reducir el número de páginas necesarias para imprimir una ruta.

Utilización del ratón en IBM SPSS Modeler

Los usos más comunes del ratón en IBM SPSS Modeler incluyen los siguientes:

- **Pulsar una vez.** Utilice el botón derecho o el izquierdo del ratón para seleccionar las opciones de los menús, abrir menús emergentes y acceder a otros controles y opciones estándar. Pulsar y mantener pulsado el botón para mover y arrastrar nodos.
- **Pulsar dos veces.** Pulse dos veces con el botón izquierdo del ratón para colocar nodos en el lienzo de rutas y editar nodos existentes.
- **Pulsar con el botón central.** Pulse con el botón central del ratón y arrastre el cursor para conectar nodos en el lienzo de rutas. Pulse dos veces con el botón central del ratón para desconectar un nodo. Si el ratón no tiene un botón central, se puede simular esta característica pulsando la tecla Alt a la vez que pulsa con el ratón y se arrastra.

Uso de teclas de acceso directo

Muchas operaciones de programación visual de IBM SPSS Modeler poseen teclas de acceso rápido asociadas. Por ejemplo, se puede eliminar un nodo pulsando en el nodo y en la tecla Supr del teclado. Del mismo modo, se puede guardar una ruta de forma rápida manteniendo pulsada la tecla Ctrl y pulsando la tecla S. Comandos de control como éste se indican con una combinación de Ctrl con otra tecla; por ejemplo, Ctrl+S.

En las operaciones estándar de Windows se utilizan varias teclas de acceso directo, tales como Ctrl+X para cortar. Estos atajos son compatibles con IBM SPSS Modeler junto con los siguientes atajos de aplicaciones específicas.

Note: En algunos casos, las teclas de acceso directo antiguas de IBM SPSS Modeler entran en conflicto con las de Windows. Estos atajos antiguos son compatibles si además se pulsa la tecla Alt. Por ejemplo, se puede utilizar Ctrl+Alt+C para activar y desactivar la caché.

Tabla 1. Teclas de acceso directo compatibles

Tecla de acceso directo	Función
Ctrl+A	Seleccionar todo
Ctrl+X	Cortar
Ctrl+N	Nueva ruta
Ctrl+O	Abrir una ruta existente
Ctrl+P	Imprimir
Ctrl+C	Copiar
Ctrl+V	Pegar
Ctrl + Z	Deshacer
Ctrl+Q	Selecciona todos los nodos que se encuentren por debajo del nodo seleccionado
Ctrl+W	Anule la selección de todos los nodos posteriores en la ruta (se conmuta con Ctrl+Q)
Ctrl+E	Ejecutar desde el nodo seleccionado
Ctrl+S	Guarda la ruta actual
Alt+Teclas de flecha	Mueve los nodos seleccionados en el lienzo de rutas en la dirección de la flecha utilizada.
Mayús+F10	Abre el menú emergente del nodo seleccionado

Tabla 2. Atajos compatibles para teclas de acceso rápido anteriores

Tecla de acceso directo	Función
Ctrl+Alt+D	Duplica el nodo
Ctrl+Alt+L	Carga el nodo
Ctrl+Alt+R	Cambia el nombre del nodo
Ctrl+Alt+U	Crea un nodo Datos Usuario
Ctrl+Alt+C	Conmutar caché activada/desactivada
Ctrl+Alt+F	Vacía la caché
Ctrl+Alt+X	Expandir Supernodo
Ctrl+Alt+Z	Acercar/alejar
Suprimir	Elimina el nodo o la conexión

Impresión

Se pueden imprimir los siguientes objetos en IBM SPSS Modeler:

- Diagramas de ruta
- Gráficos
- Tablas
- Informes (del nodo Informe y de los informes de proyectos)

- Scripts (desde los cuadros de diálogo de propiedades de la ruta, Script autónomo o Script de Supernodo)
- Modelos (exploradores de modelos, pestañas de cuadros de diálogo con la vista actual, visores de árboles)
- Anotaciones (mediante la pestaña Anotaciones de resultados)

Para imprimir un objeto:

- Para imprimir sin presentación preliminar, pulse en el botón Imprimir de la barra de herramientas.
- Para configurar la página antes de imprimir, seleccione **Configurar página** en el menú Archivo.
- Para mostrar la representación preliminar, seleccione **Presentación preliminar** en el menú Archivo.
- Para que se muestre el cuadro de diálogo de impresión estándar con las opciones para seleccionar las impresoras y especificar las opciones de aspecto, seleccione **Imprimir** en el menú Archivo.

Automatización de IBM SPSS Modeler

Debido a que la minería de datos avanzada puede ser un proceso complejo y a menudo largo, IBM SPSS Modeler incluye varios tipos de soporte de codificación y automatización.

- **Control Language for Expression Manipulation (CLEM)** es un lenguaje para analizar y manipular los datos que fluyen en las rutas de IBM SPSS Modeler. Los analistas de datos suelen utilizar CLEM en las operaciones de rutas para realizar tareas tan simples como derivar beneficios de datos de costes e ingresos, o tan complejas como transformar datos del registro Web en un conjunto de campos y registros con información útil.
- **El procesamiento en** es una herramienta potente para automatizar procesos en la interfaz de usuario. Los scripts pueden realizar las mismas acciones que los usuarios llevan a cabo con un ratón o un teclado. También pueden especificar los resultados y manipular los modelos generados.

Capítulo 4. Comprensión de la minería de datos

Conceptos básicos de la minería de datos

A través de variadas técnicas, la **minería de datos** identifica los nugget de información en los cuerpos de datos. La minería de datos extrae información de manera que pueda ser utilizada en áreas como la toma de decisiones, las predicciones, las previsiones y las estimaciones. Los datos suelen ocupar mucho espacio, aunque tengan un valor bajo y con poca utilidad directa en su forma sin procesar. Es la información oculta la que dispone del valor.

En la minería de datos, los mejores resultados se obtienen de la combinación de sus conocimientos sobre los datos (o los del experto) con las avanzadas técnicas activas de análisis, donde el equipo identifica las relaciones subyacentes y las características de los datos. El proceso de minería de datos genera modelos de datos históricos que se utilizan más tarde en las predicciones, la detección de los patrones y otras funciones. La técnica de construcción de estos modelos se llama **aprendizaje de las máquinas o modelado**.

Técnicas de modelado

IBM SPSS Modeler contiene varias tecnologías de aprendizaje de las máquinas y de modelado, que pueden más o menos agruparse según los tipos de problemas que pretenden resolver.

- Los métodos de modelado predictivo contienen árboles de decisión, redes neuronales y modelos estadísticos.
- Los modelos de agrupación en clústeres se centran en la identificación de grupos de registros similares y en el etiquetado de registros según el grupo al que pertenecen. Los métodos de clúster incluyen Kohonen, K-Medias y Bietápico.
- Las reglas de asociación asocian una conclusión concreta (como, por ejemplo, la compra de un producto en especial) con un conjunto de condiciones (la compra de varios productos).
- Los modelos de cribado se pueden utilizar para cribar datos para ubicar campos y registros con más probabilidad de ser de interés para el modelado e identificar valores atípicos que pueden no ajustarse a los patrones conocidos. Los métodos disponibles incluyen la selección de características y la detección de anomalías.

Manipulación y descubrimiento de datos

IBM SPSS Modeler también contiene diversos recursos que le permiten aplicar sus conocimientos a los datos:

- **Manipulación de datos.** Construye nuevos elementos de datos derivados de los ya existentes y desglosa los datos en subconjuntos significativos. Es posible fusionar y filtrar los datos procedentes de una serie de orígenes.
- **Exploración y visualización.** Muestra aspectos de los datos mediante el nodo Auditoría de datos a fin de desarrollar una auditoría inicial incluidos los gráficos y los estadísticos. La visualización avanzada contiene gráficos interactivos, que pueden exportarse para ser incluidos en informes de proyectos.
- **Estadísticos.** Confirma las relaciones sospechosas entre las variables de los datos. Los estadísticos de IBM SPSS Statistics también pueden utilizarse en IBM SPSS Modeler.
- **Comprobación de la hipótesis.** Construye modelos que muestran la forma en que se comportan los datos, y verifica estos modelos.

Normalmente, utilizará estos recursos para identificar un conjunto halagüeño de atributos en los datos. A continuación, estos atributos pueden cargarse en las técnicas de modelado, que intentarán identificar las reglas y las relaciones subyacentes.

Aplicaciones típicas

Éstas son algunas de las aplicaciones típicas de técnicas de minería de datos:

Correo directo. Determina qué grupos demográficos tienen la tasa de respuesta más alta. Utilice esta información para maximizar la respuesta de correos futuros.

Puntuación del crédito. Utilice un historial de crédito individual para realizar las decisiones de crédito.

Recursos humanos. Comprender los procedimientos de contratación anteriores y crear reglas de decisión a fin de hacer más eficiente el proceso de contratación.

Investigación médica. Cree reglas de decisión que sugieran procedimientos adecuados basados en comprobaciones médicas.

Análisis de mercado. Determine qué variables (como, por ejemplo, geografía, precio y características de los clientes) están asociadas con las ventas.

Control de calidad. Analice los datos procedentes de la manufactura del producto e identifique las variables que determinan los defectos de éste.

Estudio de la política. Utilice los datos de la encuesta para formular la política mediante la aplicación de reglas de decisión a fin de seleccionar las variables más importantes.

Atención médica. Puede combinar las encuestas al usuario con los datos clínicos a fin de descubrir las variables que contribuyen a la salud.

Terminología

Los términos **atributo**, **campo** y **variable** se refieren a un elemento de datos único común en todos los casos que se tienen en cuenta. Se denomina **registro**, **ejemplo** o **caso** a una colección de valores de atributo referida a un caso específico.

Evaluación de los datos

No es probable que la minería de datos sea provechosa a menos que los datos que desea utilizar reúnan ciertos criterios. Las siguientes secciones presentan algunos de los aspectos de los datos y su aplicación que debe tener en cuenta.

Asegúrese de que los datos están disponibles

Ésto puede parecer obvio, pero debe ser consciente de que, a pesar de que los datos puedan estar disponibles, es posible que no se encuentren en una forma en la que sea fácil operar. IBM SPSS Modeler puede importar los datos de bases de datos (mediante ODBC) o de archivos. Sin embargo, los datos pueden estar guardados con otra forma en una máquina a la que no se pueda acceder directamente. Es necesario descargarlo o depositarlo en una forma apropiada antes de que se pueda utilizar. Es posible que se hayan dispersado entre diferentes bases de datos y orígenes, y que necesiten agruparse. Es posible que ni siquiera se encuentren en línea. Si sólo existe en papel, deberá introducir los datos antes de comenzar con la minería de datos.

Compruebe si los datos cubren los atributos relevantes

El objeto de la minería de datos es identificar los atributos relevantes, por lo que puede parecer extraño incluir esta comprobación en primer lugar. Sin embargo, es muy útil consultar qué datos están disponibles e intentar identificar los factores relevantes de probabilidad que no están registrados. A la hora de predecir, por ejemplo, las ventas de helados, es posible que disponga de mucha información

acerca del perfil de ventas, pero puede que no disponga de información acerca de la temperatura o el clima, la cual es probable que juegue un rol importante. Los atributos perdidos no implican necesariamente que la minería de datos no generará resultados útiles, aunque pueden limitar la precisión de las predicciones resultantes.

Una forma rápida de evaluar la situación es desarrollar una auditoría detallada de los datos. Antes de continuar, contemple la opción de conectar un nodo Auditoría de datos al origen de los datos y ejecutarlo para que genere un informe completo.

Preste atención a los datos con ruido

Los datos normalmente contienen errores o pueden contener juicios subjetivos y, por lo tanto, variables. El conjunto de estos fenómenos se conoce por el nombre de **ruido**. En ocasiones, el ruido en los datos es normal. Es posible que también existan reglas subyacentes, pero no serán válidas para el 100% de los casos.

Por lo general, cuanto más ruido haya en los datos, más difícil es obtener resultados exactos. Sin embargo, los métodos de aprendizaje de las máquinas de IBM SPSS Modeler pueden gestionar los datos con ruido y se han utilizado adecuadamente en conjunto de datos que contenían hasta un 50% de ruido.

Asegúrese de que hay datos suficientes

En la minería de datos, el tamaño de un conjunto de éstos no es necesariamente lo más importante. La *susceptibilidad de ser representado* de un conjunto de datos es mucho más significativa, junto con la cobertura de posibles resultados y las combinaciones de las variables.

Generalmente, cuantos más atributos se tengan en cuenta, más registros se necesitarán para lograr una cobertura representativa.

Si los datos son representativos y existen reglas subyacentes generales, es probable que una muestra de datos de unos pocos miles (o incluso cientos) de registros produzcan resultados igual de buenos que si contuviera un millón de registros y, además, se conseguirán resultados más rápidamente.

Busque a los expertos en datos

En muchos casos, trabajará con sus propios datos, por lo que, tanto el contenido como el significado de éstos le serán muy familiares. Sin embargo, si trabaja con datos de otro departamento de la organización, o para un cliente, se recomienda que disponga de acceso a expertos que conozcan los datos. Éstos pueden ayudarle a identificar los atributos relevantes, interpretar los resultados de la minería de datos, distinguir los nugget de información verdaderos de los falsos, y a reconocer los artefactos causados por anomalías en los conjuntos de datos.

Una estrategia para la minería de datos

Al igual que ocurre con la mayoría de los trabajos comerciales, la minería de datos es mucho más eficaz si se realiza de manera planificada y sistemática. Incluso con las herramientas de minería de datos de última generación, como IBM SPSS Modeler, la mayoría del trabajo de la minería de datos necesita un analista empresarial conocedor del sistema para que el proceso se realice correctamente. Estas preguntas le servirán de pauta para la planificación:

- ¿Cuál es el problema fundamental que desea resolver?
- ¿Qué orígenes de datos están disponibles y qué partes de los datos son relevantes para el problema actual?
- ¿Qué tipo de procesamiento previo y limpieza de datos son necesarios antes de comenzar con la minería de datos?
- ¿Qué técnica/s de minería de datos utilizará?

- ¿Cómo piensa evaluar los resultados de la minería de datos?
- ¿Cómo obtendrá el máximo beneficio de la información obtenida de la minería de datos?

El proceso típico de minería de datos puede complicarse muy rápidamente. Existe una gran cantidad de elementos a los que realizar el seguimiento: complejos problemas comerciales, orígenes de datos múltiples, la variación de la calidad de los datos, una matriz de técnicas de minería de datos, las formas diferentes de medir la corrección de la minería de datos, etcétera.

Para realizar un seguimiento adecuado, es aconsejable disponer de un modelo de proceso definido de manera explícita para la minería de datos. El modelo del proceso le ayuda a responder las preguntas citadas previamente en esta sección, y garantiza el tratamiento de los puntos importantes. Sirve como mapa de carreteras de minería de datos para que no se sienta perdido mientras se sumerge en las complejidades de los datos.

El proceso de minería de datos que se recomienda utilizar con SPSS Modeler es el CRISP-DM (Cross-Industry Standard Process for Data Mining). Como se deduce de su propio nombre, este modelo ha sido diseñado como un modelo general que puede aplicarse a una amplia variedad de problemas industriales y comerciales.

El modelo del proceso CRISP-DM

El modelo del proceso general CRISP-DM contiene seis frases clave dirigidas a cubrir los aspectos principales de la minería de datos. Las seis frases se encuadran en un proceso cíclico diseñado para incorporar la minería de datos a los procedimientos comerciales más grandes.

Las seis frases son:

- **Comprensión del negocio.** Probablemente la frase más importante del proceso de minería de datos. La comprensión del negocio contiene la determinación de objetivos comerciales, la evaluación de la situación, la determinación de los objetivos de la minería de datos y la producción de un plan del proyecto.
- **Comprensión de los datos.** Los datos proporcionan el "material sin procesar" de la minería de datos. Esta fase está dirigida a cubrir la necesidad de comprender cuáles son los orígenes de los datos y las características de dichos orígenes. Incluye la recopilación de los datos iniciales, la descripción, exploración y verificación de la calidad de datos. El nodo Auditoría de datos, situado en la paleta de los nodos de resultado es una herramienta indispensable para la comprensión de los datos.
- **Preparación de datos.** Después de catalogar los orígenes de los datos, será necesario que los prepare para su análisis. La preparación incluye la selección, limpieza, construcción, integración y asignación de formato de los datos.
- **Modelado.** Se trata, obviamente, de la parte más llamativa de la minería de datos, en la que se utilizan sofisticados métodos de análisis para extraer la información de los datos. Esta fase implica la selección de las técnicas de modelado, la generación de diseños de comprobación y la generación de modelos de evaluación.
- **Evaluación.** Una vez elegidos los modelos, ya está preparado para evaluar la forma en que los resultados del análisis pueden ayudarle a lograr los objetivos comerciales. Los elementos principales de esta fase son la evaluación de los resultados, la revisión del proceso de minería de datos y la determinación de los siguientes pasos.
- **Despliegue.** Una vez realizado todo este trabajo, es hora de recoger los frutos. Esta fase se centra en la integración de sus nuevos conocimientos en el proceso comercial diario a fin de resolver el problema original comercial. Esta fase incluye el despliegue, el control y el mantenimiento del plan, la producción de un informe final, así como la revisión del proyecto.

Existen algunos puntos clave en este modelo del proceso. En primer lugar, si bien es cierto que existe una tendencia general para que el proceso siga los pasos destacados en orden de los párrafos anteriores, existe un número de casos en los que las fases influyen entre sí de manera no lineal. Por ejemplo, la

preparación de datos suele preceder al modelado. Sin embargo, tanto las decisiones realizadas como la información recogida durante la fase de modelado generalmente pueden hacer que el usuario desee configurar de nuevo ciertas partes de la fase de preparación de datos, los cuales podrán, acto seguido, presentar nuevos problemas de modelado. Ambas fases se retroalimentan hasta que ambas se resuelvan de manera adecuada. De igual manera, la fase de evaluación puede hacer que el usuario desee evaluar de nuevo la comprensión comercial original y puede hacerle caer en la cuenta de que ha estado intentando responder a la pregunta equivocada. En este punto, puede revisar, ya con un mejor objetivo en mente, la comprensión del negocio e iniciar de nuevo el resto del proceso.

El segundo punto clave es la naturaleza iterativa de la minería de datos. Es muy extraño, si es que sucede alguna vez, que el usuario simplemente planifique un proyecto de minería de datos, lo finalice y, acto seguido, empaquete los datos y se vaya a casa. La utilización de la minería de datos de manera que abarque las necesidades del cliente es una tarea continuada. El conocimiento que se obtiene de un ciclo de minería de datos originará siempre nuevas preguntas, nuevos problemas y nuevas oportunidades de identificar y cumplir las necesidades del cliente. Estas nuevas preguntas, problemas y oportunidades suelen poder tratarse analizando de nuevo los datos. Este proceso de análisis e identificación de nuevas oportunidades debería convertirse en parte del proceso de análisis de la empresa, y en piedra angular de la estrategia comercial general.

Esta introducción sólo detalla brevemente los conceptos básicos del modelo de proceso CRISP-DM. Para obtener información detallada acerca del modelo, consulte los siguientes recursos:

- La *Guía de CRISP-DM*, a la que se puede acceder junto con otra documentación en la carpeta `\Documentation` del disco de instalación del producto.
- El sistema de ayuda de CRISP-DM, disponible desde el menú Inicio o pulsando **Ayuda de CRISP-DM** desde el menú Ayuda de IBM SPSS Modeler.

Tipos de modelos

IBM SPSS Modeler ofrece una gran variedad de métodos de modelado procedentes del aprendizaje automático, la inteligencia artificial y el estadístico. Los métodos disponibles en la paleta de modelado permiten derivar nueva información procedente de los datos y desarrollar modelos predictivos. Cada método tiene ciertos puntos fuertes y es más adecuado para determinados tipos de problemas.

El *Manual de aplicaciones de IBM SPSS Modeler* ofrece ejemplos para muchos de estos métodos, junto con una introducción general al proceso de modelado. Este manual está disponible como tutorial en línea y también en formato PDF. Consulte el tema “Ejemplos de aplicaciones” en la página 5 para obtener más información.

Los métodos de modelado se dividen en tres categorías:

- Clasificación
- Asociación
- Segmentación.

Modelos de clasificación

Los *modelos de clasificación* usan el valor de uno o más campos de **entrada** para predecir el valor de uno o más resultados o campos de **destino**. Algunos ejemplos de estas técnicas son: árboles de decisiones (árbol C&R, QUEST, CHAID y algoritmos C5.0), regresión (lineal, logística, lineal generalizada y algoritmos de regresión de Cox), redes neuronales, máquinas de vectores de soporte y redes bayesianas.

Los modelos de clasificación ayudan a las organizaciones a predecir un resultado conocido, como saber si un cliente comprará o se irá, o si una transacción se ajusta a un patrón conocido de fraude. Las técnicas de modelado incluyen aprendizaje automático de las máquinas, inducción de reglas, identificación de subgrupos, métodos estadísticos y generación de varios modelos.

Nodos de clasificación



El nodo Clasificador automático crea y compara varios modelos diferentes para obtener resultados binarios (sí o no, abandono o no de clientes, etc.), lo que le permite seleccionar el mejor enfoque para un análisis determinado. Son compatibles varios algoritmos de modelado, por lo que es posible seleccionar los métodos que desee utilizar, las opciones específicas para cada uno y los criterios para comparar los resultados. El nodo genera un conjunto de modelos basado en las opciones especificadas y clasifica los mejores candidatos en función de los criterios que especifique.



El nodo Autonumérico calcula y compara modelos para resultados de rango numérico continuo utilizando cierto número de métodos diferentes. El nodo funciona de la misma manera que el nodo Clasificador automático, lo que le permite seleccionar los algoritmos que desee utilizar y experimentar con varias combinaciones de opciones en una única pasada de modelado. Los algoritmos admitidos incluyen redes neuronales, C&RT, CHAID, regresión lineal, regresión lineal generalizada y máquinas de vectores de soporte (SVM). Los modelos se pueden comparar basándose en la correlación, el error relativo o el número de variables utilizado.



El nodo de árbol de clasificación y regresión (C&R) genera un árbol de decisión que permite predecir o clasificar observaciones futuras. El método utiliza la partición reiterada para dividir los registros de entrenamiento en segmentos minimizando las impurezas en cada paso, donde un nodo se considera “puro” si el 100% de los casos del nodo corresponden a una categoría específica del campo objetivo. Los campos de entrada y objetivo pueden ser continuos (rango numérico) o categóricos (nominal, ordinal o marca). Todas las divisiones son binarias (sólo se crean dos subgrupos).



El nodo QUEST proporciona un método de clasificación binario para generar árboles de decisión; está diseñado para reducir el tiempo de procesamiento necesario para realizar los análisis de C&RT y reducir la tendencia de los métodos de clasificación de árboles para favorecer a las entradas que permitan realizar más divisiones. Los campos de entrada pueden ser continuos (rango numérico), sin embargo el campo objetivo debe ser categórico. Todas las divisiones son binarias.



El nodo CHAID genera árboles de decisión utilizando estadísticos de chi-cuadrado para identificar las divisiones óptimas. A diferencia de los nodos C&RT y QUEST, CHAID puede generar árboles no binarios, lo que significa que algunas divisiones generarán más de dos ramas. Los campos de entrada y objetivo pueden ser continuos (rango numérico) o categóricos. CHAID exhaustivo es una modificación de CHAID que examina con mayor precisión todas las divisiones posibles, aunque necesita más tiempo para realizar los cálculos.



El nodo C5.0 genera un árbol de decisión o un conjunto de reglas. El modelo divide la muestra basándose en el campo que ofrece la máxima ganancia de información en cada nivel. El campo objetivo debe ser categórico. Se permiten varias divisiones en más de dos subgrupos.



El nodo Lista de decisiones identifica subgrupos, o segmentos, que muestran una mayor o menor posibilidad de proporcionar un resultado binario relacionado con la población global. Por ejemplo, puede buscar clientes que tengan menos posibilidades de abandonar o más posibilidades de responder favorablemente a una campaña. Puede incorporar su conocimiento empresarial al modelo añadiendo sus propios segmentos personalizados y previsualizando modelos alternativos uno junto a otro para comparar los resultados. Los modelos de listas de decisiones constan de una lista de reglas en las que cada regla tiene una condición y un resultado. Las reglas se aplican en orden, y la primera regla que coincide determina el resultado.



Los modelos de regresión lineal predicen un objetivo continuo tomando como base las relaciones lineales entre el destino y uno o más predictores.



El nodo PCA/Factorial proporciona técnicas eficaces de reducción de datos para reducir la complejidad de los datos. Análisis de componentes principales (PCA) busca combinaciones lineales de los campos de entrada que realizan el mejor trabajo a la hora de capturar la varianza en todo el conjunto de campos, en el que los componentes son ortogonales (perpendiculares) entre ellos. Análisis factorial intenta identificar factores subyacentes que expliquen el patrón de correlaciones dentro de un conjunto de campos observados. Para los dos métodos, el objetivo es encontrar un número pequeño de campos derivados que resume de forma eficaz la información del conjunto original de campos.



El nodo Selección de características filtra los campos de entrada para su eliminación en función de un conjunto de criterios (como el porcentaje de valores perdidos); a continuación, clasifica el grado de importancia del resto de entradas de acuerdo con un objetivo específico. Por ejemplo, a partir de un conjunto de datos dado con cientos de entradas potenciales, ¿cuáles tienen mayor probabilidad de ser útiles para el modelado de resultados de pacientes?



El análisis discriminante realiza más supuestos rigurosos que regresiones logísticas, pero puede ser una alternativa o un suplemento valioso al análisis de regresión logística si se cumplen dichos supuestos.



La regresión logística es una técnica de estadístico para clasificar los registros en función los valores de los campos de entrada. Es análoga a la regresión lineal pero toma un campo objetivo categórico en lugar de uno numérico.



El modelo lineal generalizado amplía el modelo lineal general, de manera que la variable dependiente está relacionada linealmente con los factores y las covariables mediante una determinada función de enlace. Además, el modelo permite que la variable dependiente tenga una distribución no normal. Cubre la funcionalidad de un amplio número de modelo estadísticos, incluyendo regresión lineal, regresión logística, modelos log lineales para recuento de datos y modelos de supervivencia censurados por intervalos.



Un modelo lineal mixto generalizado (GLMM) amplía el modelo lineal de modo que el objetivo pueda tener una distribución no normal, esté linealmente relacionado con los factores y covariables mediante una función de enlace especificada y las observaciones se puedan correlacionar. Los modelos lineales mixtos generalizados cubren una amplia variedad de modelos, desde modelos de regresión lineal simple hasta modelos multinivel complejos para datos longitudinales no normales.



El nodo Regresión de Cox le permite crear un modelo de supervivencia para datos de tiempo hasta el evento en presencia de registros censurados. El modelo produce una función de supervivencia que predice la probabilidad de que el evento de interés se haya producido en el momento dado (t) para valores determinados de las variables de entrada.



El nodo Máquina de vectores de soporte (SVM) le permite clasificar datos en uno o dos grupos sin que haya un ajuste por exceso. SVM funciona bien con conjuntos de datos grandes, como aquellos con un gran número de campos de entrada.



El nodo Red bayesiana le permite crear un modelo de probabilidad combinando pruebas observadas y registradas con conocimiento del mundo real para establecer la probabilidad de instancias. El nodo se centra en las redes Naïve Bayes aumentado a árbol (TAN) y de manto de Markov que se utilizan principalmente para la clasificación.



El nodo Modelo de respuesta de autoaprendizaje (SLRM) permite crear un modelo en el que un solo caso nuevo o un pequeño número de casos nuevos se pueden utilizar para volver a calcular el modelo sin tener que entrenar de nuevo el modelo utilizando todos los datos.



El nodo Serie temporal estima modelos de suavizado exponencial, modelos autorregresivos integrados de media móvil (ARIMA) univariados y modelos ARIMA (o de función de transferencia) multivariados para series temporales y genera previsiones. Un nodo Serie temporal debe ir siempre precedido por un nodo Intervalos de tiempo.



El nodo k de modelado de vecino (KNN) asocia el nuevo caso con la categoría o valor de los objetos k junto a él en el espacio de predictores, donde k es un entero. Los casos parecidos están próximos y los que no lo son están alejados entre sí.

Modelos de asociación

Los *modelos de asociación* encuentran patrones en los datos en los que una o más entidades (como eventos, compras o atributos) se asocian con una o más entidades. Los modelos construyen conjuntos de reglas que definen estas relaciones. Aquí los campos de los datos pueden funcionar como entradas y destinos. Podría encontrar estas asociaciones manualmente, pero los algoritmos de reglas de asociaciones lo hacen mucho más rápido, y pueden explorar patrones más complejos. Los modelos Apriori y Carma son ejemplos del uso de estos algoritmos. Otro tipo de modelo de asociación es el modelo de detección de secuencias, que encuentra patrones secuenciales en datos estructurados temporalmente.

Los modelos de asociación son los más útiles si se desean predecir varios resultados; por ejemplo, los clientes que adquirieron el producto X también adquirieron Y y Z. Los modelos de asociación relacionan una conclusión específica (como la decisión de adquirir un producto) con un conjunto de condiciones. La ventaja de los algoritmos de reglas de asociación sobre los algoritmos más estándar de árboles de decisión (C5.0 y Árbol C&R) es que las asociaciones pueden existir entre cualquiera de los atributos. Un algoritmo de árbol de decisión generará reglas con una única conclusión, mientras que los algoritmos de asociación tratan de buscar muchas reglas, cada una de las cuales puede tener una conclusión diferente.

Nodos de asociación



El nodo Apriori extrae un conjunto de reglas de los datos y destaca aquellas reglas con un mayor contenido de información. Apriori ofrece cinco métodos diferentes para la selección de reglas y utiliza un sofisticado esquema de indización para procesar eficientemente grandes conjuntos de datos. En los problemas de mucho volumen, Apriori se entrena más rápidamente, no tiene un límite arbitrario para el número de reglas que puede retener y puede gestionar reglas que tengan hasta 32 precondiciones. Apriori requiere que todos los campos de entrada y salida sean categóricos, pero ofrece un mejor rendimiento ya que está optimizado para este tipo de datos.



El modelo CARMA extrae un conjunto de reglas de los datos sin necesidad de especificar campos de entrada ni de objetivo. A diferencia de Apriori el nodo CARMA ofrece configuraciones de generación basadas en el soporte de las reglas (soporte tanto para el antecedente como el consecuente) en lugar de hacerlo sólo respecto al soporte del antecedente. Esto significa que las reglas generadas se pueden utilizar en una gama de aplicaciones más amplia, por ejemplo, para buscar una lista de productos o servicios (antecedentes) cuyo consecuente es el elemento que se desea promocionar durante esta temporada de vacaciones.



El nodo Secuencia encuentra reglas de asociación en datos secuenciales o en datos ordenados en el tiempo. Una secuencia es una lista de conjuntos de elementos que tiende a producirse en un orden previsible. Por ejemplo, si un cliente compra una cuchilla y una loción para después del afeitado, probablemente comprará crema para afeitarse la próxima vez que vaya a comprar. El nodo Secuencia se basa en el algoritmo de reglas de asociación de CARMA, que utiliza un método de dos pasos para encontrar las secuencias.

Modelos de segmentación

Los *modelos de segmentación* dividen los datos en segmentos o clústeres de registros que tienen patrones similares de campos de entrada. Como sólo se interesan por los campos de entrada, los modelos de segmentación no contemplan el concepto de campos de salida o destino. Ejemplos de modelos de segmentación son las redes Kohonen, la agrupación en clústeres de K-medias, la agrupación en clústeres en dos pasos y la detección de anomalías.

Los modelos de segmentación (también conocidos como "modelos de agrupación en clústeres") son útiles en aquellos casos en los que se desconoce el resultado específico (por ejemplo a la hora de detectar nuevos patrones de fraude o de identificar grupos de interés en la base de clientes). Los modelos de agrupación en clústeres se centran en la identificación de grupos de registros similares y en el etiquetado de registros según el grupo al que pertenecen. Esto se lleva a cabo sin la ventaja que ofrece el conocimiento previo sobre los grupos y sus características, y diferencia a los modelos de clústeres de otras técnicas de modelado en que no hay campos de salida u objetivo predefinidos para el modelo que se va a predecir. No hay respuestas correctas o incorrectas para estos modelos. Su valor viene determinado por su capacidad de capturar agrupaciones interesantes en los datos y proporcionar descripciones útiles de dichas agrupaciones. Los modelos de clúster se usan a menudo para crear clústeres o segmentos que se usan posteriormente como entradas en análisis posteriores, (por ejemplo mediante la segmentación de clientes potenciales en subgrupos homogéneos).

Nodos de segmentación



El nodo Agrupación en clústeres automática calcula y compara los modelos de agrupación en clústeres que identifican grupos de registros con características similares. El nodo funciona de la misma manera que otros nodos de modelado automático, permitiéndole experimentar con múltiples combinaciones de opciones en una única pasada de modelado. Los modelos se pueden comparar utilizando medidas básicas con las que se intenta filtrar y definir la utilidad de los modelos de clúster y proporcionar una medida según la importancia de campos concretos.



El nodo K-medias agrupa conjuntos de datos en grupos distintos (o clústeres). El método define un número fijo de clústeres, de forma iterativa asigna registros a los clústeres y ajusta los centros de los clústeres hasta que no se pueda mejorar el modelo. En lugar de intentar predecir un resultado, los modelos de *k*-medias utilizan un proceso conocido como aprendizaje no supervisado para revelar los patrones del conjunto de campos de entrada.



El nodo Kohonen genera un tipo de red neuronal que se puede usar para agrupar un conjunto de datos en grupos distintos. Cuando la red se termina de entrenar, los registros que son similares se deberían cerrar juntos en el mapa de resultados, mientras que los registros que son diferentes aparecerían aparte. Puede observar el número de observaciones capturadas por cada unidad en el nugget de modelo para identificar unidades fuertes. Esto le proporcionará una idea del número apropiado de clústeres.



El nodo Bietápico es un método de agrupación en clústeres de dos pasos. El primer paso es hacer una única pasada por los datos para comprimir los datos de entrada de la fila en un conjunto de subclústeres administrable. El segundo paso utiliza un método de agrupación en clústeres jerárquica para fundir progresivamente los subclústeres en clústeres cada vez más grandes. El bietápico tiene la ventaja de estimar automáticamente el número óptimo de clústeres para los datos de entrenamiento. Puede gestionar tipos de campos mixtos y grandes conjuntos de datos eficazmente.



El nodo Detección de anomalías identifica casos extraños, o valores atípicos, que no se ajustan a patrones de datos “normales”. Con este nodo, es posible identificar valores atípicos aunque no se ajusten a ningún patrón previamente conocido o no se realice una búsqueda exacta.

Modelos de minería interna de bases de datos

IBM SPSS Modeler admite la integración con herramientas de modelado y minería de datos que están disponibles en proveedores de bases de datos como Oracle Data Miner, IBM DB2 InfoSphere Warehouse y Microsoft Analysis Services. Podrá crear, puntuar y almacenar modelos dentro de la base de datos, todo desde la aplicación IBM SPSS Modeler. Para obtener los detalles completos, consulte el *IBM SPSS Modeler Manual de minería interna de bases de datos*, disponible en DVD.

Modelos de IBM SPSS Statistics

Si dispone de una copia de IBM SPSS Statistics instalada y con la licencia necesaria en su ordenador, puede acceder y ejecutar determinadas rutinas de IBM SPSS Statistics en IBM SPSS Modeler para generar y puntuar modelos.

Más información

También hay disponible información detallada sobre el modelado de algoritmos. Si desea obtener más información, consulte el *Manual de algoritmos de IBM SPSS Modeler*, disponible en el DVD del producto.

Ejemplos de minería de datos

La mejor forma de aprender a analizar los datos en la práctica es comenzar con un ejemplo. En el *Manual de aplicaciones de IBM SPSS Modeler* se incluyen varios ejemplos de aplicaciones, que le ofrecen introducciones breves y determinadas a métodos y técnicas de modelado específicos. Consulte el tema “Ejemplos de aplicaciones” en la página 5 para obtener más información.

Capítulo 5. Generación de rutas

Conceptos básicos de la generación de rutas

La minería de datos que usa IBM SPSS Modeler se centra en el proceso de ejecución de datos en una serie de nodos, que se denomina **ruta**. Esta serie de nodos representa las operaciones que van a realizarse en los datos, mientras que los enlaces entre los nodos indican la dirección del flujo de datos. Normalmente, se utiliza una ruta de datos para leer datos en IBM SPSS Modeler, ejecutarla a través de una serie de manipulaciones y, a continuación, enviarla a su destino, como puede ser una tabla o un visor.

Por ejemplo, imagine que desea abrir un origen de datos, añadir un campo nuevo, seleccionar los registros basados en los valores del campo nuevo y, después, mostrar los resultados en una tabla. En este caso, la ruta de datos estaría compuesta de cuatro nodos:



Un nodo Archivo var. configurado para leer los datos del origen de datos.



Un nodo Derivar que se utiliza para añadir al conjunto de datos el campo nuevo calculado.



Un nodo Seleccionar que se utiliza para establecer los criterios de selección y excluir los registros de la ruta de datos.



Un nodo Tabla que se utiliza para mostrar los resultados de las manipulaciones en pantalla.

Generación de rutas de datos

La interfaz exclusiva de IBM SPSS Modeler permite analizar los datos visualmente mediante diagramas de rutas de datos. En el nivel más básico, se puede generar una ruta de datos si se siguen los pasos siguientes:

- Añadir nodos al lienzo de rutas.
- Conectar los nodos para formar una ruta.
- Especificar cualquier opción del nodo o de la ruta.
- Ejecute la ruta.

Esta sección contiene información más detallada acerca del trabajo con nodos para crear rutas de datos más complejas. También describe opciones y configuraciones de los nodos y de las rutas. Para ver los ejemplos paso a paso de la generación de rutas a través de los datos que se incluyen con IBM SPSS Modeler (en la carpeta *Demos* de la instalación del programa), consulte “Ejemplos de aplicaciones” en la página 5

Cómo trabajar con nodos

Los nodos se utilizan en IBM SPSS Modeler para ayudar en la exploración de datos. En el espacio de trabajo hay distintos nodos que representan objetos y acciones diferentes. La paleta que se encuentra en la parte inferior de la ventana de IBM SPSS Modeler contiene todos los nodos posibles empleados para la generación de rutas.

Existen varios tipos de nodos. Los **Nodos de origen** introducen datos en la ruta, y se encuentran en la pestaña Orígenes de la paleta de nodos. Los **Nodos de proceso** realizan operaciones sobre registros y campos de datos individuales, y pueden encontrarse en las pestañas Operaciones con registros y Operaciones con campos de la paleta. Los **Nodos de resultado** generan una variedad de resultados para los modelos de datos, gráficos y resultados, y se incluyen en las pestañas Gráfico, Resultado y Exportar de la paleta de nodos. Los **Nodos de modelado** utilizan algoritmos estadísticos para crear nuggets de modelos y aparecen en la pestaña Modelado, y (si está activada) en la pestaña Modelado de bases de datos de la paleta de nodos. Consulte el tema “Paleta de nodos” en la página 16 para obtener más información.

Los nodos se conectan para formar rutas que, cuando se ejecutan, permiten visualizar relaciones y extraer conclusiones. Las rutas son como scripts: se pueden guardar y reutilizar con archivos de datos distintos.

Un nodo ejecutable que procesa los datos de ruta se conoce como un **nodo terminal**. Un nodo de modelado o de resultado es un nodo terminal si aparece al final de una ruta o una rama de ruta. No se pueden conectar más nodos a un nodo terminal.

Note: puede personalizar la paleta nodos. Consulte el tema “Personalización de la paleta de nodos” en la página 156 para obtener más información.

Adición de nodos a una ruta

Hay diversas formas de añadir nodos a una ruta desde la paleta de nodos:

- Pulsar dos veces un nodo de la paleta. *Note:* cuando se pulsa dos veces en un nodo, éste se conecta automáticamente a la ruta actual. Consulte el tema “Conexión de nodos en una ruta” para obtener más información.
- Arrastrar y soltar un nodo de la paleta al lienzo de rutas.
- Pulsar en un nodo de la paleta y, a continuación, pulsar en el lienzo de rutas.
- Seleccione una opción apropiada desde el menú Insertar de IBM SPSS Modeler.

Una vez que haya añadido un nodo al lienzo de rutas, pulse dos veces en el nodo para mostrar su cuadro de diálogo correspondiente. La disponibilidad de opciones depende del tipo de nodo que desee añadir. Si desea obtener información sobre controles específicos del cuadro de diálogo, pulse en su botón de **Ayuda**.

Eliminación de nodos

Para eliminar un nodo de la ruta de datos, pulse en la ruta y pulse la tecla Supr o pulse con el botón derecho en la ruta y seleccione **Eliminar** del menú.

Conexión de nodos en una ruta

Los nodos añadidos al lienzo de rutas no forman una ruta de datos a menos que se conecten. Las conexiones entre nodos indican la dirección de los datos a medida que fluyen de una operación a la siguiente. Existen varias maneras de conectar los nodos para formar una ruta: efectuando una doble pulsación, mediante el botón central del ratón o de forma manual. using the middle mouse button, or manually.

Para añadir y conectar nodos pulsando dos veces

La manera más sencilla de formar una ruta es pulsar dos veces con el ratón en los nodos de la paleta. Este método conecta automáticamente el nuevo nodo con el nodo seleccionado en el lienzo de rutas. Por ejemplo, si el lienzo contiene un nodo Base de datos, se puede seleccionar este nodo y, a continuación, pulsar dos veces en el nodo siguiente de la paleta, como el nodo Derivar. Esta acción conecta automáticamente el nodo Derivar al nodo Base de datos existente. Se puede repetir el proceso hasta que se llegue a un nodo terminal, como es un nodo Histograma o Tabla, momento en el que se conectará cualquier nodo nuevo a la última ruta del nodo terminal.

Conectar nodos mediante el botón central del ratón

En el lienzo de rutas, se puede pulsar y arrastrar desde un nodo a otro con el botón central del ratón. (Si el ratón no tiene un botón central, se puede simular esta acción pulsando la tecla Alt a la vez que arrastra con el ratón de un nodo a otro.)

Para conectar nodos manualmente

Si el ratón no tiene botón central y prefiere conectar nodos manualmente, puede utilizar el menú emergente de un nodo para conectarlo a otro nodo que ya se encuentra en el lienzo.

1. Pulse con el botón derecho en el nodo desde el que desea iniciar la conexión. Se abrirá el menú del nodo.
2. En el menú, pulse **Conectar**.
3. Aparecerá un icono de conexión en el nodo de inicio y en el cursor. Pulse en otro nodo del lienzo para conectar los dos nodos.

Se pueden seguir distintas directrices para conectar nodos. Si se intenta realizar cualquiera de los siguientes tipos de conexiones, aparecerá un mensaje de error:

- Una conexión con un nodo de origen
- Una conexión desde un nodo terminal
- Un nodo que posee más conexiones de entrada de las permitidas por el número máximo establecido
- Conexión de dos nodos que ya están conectados
- Circularidad (el dato vuelve a un nodo desde el que ya ha fluido)

Omisión de nodos en una ruta

Cuando se omite un nodo en la ruta de datos, todas las conexiones de entrada y de salida se sustituyen por conexiones que van directamente de los nodos de entrada a los de resultados. Todas las conexiones del nodo se eliminan si el nodo no tiene ni conexiones de entrada ni de salida, en lugar de volver a distribuirlos.

Por ejemplo, puede que haya una ruta que proporciona un nuevo campo, filtra los campos existentes y, a continuación, explora los resultados en un histograma y en una tabla. Si también desea ver el mismo gráfico y la misma tabla para los datos *antes* de que se filtren los campos, se pueden añadir más nodos Histograma y Tabla a la ruta o se puede omitir el nodo Filtrar. Cuando se omite un nodo Filtrar, las conexiones con el gráfico y la tabla pasan directamente desde el nodo Derivar. El nodo Filtrar se desconecta de la ruta.

Omitir un nodo

1. En el lienzo de rutas, pulse dos veces con el botón central del ratón en el nodo que desea omitir. También puede utilizar Alt+pulsar dos veces.

Note: se puede deshacer esta acción pulsando en **Deshacer** en el menú Edición o en Ctrl+Z.

Desactivación de nodos en una ruta

Los nodos de proceso con una única entrada en rutas se pueden desactivar, dando como resultado que el nodo se ignora durante la ejecución de la ruta. De esta forma se evita que tenga que eliminar u omitir el

nodo y podrá dejarlo conectado al resto de nodos. Podrá abrir y editar la configuración del nodo; sin embargo, las modificaciones no surtirán efecto hasta que vuelva a activar el nodo.

Por ejemplo, es posible que tenga una ruta que filtre varios campos y que cree modelos con el conjunto de datos reducidos. Si también desea crear los mismos modelos *sin* filtrar los campos, para ver si mejoran los resultados del modelo, puede desactivar el nodo Filtrar. Si desactiva el nodo Filtrar, las conexiones de los nodos de modelado pasan directamente desde el nodo Derivar al nodo Tipo.

Para desactivar un nodo

1. En el lienzo de rutas, pulse con el botón derecho en el nodo que desee desactivar.
2. En el menú emergente, pulse en **Desactivar nodo**.

También puede pulsar en **Nodo > Desactivar nodo** en el menú Editar. Si desea volver a incluir el nodo en la ruta, pulse **Activar nodo** de la misma manera.

Note: se puede deshacer esta acción pulsando en **Deshacer** en el menú Edición o en Ctrl+Z.

Adición de nodos a conexiones existentes

Se puede añadir un nuevo nodo entre dos nodos conectados arrastrando la flecha que conecta ambos nodos.

1. Pulse y arrastre con el botón central del ratón la flecha de conexión donde desea insertar el nodo. Si lo prefiere, para simular un botón central, puede mantener pulsada la tecla Alt a la vez que pulsa y arrastra el ratón.
2. Arrastre la conexión hasta el nodo que desea incluir y suelte el botón del ratón.

Note: se pueden eliminar las conexiones nuevas del nodo y restaurar la original mediante la **omisión** del nodo.

Eliminación de conexiones entre nodos

Para eliminar la conexión entre dos nodos:

1. Pulse con el botón derecho en la flecha de conexión.
2. En el menú, pulse **Eliminar conexión**.

Para eliminar todas las conexiones que van y proceden de un nodo, realice una de las siguientes acciones:

- Seleccione el nodo y pulse F3.
- Seleccione el nodo y, en el menú principal, pulse:

Editar > Nodo > Desconectar

Opciones de configuración de los nodos

Existen distintas opciones para personalizar nodos una vez que se han creado y conectado. Pulse con el botón derecho en un nodo y seleccione una de las opciones del menú.

- Pulse **Edición** para abrir el cuadro de diálogo del nodo seleccionado.
- Pulse en **Conectar** para conectar manualmente un nodo con otro.
- Pulse en **Desconectar** para eliminar todos los enlaces desde el nodo y hacia el nodo.
- Pulse en **Cambiar nombre y anotar** para abrir la pestaña Anotaciones del cuadro de diálogo de edición.
- Pulse en **Nuevo comentario** para añadir un comentario relacionado con el nodo. Consulte el tema “Adición de comentarios y anotaciones a nodos y rutas” en la página 55 para obtener más información.
- Pulse en **Desactivar nodo** para “ocultar” el nodo durante el procesamiento. Para que el nodo vuelva a ser divisible para su procesamiento, pulse en **Activar nodo**. Consulte el tema “Desactivación de nodos en una ruta” en la página 37 para obtener más información.

- Pulse en **Cortar** o **Eliminar** para eliminar los nodos seleccionados del lienzo de rutas. *Note:* si pulsa en **Cortar** se pueden pegar nodos, mientras que la opción **Eliminar** no permite esta acción.
- Pulse en **Copiar nodo** para realizar una copia del nodo sin conexiones. Este nodo puede añadirse a una ruta nueva o a una existente.
- Pulse en **Cargar nodo** para abrir un nodo guardado anteriormente y cargar las opciones en el nodo que se ha seleccionado. *Note:* los nodos tienen que ser del mismo tipo.
- Pulse en **Recuperar nodo** para recuperar un nodo de un IBM SPSS Collaboration and Deployment Services Repository conectado.
- Pulse en **Guardar nodo** para guardar los detalles del nodo en un archivo. Los detalles de un nodo se pueden cargar en otro nodo del mismo tipo.
- Pulse en **Almacenar nodo** para guardar el nodo seleccionado en un IBM SPSS Collaboration and Deployment Services Repository conectado.
- Pulse en **Caché** para expandir el menú con las opciones de almacenamiento en caché del nodo seleccionado.
- Pulse en **Correlación de datos** para expandir el menú con las opciones para correlacionar datos a un origen nuevo o para especificar campos obligatorios.
- Pulse en **Crear Supernodo** para expandir el menú con las opciones de creación de un Supernodo en la ruta actual.
- Pulse en **Generar nodo de datos de usuario** para sustituir el nodo seleccionado. Los ejemplos que genere este nodo tendrán los mismos campos que el nodo actual.
- Pulse en **Ejecutar desde aquí** para ejecutar todos los nodos terminales desde el nodo seleccionado.

Opciones de caché de los nodos

Para optimizar la ejecución de la ruta, se puede configurar una **caché** en cualquier nodo no terminal. Cuando se configura una caché en un nodo, ésta se rellena con los datos que pasan a través del nodo la próxima vez que se ejecuta la ruta de datos. En adelante, los datos se leerán de la caché (que está almacenada en disco en un directorio temporal) en lugar del origen de datos.

El almacenamiento en caché es más útil tras una operación que exige mucho tiempo de ejecución, como la ordenación, fusión o agregación. Por ejemplo, supongamos que tiene un nodo de origen configurado para leer los datos de ventas desde una base de datos y un nodo Agregar que resume las ventas por ubicación. Se puede configurar una caché en el nodo Agregar en lugar de hacerlo en el nodo de origen, ya que se pretende que la caché almacene los datos agregados, no todo el conjunto de datos.

Note: El almacenamiento en caché en nodos de origen, que simplemente guarda una copia de los datos originales a medida que se leen en IBM SPSS Modeler, no mejorará el rendimiento en la mayoría de circunstancias.

Los nodos con el almacenamiento en caché activado se muestran con un pequeño icono de documento en la esquina superior derecha. Cuando los datos se almacenan en caché en el nodo, el icono del documento es verde.

Para activar una caché

1. En el lienzo de rutas, pulse con el botón derecho del ratón en el nodo y pulse en **Caché** en el menú.
2. En el submenú de caché, pulse en **Activar**.
3. Para desactivar la caché, pulse con el botón derecho del ratón en el nodo y pulse **Desactivar** del submenú de caché.

Almacenamiento en caché de nodos en una base de datos

Para las rutas ejecutadas en una base de datos, los datos se pueden almacenar en caché en medio de la ruta en una tabla temporal en la base de datos en lugar de en el sistema de archivos. Al combinarlo con la optimización de SQL, se puede mejorar considerablemente el rendimiento. Por ejemplo, el resultado de

una ruta que fusiona varias tablas para crear una vista de minería de datos se puede guardar en caché y reutilizar cuando sea necesario. Al generar automáticamente SQL para todos los nodos posteriores en la ruta, el rendimiento se puede mejorar mucho más.

Si utiliza el almacenamiento en caché de la base de datos con cadenas con más de 255 caracteres, asegúrese de que hay un nodo Tipo anterior desde donde se lee el nodo de caché y los valores de campo, o bien, defina la longitud de la cadena mediante el parámetro `default_sql_string_length` en el archivo `options.cfg`. Al hacerlo, se asegura de que la columna correspondiente de la tabla temporal se define con la anchura correcta para acomodar las cadenas.

Para aprovechar el almacenamiento en caché en una base de datos, se debe activar el almacenamiento en caché de la base de datos y la optimización de SQL. Tenga en cuenta que la configuración de optimización de Server reemplaza la de Client. Consulte el tema “Configuración de opciones de optimización de las rutas” en la página 44 para obtener más información.

Con el almacenamiento en caché de la base de datos activado, sólo tiene que pulsar con el botón derecho en cualquier nodo no terminal para almacenar en caché los datos en ese punto, y la caché se creará automáticamente de forma directa en la base de datos la próxima vez que se ejecute la ruta. Si no se activa el almacenamiento en caché de la base de datos o la optimización de SQL, la caché se escribirá en el sistema de archivos en lugar de en la base de datos.

Nota: las bases de datos siguientes admiten tablas temporales con el objetivo de almacenar en caché: DB2, Netezza, Oracle, SQL Server y Teradata. Otras bases de datos utilizarán una tabla normal para el almacenamiento en caché de la base de datos. El código SQL puede personalizarse para bases de datos específicas; póngase en contacto con la asistencia técnica para obtener ayuda.

Para vaciar una caché

Un icono blanco de documento en un nodo indica que la caché está vacía. Cuando la caché está llena, el icono de documento aparece en color verde oscuro. Si desea reemplazar el contenido de la caché, debe vaciar la caché en primer lugar y, después, volver a ejecutar la ruta de datos para rellenarla.

1. En el lienzo de rutas, pulse con el botón derecho del ratón en el nodo y pulse en **Caché** en el menú.
2. En el submenú de caché, pulse en **Vaciar**.

Para guardar una caché

Se puede guardar el contenido de una caché como un archivo de datos IBM SPSS Statistics (*.sav). Se puede volver a cargar el archivo como una caché o configurar el nodo que utiliza el archivo caché como origen de datos. También se puede cargar una caché guardada perteneciente a otro proyecto.

1. En el lienzo de rutas, pulse con el botón derecho del ratón en el nodo y pulse en **Caché** en el menú.
2. En el submenú de caché, pulse en **Guardar caché**.
3. En el cuadro de diálogo Guardar caché, busque el lugar en que desea guardar el archivo caché.
4. Introduzca un nombre en el cuadro de texto Nombre de archivo.
5. Asegúrese de que está seleccionado *.sav en la lista Archivos de tipo y pulse en **Guardar**.

Para cargar una caché

Si se ha guardado un archivo caché antes de eliminarlo del nodo, es posible volver a cargarlo.

1. En el lienzo de rutas, pulse con el botón derecho del ratón en el nodo y pulse en **Caché** en el menú.
2. En el submenú de caché, pulse en **Cargar caché**.
3. En el cuadro de diálogo Cargar caché, busque la ubicación del archivo caché, selecciónelo y pulse en **Cargar**.

Vista previa de datos de nodos

Para garantizar que los datos se cambian de la manera esperada al crear una ruta, se pueden ejecutar los datos a través de un nodo de tabla en cada paso significativo. Para que no tenga que hacerlo, puede generar una vista previa de cada nodo en la que aparezca una muestra de los datos que se crearán, reduciendo con ello el tiempo que se tarda en crear cada nodo.

Para nodos anteriores de un nugget de modelo, la vista previa muestra los campos de entrada; para un nugget de modelo o nodos posteriores al nugget (salvo nodos terminales), la vista previa muestra campos de entrada y campos generados.

El número predeterminado de filas visualizadas es 10; sin embargo, puede cambiarlo en las propiedades de la ruta. Consulte el tema “Configuración de opciones generales de las rutas” en la página 42 para obtener más información.

En el menú Generar, puede crear varios tipos de nodos.

Bloqueo de nodos

Para evitar que otros usuarios cambien la configuración de un o más nodos en una ruta, puede encapsular el nodo o nodos en un tipo especial de nodo llamado Supernodo, y a continuación bloquear el supernodo aplicando una protección por contraseña.

Cómo trabajar con rutas

Una vez conectados los nodos de origen, de proceso y terminales en el lienzo de rutas, habrá creado una ruta. Al igual que una colección de nodos, las rutas se pueden guardar, anotar y añadir a los proyectos. También se pueden configurar varias opciones para las rutas, como optimización, configuración de fecha y hora, parámetros y scripts. Estas propiedades se describen en el tema siguiente.

En IBM SPSS Modeler, puede utilizar y modificar más de una ruta de datos en la misma sesión de IBM SPSS Modeler. La parte derecha de la ventana contiene el panel de gestores que le ayudará a desplazarse por las rutas, los resultados y los modelos que estén abiertos. Si no puede ver el panel de gestores, pulse en **Gestores** en el menú Ver y, a continuación, pulse en la pestaña **Rutas**.

Desde esta pestaña, podrá:

- Acceder a las rutas.
- Guardar rutas.
- Guardar rutas en el proyecto actual.
- Cerrar rutas.
- Abrir nuevas rutas.
- Almacenar y recuperar rutas desde un repositorio de IBM SPSS Collaboration and Deployment Services (si está disponible en su lugar de trabajo). Consulte el tema “Acerca de IBM SPSS Collaboration and Deployment Services Repository” en la página 119 para obtener más información.

Pulse con el botón derecho del ratón en la ruta de la pestaña Rutas para acceder a estas opciones.

Opciones de configuración de las rutas

Puede especificar diferentes opciones que se aplicarán a la ruta actual. También puede guardar estas opciones como valores predeterminados que se aplicarán a todas las rutas. Las opciones son las siguientes:

- **General.** Opciones generales como símbolos y codificación de texto que se utilizará en la ruta. Consulte el tema “Configuración de opciones generales de las rutas” en la página 42 para obtener más información.

- **Fecha/hora.** Opciones relacionadas con el formato de las expresiones de fecha y hora. Consulte el tema “Configuración de opciones de fecha y hora de las rutas” en la página 43 para obtener más información.
- **Formatos de numeración.** Opciones que controlan el formato de expresiones numéricas. Consulte el tema “Configuración de opciones de formato de número para rutas” en la página 44 para obtener más información.
- **Optimización.** Opciones de optimización del rendimiento de la ruta. Consulte el tema “Configuración de opciones de optimización de las rutas” en la página 44 para obtener más información.
- **Registro y estado.** Opciones de control de registro de SQL y estado del registro. Consulte el tema “Configuración de registro de SQL y opciones de estado de registro para rutas” en la página 46 para obtener más información.
- **Diseño.** Opciones relacionadas con el diseño de la ruta en el lienzo. Consulte el tema “Configuración de opciones de diseño de las rutas” en la página 46 para obtener más información.

Para configurar las opciones de las rutas

1. En el menú Archivo, pulse en **Propiedades de ruta** (o seleccione la ruta de la pestaña Rutas del panel de gestores, pulse con el botón derecho del ratón y, a continuación, seleccione **Propiedades de ruta** en el menú emergente).
2. Pulse en la pestaña **Opciones**.

En el menú Herramientas, también se puede pulsar en:

Propiedades de ruta > Opciones

Configuración de opciones generales de las rutas: Las opciones generales son un conjunto de opciones que se aplican a diferentes aspectos de la ruta actual.

Símbolo decimal. Seleccione coma (,) o punto (.) como separador decimal.

Símbolo de agrupación. Para los formatos de presentación de los números, seleccione el símbolo utilizado para agrupar valores (por ejemplo, el punto en 3.000,00). Entre las opciones se incluyen la ausencia de puntuación, el punto, la coma, el espacio y definido por el entorno local (en cuyo caso se utiliza el valor predeterminado del entorno local actual).

Codificación. Especifique el método predeterminado de la ruta para la codificación de texto. (*Nota:* se aplica a Var. Sólo el nodo origen Archivo y el nodo de exportación de Archivo sin formato. Ningún otro nodo utiliza este parámetro; la mayoría de los archivos de datos han incrustado información de codificación.) Puede elegir entre el valor predeterminado del sistema o UTF-8. El valor predeterminado del sistema se especifica en el Panel de control de Windows o, si lo ejecuta en modo distribuido, en el equipo servidor. Consulte el tema “Compatibilidad con Unicode en IBM SPSS Modeler” en la página 179 para obtener más información.

Evaluación de conjunto de reglas. Determina cómo se evalúan los modelos de conjuntos de reglas. Los conjuntos de reglas utilizan de forma predeterminada la opción **Elección** para combinar predicciones de reglas individuales y determinar la predicción final. Para garantizar que los conjuntos de reglas utilizan la regla de primer acierto de forma predeterminada, seleccione **Primer acierto**. Observe que esta opción no se aplica a los modelos de listas de decisiones, que siempre utilizan el primer acierto como define el algoritmo.

Número máximo de filas que aparecen en la vista previa de datos. Especifique el número de filas que aparecerá cuando se solicite una vista previa de los datos de un nodo. Consulte el tema “Vista previa de datos de nodos” en la página 41 para obtener más información.

Número máximo de miembros para los campos nominales. Permite seleccionar si se desea especificar un número máximo de miembros en los campos nominales (conjuntos) tras el cual el tipo de campo se

convierte en **sin tipo**. Esta opción puede resultar de utilidad cuando se trabaja con grandes campos nominales. *Note:* Cuando el nivel de medición de un campo está establecido en **sin tipo**, su rol se define directamente en **Ninguno**. Esto significa que los campos no están disponibles para el modelado.

Limitar tamaño de conjunto para creación de modelos de Kohonen y de K-medias. Permite seleccionar si se desea especificar un número máximo de miembros en los campos nominales utilizados en redes de Kohonen y de K-medias. El tamaño de conjunto predeterminado es 20; después de este valor, se ignora el campo y aparece una advertencia que proporciona información sobre el campo en cuestión.

Tenga en cuenta que, para obtener compatibilidad, esta opción se aplica también al nodo Red neuronal anterior que se ha sustituido en la versión 14 de IBM SPSS Modeler; algunas rutas heredadas pueden contener aún este nodo.

Renovar nodos de origen en la ejecución. Permite seleccionar la actualización automática de todos los nodos de origen cuando se ejecuta la ruta actual. Esta acción equivale a pulsar en el botón **Actualizar** de un nodo de origen, salvo que con ésta se actualizan automáticamente todos los nodos de origen (excepto los nodos Datos Usuario) de la ruta actual.

Note: si selecciona esta opción, se vaciarán las cachés de los nodos posteriores en la ruta, incluso sin haber modificado los datos. El vaciado tiene lugar una vez por ejecución de la ruta, lo que significa que aún se pueden utilizar las cachés que se encuentran por debajo en la ruta como almacenamiento temporal para una ejecución. Por ejemplo, imagine que ha definido una caché en medio de la ruta después de una compleja operación de derivación y que tiene varios gráficos e informes adjuntos por debajo de este nodo Derivar. Cuando se ejecute la ruta, la caché en el nodo Derivar se vaciará y rellenará, aunque solamente en el primer gráfico o informe. Los siguientes nodos terminales leerán los datos de la caché del nodo Derivar.

Mostrar etiquetas de valor y de campo en resultados. Muestra etiquetas de valor y de campo en tablas, gráficos y otros resultados. Si no hay etiquetas, se mostrarán en su lugar los valores de los datos y de los nombres de los campos. Las etiquetas se desactivan de forma predeterminada; no obstante, se pueden conmutar de forma individual en cualquier lugar de IBM SPSS Modeler. También se puede optar por mostrar las etiquetas en la ventana de resultados mediante un botón de gafas de la barra de herramientas.



Figura 11. Icono de la barra de herramientas utilizado para conmutar etiquetas de campos y de valores

Mostrar tiempos de ejecución. Muestra los tiempos de ejecución individuales de los nodos de ruta en la pestaña Tiempos de ejecución después de haber ejecutado la ruta. Consulte el tema “Visualización de tiempos de ejecución de nodos” en la página 48 para obtener más información.

Guardar como valor predeterminado. Las opciones especificadas se aplican solamente a la ruta actual. Pulse en este botón para establecer estas opciones como el valor predeterminado para todas las rutas.

Configuración de opciones de fecha y hora de las rutas: Estas opciones especifican el formato que se utilizará para diferentes expresiones de fecha y hora en la ruta actual.

Importar fecha/hora como. Seleccione si desea utilizar el almacenamiento de fecha/hora para los campos de fecha/hora o si desea importarlos como variables de cadena.

Formato de fecha. Seleccione un formato de fecha que usar en los campos de almacenamiento de fechas o cuando las funciones de fecha de CLEM interpreten las cadenas como fechas.

Formato de hora. Seleccione un formato de hora que utilizar en los campos de almacenamiento de horas o cuando las funciones de hora de CLEM interpreten las cadenas como horas.

Admitir fecha/mín. negativos. En el caso de los formatos de hora, seleccione si desea que las diferencias de fecha negativas se interpreten en relación al día o la hora anterior.

Línea base de fecha (1 de enero). Esta opción permite seleccionar los años de línea base (siempre 1 de enero) que utilizarán las funciones de fecha de CLEM que trabajan con una sola fecha.

Fechas de 2 dígitos comienzan a partir de. Esta opción especifica al año de corte para añadir dígitos de centenas para aquellos años expresados únicamente con dos dígitos. Por ejemplo, si se especifica 1930 como el año de corte, se asumirá que la fecha 05/11/02 pertenece al año 2002. Ocurrirá lo mismo con el siglo XX para las fechas posteriores al 30; por lo tanto, se asume que la fecha 05/11/73 corresponda al año 1973.

Guardar como valor predeterminado. Las opciones especificadas se aplican solamente a la ruta actual. Pulse en este botón para establecer estas opciones como el valor predeterminado para todas las rutas.

Configuración de opciones de formato de número para rutas: Estas opciones especifican el formato que se utilizará para diferentes expresiones numéricas en la ruta actual.

Formato de presentación de los números. Se puede elegir entre los formatos de presentación estándar (####,###), científico (#,###E+##) o de moneda (\$###,##).

Cifras decimales (estándar, científico, moneda). Para los formatos de presentación de los números, esta opción especifica el número de cifras decimales que se utilizan cuando se presentan o imprimen números reales. Esta opción se especifica de forma independiente en cada formato de presentación.

Cálculos en. Seleccione **Radianes** o **Grados** como unidad de medida utilizada en las expresiones trigonométricas de CLEM. Consulte el tema “Funciones trigonométricas” en la página 98 para obtener más información.

Guardar como valor predeterminado. Las opciones especificadas se aplican solamente a la ruta actual. Pulse en este botón para establecer estas opciones como el valor predeterminado para todas las rutas.

Configuración de opciones de optimización de las rutas: Puede utilizar la configuración de optimización para optimizar el rendimiento de la ruta. Tenga en cuenta que la configuración de la optimización y el rendimiento en IBM SPSS Modeler Server (si se utiliza) reemplaza cualquier configuración equivalente en el cliente.

Note: Optimización de SQL y modelado de bases de datos requieren que la conectividad de IBM SPSS Modeler Server esté activada en el equipo con IBM SPSS Modeler. Con esta configuración activada, puede acceder a los algoritmos de bases de datos, devolver SQL directamente desde IBM SPSS Modeler y acceder a IBM SPSS Modeler Server. Para verificar el estado de la licencia actual, seleccione las siguientes opciones en el menú de IBM SPSS Modeler.

Ayuda > Acerca de > Detalles adicionales

Si la conectividad está activada, verá la opción **Activación de servidor** en la pestaña Estado de licencia.

Consulte el tema “Conexión con IBM SPSS Modeler Server” en la página 12 para obtener más información.

Nota: la compatibilidad con retrotracción y optimización de SQL depende del tipo de base de datos en uso. Si desea obtener la información más reciente acerca de las bases de datos y los controladores ODBC compatibles y cuyo uso se ha comprobado con IBM SPSS Modeler 16, visite el sitio de asistencia corporativo <http://www.ibm.com/support>.

Activar reescritura de rutas. Seleccione esta opción para activar la reescritura de rutas en IBM SPSS Modeler. Hay cuatro tipos de reescritura disponibles, y puede seleccionar una o ambas. La reescritura reordena los nodos de una ruta en segundo plano para obtener un funcionamiento más eficiente sin alterar la semántica de la ruta.

- **Optimizar generación de SQL.** Esta opción permite reordenar los nodos de la ruta de manera que se puedan retrotraer más operaciones usando la generación de SQL para su ejecución en la base de datos. Si encuentra un nodo que no se puede representar en SQL, el optimizador leerá previamente para comprobar si hay algún nodo posterior en la ruta que se pueda representar en SQL y mover con seguridad al nodo problema sin que la semántica de la ruta se vea afectada. No sólo la base de datos puede realizar operaciones de un modo más eficiente que IBM SPSS Modeler, sino que los puntos de retrotracción actúan para reducir el tamaño del conjunto de datos que se devuelve a IBM SPSS Modeler para su procesamiento. A su vez, esto puede reducir el tráfico de red y acelerar las operaciones de las rutas. Tenga en cuenta que la casilla de verificación **Generar SQL** debe estar seleccionada para que la optimización de SQL tenga algún efecto.
- **Optimizar expresión CLEM.** Esta opción permite al optimizador buscar las expresiones CLEM que se pueden preprocesar antes de que se ejecute la ruta para poder aumentar la velocidad de proceso. Como ejemplo sencillo, si tiene una expresión tal como $\log(\text{salario})$, el optimizador calculará el valor del salario real y lo pasará al proceso. Esto se puede utilizar para mejorar los puntos de retrotracción de SQL y el rendimiento de IBM SPSS Modeler Server.
- **Optimizar ejecución de sintaxis.** Este método de reescritura de rutas incrementa la eficiencia de las operaciones que incorporan más de un nodo con la sintaxis de IBM SPSS Statistics. La optimización se logra combinando los comandos de sintaxis en una única operación, en lugar de ejecutar cada operación como separada.
- **Optimizar otra ejecución.** Este método de reescritura de rutas incrementa la eficiencia de las operaciones que no se pueden delegar a la base de datos. La optimización se consigue reduciendo la cantidad de datos de la ruta lo antes posible. Al tiempo que se mantiene la integridad de los datos, la ruta se reescribe para colocar las operaciones más cerca del origen de datos, reduciendo de este modo los datos de la parte de abajo en la ruta en el caso de operaciones costosas, como las uniones.

Activar procesamiento paralelo. Cuando trabaje en un ordenador con varios procesadores, esta opción permite al sistema equilibrar la carga a través de estos procesadores, que puede resultar en rendimiento más rápido. El uso de varios nodos, o de los siguientes nodos individuales, puede mejorar gracias al procesamiento paralelo: C5.0, Fundir (por clave), Ordenar, Intervalo (métodos de mosaico y rango) y Agregar (utilizando uno o más campos clave).

Generar SQL. Seleccione esta opción para activar la optimización de SQL, permitiendo que las operaciones de ruta se retrotraigan a la base de datos mediante el código SQL para generar los procesos de ejecución, que puede mejorar el rendimiento. Para mejorar aún más el rendimiento, se puede seleccionar **Optimizar generación de SQL** para maximizar el número de operaciones retrotraídas a la base de datos. Si las operaciones de un nodo se han retrotraído a la base de datos, el nodo se resaltará en color púrpura cuando se ejecute la ruta.

- **Almacenamiento en caché de base de datos.** Para que se ejecuten las rutas que generan SQL en la base de datos, los datos se pueden almacenar en caché en medio de la ruta en una tabla temporal en la base de datos en lugar de en el sistema de archivos. Al combinarlo con la optimización de SQL, se puede mejorar considerablemente el rendimiento. Por ejemplo, el resultado de una ruta que fusiona varias tablas para crear una vista de minería de datos se puede guardar en caché y reutilizar cuando sea necesario. Con el almacenamiento en caché de la base de datos activado, sólo tiene que pulsar con el botón derecho del ratón en cualquier nodo no terminal para almacenar en caché los datos en ese punto, y la caché se creará de forma directa y automática en la base de datos la próxima vez que se ejecute la ruta. Esto permite generar SQL para nodos posteriores en la ruta, mejorando aún más el

rendimiento. Si lo prefiere, esta opción se puede desactivar si es necesario, por ejemplo, cuando las directivas o los permisos excluyen los datos que se escriben en la base de datos. Si no se activa el almacenamiento en caché de la base de datos o la optimización de SQL, la caché se escribirá en el sistema de archivos en lugar de en la base de datos. Consulte el tema “Opciones de caché de los nodos” en la página 39 para obtener más información.

- **Utilizar conversión relajada.** Esta opción permite la conversión de datos de cadenas a números y viceversa si están almacenados en un formato adecuado. Por ejemplo, si los datos se almacenan como cadena en una base de datos, pero contienen un número significativo, los datos se pueden convertir para utilizarlos cuando se realice la retrotracción.

Note: Debido a diferencias menores en la implementación de SQL, las rutas ejecutadas en una base de datos pueden devolver resultados ligeramente diferentes cuando se ejecutan en IBM SPSS Modeler. Por motivos similares, estas diferencias también pueden variar en función del proveedor de la base de datos.

Guardar como valor predeterminado. Las opciones especificadas se aplican solamente a la ruta actual. Pulse en este botón para establecer estas opciones como el valor predeterminado para todas las rutas.

Configuración de registro de SQL y opciones de estado de registro para rutas: Esta configuración incluye las diferentes opciones que controlan la visualización de declaraciones SQL que genera la ruta y la visualización del número de registros que procesa la ruta.

Mostrar SQL en el registro de mensajes durante la ejecución de la ruta. Especifica si el SQL generado durante la ejecución de la ruta se debe pasar al registro de mensajes.

Mostrar detalles de generación de SQL en el registro de mensajes durante la preparación de la ruta. Durante la presentación preliminar de la ruta, especifica si una presentación preliminar del SQL que se generaría se pasa al registro de mensajes.

Mostrar SQL. Especifica si cualquier SQL que aparece en el registro debe contener funciones SQL nativas o funciones ODBC estándar del tipo {fn FUNC(...)} tal como las genera IBM SPSS Modeler. La primera confía en la funcionalidad del controlador ODBC que puede no implementarse. Por ejemplo, este control no tendría efecto alguno para SQL Server.

Volver a dar formato a SQL para mejorar la legibilidad. Especifica si se debe dar formato a SQL que se muestre en el registro para mejorar la legibilidad.

Mostrar estado de registros. Especifica si los registros se deben generar cuando lleguen a los nodos terminales. Especifique un número que se utilizará para actualizar el estado cada *N* registros.

Guardar como valor predeterminado. Las opciones especificadas se aplican solamente a la ruta actual. Pulse en este botón para establecer estas opciones como el valor predeterminado para todas las rutas.

Configuración de opciones de diseño de las rutas: Esta configuración proporciona diferentes opciones relacionadas con la visualización y uso del lienzo de rutas.

Anchura mínima del lienzo de rutas. Permite especificar la anchura mínima del lienzo de rutas en píxeles.

Altura mínima del lienzo de rutas. Permite especificar la altura mínima del lienzo de rutas en píxeles.

Velocidad de desplazamiento de ruta. Especifique la velocidad de desplazamiento del lienzo de rutas para controlar la rapidez con la que el panel del lienzo de rutas se desplaza si se arrastra un nodo desde un lugar a otro en el lienzo. Los números más altos determinan una velocidad de desplazamiento superior.

Máximo para nombre de icono. Permite especificar un límite de caracteres para los nombres de los nodos del lienzo de rutas.

Tamaño de icono. Seleccione una opción para adaptar la totalidad de la vista de ruta a uno de los tamaños disponibles entre el 8% y el 200% del tamaño de icono estándar.

Tamaño de casilla de cuadrícula. Seleccione un tamaño de casilla de cuadrícula en la lista. Este número se utiliza para la alineación de nodos en el lienzo de rutas mediante una cuadrícula invisible. El tamaño de casilla de cuadrícula predeterminado es 0,25.

Ajustar a la cuadrícula. Permite seleccionar la alineación de iconos con un patrón de cuadrícula invisible (seleccionado de forma predeterminada).

Ubicación del icono generado. Seleccione en qué lugar del lienzo deben colocarse los iconos de los nodos generados a partir de nuggets de modelo. El valor predeterminado es la parte superior izquierda.

Guardar como valor predeterminado. Las opciones especificadas se aplican solamente a la ruta actual. Pulse en este botón para establecer estas opciones como el valor predeterminado para todas las rutas.

Propiedades de la ruta de Analytic Server: Los siguientes valores proporcionan una serie de opciones para trabajar con Analytic Server.

Número máximo de registros que se procesan fuera de Analytic Server

Especifica el número máximo de registros que se importan en el servidor de SPSS Modeler procedentes de un origen de datos de Analytic Server.

Notificación cuando un nodo no puede procesarse en Analytic Server

Este valor determina lo que ocurre cuando una secuencia que se va a enviar a Analytic Server contiene un nodo que no puede procesarse en Analytic Server. Especifica si debe emitirse un aviso y continuar procesándose la secuencia, o si se genera un error y se detiene el procesamiento.

Valores de almacenamiento de modelos divididos

Almacenar modelos divididos por referencia en Analytic Server cuando el tamaño (MB) del modelo sobrepase

Los nuggets (fragmentos) de modelo se almacenan habitualmente como parte de la secuencia. Los modelos divididos en muchas partes pueden dar lugar a nuggets de gran tamaño, y el trasiego de dichos nuggets entre la secuencia y Analytic Server puede tener un impacto negativo en el rendimiento. Para resolver este problema, cuando un modelo dividido sobrepasa el tamaño especificado, se almacena en Analytic Server, y el nugget que está en SPSS Modeler contiene una referencia al modelo.

Carpeta predeterminada para almacenar modelos por referencia en Analytic Server una vez finalizada la ejecución

Especifica la ruta predeterminada donde se almacenan los modelos divididos en Analytic Server. La ruta debe comenzar con un nombre válido de proyecto de Analytic Server.

Carpeta para almacenar los modelos ascendidos

Especifica la ruta predeterminada en la que se desean almacenar los modelos "ascendidos". Un modelo ascendido no se limpia cuando termina la sesión de SPSS Modeler.

Visualización de los mensajes de la operación de una ruta

Los mensajes relacionados con las operaciones de rutas, como la ejecución, la optimización y el tiempo transcurrido para la generación y evaluación de modelos, se pueden ver fácilmente mediante la pestaña Mensajes del cuadro de diálogo de propiedades de la ruta. Los mensajes de error también se notifican en esta tabla.

Para ver los mensajes de ruta

1. En el menú Archivo, pulse en **Propiedades de ruta** (o seleccione la ruta de la pestaña Rutas del panel de gestores, pulse con el botón derecho del ratón y, a continuación, seleccione **Propiedades de ruta** en el menú emergente).
2. Pulse en la pestaña **Mensajes**.

En el menú Herramientas, también se puede pulsar en:

Propiedades de ruta > Mensajes

Además de los mensajes relacionados con las operaciones de ruta, los mensajes de error también se notifican aquí. Cuando se detiene la ejecución de la ruta debido a un error, este cuadro de diálogo se abrirá en la pestaña Mensajes con el mensaje de error visible. Además, el nodo con errores se resalta en rojo en el lienzo de rutas.

Si se activan las opciones de optimización y registro de SQL en el cuadro de diálogo Opciones de usuario, también se mostrará la información generada en SQL. Consulte el tema “Configuración de opciones de optimización de las rutas” en la página 44 para obtener más información.

Se pueden guardar mensajes que se hayan notificado aquí para una ruta pulsando **Guardar mensajes** en la lista desplegable del botón Guardar (a la izquierda, justo debajo de la pestaña Mensajes). También se pueden eliminar los mensajes de una ruta determinada pulsando **Borrar todos los mensajes** en la lista del botón Guardar.

Visualización de tiempos de ejecución de nodos

En la pestaña Mensajes puede elegir que se muestre la pestaña Tiempos de ejecución, donde podrá ver los tiempos de ejecución individuales de todos los nodos de la ruta que se ejecutan en IBM SPSS Modeler Server. Tenga en cuenta que los tiempos pueden no ser precisos para las rutas que se ejecutan en otras áreas, tales como R o Analytic Server.

Note: Para que esta característica funcione, la casilla de verificación **Mostrar tiempos de ejecución** debe estar seleccionada en el ajuste **General** de la pestaña **Opciones**.

En la tabla de tiempos de ejecución de nodos, las columnas son las siguientes. Pulse en la cabecera de una columna para disponer las entradas por orden ascendente o descendente (por ejemplo, para ver qué nodos tienen los tiempos de ejecución más largos).

Nodo terminal. El identificador de la rama a la que pertenece el nodo. El identificador es el nombre del nodo terminal del extremo de la rama.

Etiqueta de nodo. El nombre del nodo al que hace referencia el tiempo de ejecución.

ID de nodo. El identificador exclusivo del nodo al que hace referencia el tiempo de ejecución. Este identificador se genera por parte del sistema cuando se crea el nodo.

Tiempos de ejecución. El tiempo en segundos que se tarda en ejecutar este nodo.

Cómo configurar parámetros de sesión y ruta.

Se pueden definir los parámetros para utilizarlos en scripts y expresiones CLEM. Son, de hecho, variables definidas por el usuario que se guardan y conservan con la ruta actual, sesión, o Supernodo, y a los que se puede acceder tanto desde la interfaz de usuario como a través de scripts. Si, por ejemplo, se guarda una ruta, cualquier conjunto de parámetros para esa ruta también se guarda. (Así se distinguen de las variables de script local, que sólo se pueden utilizar en el script en que se declaran.) Con frecuencia los parámetros se utilizan en los scripts para controlar el comportamiento del script, proporcionando información sobre los campos y valores que no necesitan estar codificados internamente en el script.

El ámbito de un parámetro depende de dónde se establezca:

- los parámetros de ruta se pueden establecer en un script de ruta o en el cuadro de diálogo de propiedades de ruta, y están disponibles para todos los nodos de la ruta. Se muestran en la lista Parámetros del generador de expresiones.
- Los parámetros de sesión se pueden establecer en un script autónomo o en el cuadro de diálogo de parámetros de sesión. Están disponibles para todas las rutas utilizadas en la sesión actual (todas las rutas enumeradas en la pestaña Rutas del panel de gestores).

También se pueden configurar los parámetros para Supernodos, donde sólo estarán visibles para los nodos encapsulados dentro del Supernodo.

Para configurar los parámetros de sesión y ruta a través de la interfaz de usuario

1. Para configurar los parámetros de ruta, en el menú principal, pulse:
Herramientas > Propiedades de ruta > Parámetros
2. Para configurar los parámetros de sesión, pulse en **Definir parámetros de sesión** en el menú Herramientas.

¿Solicitar? Marque esta casilla si desea que se le solicite a un usuario en tiempo de ejecución que introduzca un valor para este parámetro.

Nombre. Los nombres de los parámetros se enumeran aquí. Se puede crear un parámetro nuevo escribiendo un nombre en este campo. Por ejemplo, si desea crear un parámetro para la temperatura mínima, puede escribir `minvalue`. No incluya el prefijo `$P-` que denota un parámetro en las expresiones CLEM. Este nombre se utiliza para su representación en el generador de expresiones de CLEM.

Nombre largo. Indica el nombre descriptivo de cada parámetro creado.

Almacenamiento. Seleccione un tipo de almacenamiento de la lista. Indica cómo se almacenan los valores de datos en el parámetro. Por ejemplo, cuando trabaje con valores con ceros iniciales que desee conservar (como 008), debe seleccionar **Cadena** como el tipo de almacenamiento. De lo contrario, los ceros se eliminarán del valor. Los tipos de almacenamiento disponibles son cadena, entero, real, hora, fecha y marca de tiempo. Tenga en cuenta que para los parámetros de fecha, los valores se deben especificar utilizando la notación estándar ISO tal y como se detalla en el siguiente párrafo.

Valor. Indica el valor actual para cada parámetro. Ajuste el parámetro como desee. Tenga en cuenta que para los parámetros de fecha, los valores se deben especificar en la notación estándar ISO (que es AAAA-MM-DD). No se aceptan fechas especificadas en otros formatos.

Tipo (opcional). Si desea desplegar la ruta en una aplicación externa, seleccione un nivel de medición de la lista. En caso contrario, se recomienda dejar la columna *Tipo* tal como está. Si desea especificar restricciones de valores para el parámetro, como límites superiores e inferiores para un rango numérico, seleccione **Especificar** en la lista.

Tenga en cuenta que las opciones de tipo, almacenamiento y nombre largo se pueden establecer para los parámetros sólo a través de la interfaz de usuario. Estas opciones no se pueden establecer utilizando scripts.

Pulse en las flechas de la derecha para mover el parámetro seleccionado hacia arriba o hacia abajo en la lista de parámetros disponibles. Utilice el botón de eliminación (marcado con una X) para eliminar el parámetro seleccionado.

Especificación de solicitudes en tiempo de ejecución para valores de parámetros

Si tiene rutas donde es posible que necesite introducir valores diferentes para el mismo parámetro en diferentes ocasiones, puede especificar solicitudes en tiempo de ejecución para una o más rutas o valores de parámetros de sesión.

Parámetros. (Opcional) Introduzca un valor para el parámetro o deje el valor predeterminado si ya hay uno.

Desactivación de estas solicitudes. Seleccione esta casilla si no desea que estas solicitudes aparezcan cuando ejecute la ruta. Puede provocar que se vuelvan a mostrar seleccionando la casilla **¿Solicitar?** en las propiedades de la ruta o en el cuadro de diálogo de propiedades de la sesión donde se definieron los parámetros. Consulte el tema “Cómo configurar parámetros de sesión y ruta.” en la página 48 para obtener más información.

Especificación de restricciones de valores para un tipo de parámetro

Puede realizar restricciones de valores para un parámetro disponible durante el despliegue de una ruta en una aplicación externa que lea las rutas de modelado de datos. Este cuadro de diálogo permite especificar los valores disponibles para un usuario externo que ejecute la ruta. Las restricciones de los valores varían de forma dinámica en el cuadro de diálogo en función del tipo de datos. Las opciones que aparecen aquí son idénticas a las opciones disponibles para los valores del nodo Tipo.

Tipo. Muestra el nivel de medición seleccionado actualmente. Se pueden cambiar este valor para reflejar la forma en la que piensa utilizar el parámetro en IBM SPSS Modeler.

Almacenamiento. Muestra el tipo de almacenamiento si se conoce. Los tipos de almacenamiento no resultan afectados por el nivel de medición (continuo, nominal o marca) escogido para trabajar en IBM SPSS Modeler. El tipo de almacenamiento se puede modificar en la pestaña Parámetros principal.

La mitad inferior del cuadro de diálogo cambia de forma dinámica en función del nivel de medición seleccionado en el campo **Tipo**.

Niveles de medición continuos

Inferior. Especifica un límite inferior para los valores del parámetro.

Superior. Especifica un límite superior para los valores del parámetro.

Etiquetas Puede especificar etiquetas para cualquier valor de un campo de rango. Pulse en el botón **Etiquetas** para abrir un cuadro de diálogo independiente con el fin de especificar las etiquetas de valores.

Niveles de medición nominales

Valores. Esta opción permite especificar los valores de un parámetro que se va a utilizar como un campo nominal. En la ruta de IBM SPSS Modeler no se fuerzan los valores pero se incluyen en una lista desplegable para aplicaciones de despliegue externas. Puede modificar valores existentes y reordenar o eliminar valores con los botones de flecha o de eliminación.

Niveles de medición marca

Verdadero. Especifica un valor de marca para el parámetro cuando se cumple la condición.

Falso. Especifica un valor de marca para el parámetro cuando no se cumple la condición.

Etiquetas Puede especificar etiquetas para los valores de un campo de marca.

Opciones de despliegue de rutas

La pestaña Despliegue del cuadro de diálogo de propiedades de la ruta le permite especificar las opciones para desplegar la ruta como un escenario dentro de IBM SPSS Collaboration and Deployment Services para la actualización de modelos, la planificación de trabajos automatizados o para su uso en IBM Analytical Decision Management. Antes de su despliegue, todas las rutas requieren una rama de

puntuación específica; las opciones y requisitos adicionales dependen del tipo de despliegue. Consulte el tema “Almacenamiento y despliegue de objetos de repositorio” en la página 120 para obtener más información.

Ejecución de bucles para rutas

Utilizando la pestaña Ejecución en el cuadro de diálogo de propiedades de ruta, podrá configurar condiciones de bucle para automatizar tareas repetitivas en la ruta actual.

Una vez que ha definido estas condiciones, podrá utilizarlo como introducción para los scripts ya que llena la ventana de script con scripts básicos para la ruta con los que pueden modificar, quizás para utilizar como base desde la cual generar mejores scripts. Consulte el tema “Funciones globales” en la página 114 para obtener más información.

Para definir bucles para una ruta

1. En el menú Archivo, pulse en **Propiedades de ruta** (o seleccione la ruta de la pestaña Rutas del panel de gestores, pulse con el botón derecho del ratón y, a continuación, seleccione **Propiedades de ruta** en el menú emergente).
2. Pulse en la pestaña **Ejecución**.
3. Seleccione el modo de ejecución **Ejecución de bucles / condicional**.
4. Pulse la pestaña **Bucles**.

En el menú Herramientas, también se puede pulsar en:

Propiedades de ruta > Ejecución

Como alternativa adicional, pulse con el botón derecho del ratón en el nodo y en el menú contextual, pulse:

Ejecución de bucles / condicional > Editar configuración de bucles

Iteración. No puede editar este valor de número de fila, pero puede añadir, suprimir o subir o bajar una iteración utilizando los botones de la derecha de la tabla.

Cabeceras de tabla. Reflejan la clave de iteración y cualquiera de las variables creadas al configurar el bucle.

Visualización de valores globales para rutas

A través de la pestaña Valores globales del cuadro de diálogo de propiedades de la ruta, se pueden ver los valores globales establecidos para la ruta actual. Los valores globales se crean utilizando el nodo Val. globales para determinar estadísticos como la media, la suma o la desviación estándar de los campos seleccionados.

Una vez ejecutado el nodo Val. globales, estos valores pueden utilizarse de distintas formas en las operaciones de ruta. Consulte el tema “Funciones globales” en la página 114 para obtener más información.

Para ver los valores globales de una ruta

1. En el menú Archivo, pulse en **Propiedades de ruta** (o seleccione la ruta de la pestaña Rutas del panel de gestores, pulse con el botón derecho del ratón y, a continuación, seleccione **Propiedades de ruta** en el menú emergente).
2. Pulse en la pestaña **Valores globales**.

En el menú Herramientas, también se puede pulsar en:

Propiedades de ruta > Valores globales

Valores globales disponibles. Los valores globales disponibles se enumeran en esta tabla. Aquí no se pueden editar valores globales, aunque sí se pueden eliminar de una ruta mediante el botón Borrar todos los valores globales, situado en la parte derecha de la tabla.

Búsqueda de nodos en una ruta

Puede buscar nodos en una ruta especificando un número de criterios de búsqueda, como el nombre, categoría e identificador de nodo. Esta característica puede ser de especial utilidad en el caso de rutas complejas que contengan un gran número de nodos.

Buscar nodos en una ruta

1. En el menú Archivo, pulse en **Propiedades de ruta** (o seleccione la ruta de la pestaña Rutas del panel de gestores, pulse con el botón derecho del ratón y, a continuación, seleccione **Propiedades de ruta** en el menú emergente).
2. Pulse en la pestaña **Buscar**.

En el menú Herramientas, también se puede pulsar en:

Propiedades de ruta > Buscar

Puede especificar más de una opción para limitar la búsqueda, a excepción de que la búsqueda por el ID de nodo (mediante el campo **ID es igual a**) excluye las otras opciones.

La etiqueta de nodo contiene. Marque esta casilla e introduzca todo o parte de una etiqueta de nodo para buscar un nodo en particular. Las búsquedas no distinguen entre mayúsculas y minúsculas, y varias palabras se consideran una única porción de texto.

Categoría de nodo. Marque esta casilla y seleccione una categoría de la lista para buscar un tipo de nodo en particular. Un **Nodo de proceso** es un nodo de la pestaña Operaciones con campos o la pestaña Operaciones con registros de la paleta de nodos; **Aplicar nodo de modelo** hace referencia a un nugget de modelo.

Las palabras clave incluyen. Marque esta casilla e introduzca una o más palabras clave completas para buscar nodos en los que se hayan introducido texto en el campo Palabras clave de la pestaña Anotaciones del cuadro de diálogo de nodo. El texto que introduzca en Palabra clave debe ser una cadena exacta. Separe varias palabras clave con punto y coma para buscar alternativas (por ejemplo, si introduce protón;neutrón se buscarán todos los nodos con alguna de estas palabras clave. Consulte el tema "Anotaciones" en la página 60 para obtener más información.

La anotación contiene. Marque esta casilla e introduzca una o más palabras para buscar nodos que contengan este texto en el área de texto principal de la pestaña Anotaciones del cuadro de diálogo de nodo. Las búsquedas no distinguen entre mayúsculas y minúsculas, y varias palabras se consideran una única porción de texto. Consulte el tema "Anotaciones" en la página 60 para obtener más información.

Campo generado denominado. Marque esta casilla e introduzca el nombre de un campo generado (por ejemplo, \$C-Drug). Puede utilizar esta opción para buscar nodos de modelado que generen un campo específico. Introduzca solamente un nombre de campo, que debe coincidir exactamente.

ID es igual a. Marque esta casilla e introduzca un ID de nodo para buscar un nodo en particular que tenga dicho identificador (si selecciona esta opción, desactivará todas las opciones anteriores). El sistema asigna los ID de nodo cuando se crea el nodo. Estos pueden utilizarse para hacer referencia al nodo para scripts o automatización. Introduzca solamente un ID de nodo, que debe coincidir exactamente. Consulte el tema "Anotaciones" en la página 60 para obtener más información.

Buscar en supernodos. Esta casilla está marcada de forma predeterminada, lo que significa que la búsqueda se realizará tanto en los nodos internos como en los nodos externos a los supernodos. Elimine la selección de la casilla si desea realizar la búsqueda solamente en los nodos externos a los supernodos, en el nivel superior de la ruta.

Buscar. Cuando haya especificado todas las opciones que desee, pulse en este botón para iniciar la búsqueda.

Los nodos que tengan las opciones especificadas aparecerán en la parte inferior del cuadro de diálogo. Seleccione un nodo de la lista para resaltarlo en el lienzo de rutas.

Cambio de nombres de rutas

Si se utiliza la pestaña Anotaciones del cuadro de diálogo de propiedades de la ruta, se pueden añadir anotaciones descriptivas y crear un nombre personalizado para la ruta. Estas opciones resultan especialmente útiles cuando se generan informes para rutas añadidas al panel de proyectos. Consulte el tema “Anotaciones” en la página 60 para obtener más información.

Descripciones de ruta

Por cada ruta que cree, IBM SPSS Modeler produce una descripción de ruta con información del contenido de la ruta. Esto puede resultar de utilidad si está intentando ver lo que hace una ruta pero no tiene IBM SPSS Modeler instalado, por ejemplo, cuando accede a una ruta a través de IBM SPSS Collaboration and Deployment Services.

La descripción de ruta se muestra en forma de documento HTML compuesto por un cierto número de secciones.

Información general de ruta

Esta sección contiene el nombre de la ruta, junto con los detalles de cuándo se creó la ruta y cuál fue la última vez que se guardó.

Descripción y comentarios

Esta sección incluye:

- Anotaciones de la ruta (consulte “Anotaciones” en la página 60)
- Comentarios no relacionados con nodos específicos
- Comentarios relacionados con nodos en las ramas de modelado y puntuación de la ruta

Información de puntuación

Esta sección contiene información bajo diversas cabeceras relativas a la rama de puntuación de la ruta.

- **Comentarios.** Incluye comentarios vinculados únicamente a nodos de la rama de puntuación.
- **Entradas.** Enumera los campos de entrada junto con sus tipos de almacenamiento (por ejemplo, cadena, entero, real, etc.).
- **Resultados.** Enumera los campos de resultados, incluidos los campos adicionales generados por el nodo de modelado, junto con sus tipos de almacenamiento.
- **Parámetros.** Enumera los parámetros relativos a la rama de puntuación de la ruta que pueden visualizarse o editarse cada vez que se puntúa el modelo. Estos parámetros se identifican cuando pulsa en el botón **Parámetros de puntuación** de la pestaña **Despliegue** del cuadro de diálogo Propiedades de ruta.
- **Nodo Modelo.** Muestra el nombre y el tipo del modelo (por ejemplo, Red neuronal, C&RT, etc.). Éste es el nugget de modelo seleccionado para el campo **Nodo de modelo** de la pestaña **Despliegue** del cuadro de diálogo Propiedades de ruta.

- **Detalles del modelo.** Muestra detalles del nugget de modelo identificado en la cabecera anterior. Cuando sea posible, se incluirán gráficos de importancia de predictores y diagramas de evaluación para el modelo.

Información de modelado

Contiene información relativa a la rama de modelado de la ruta.

- **Comentarios.** Enumera los comentarios o anotaciones conectados con los nodos de la rama de modelado.
- **Entradas.** Enumera los campos de entrada junto con su rol en la rama de modelado (con la forma del valor de rol del campo, por ejemplo, Entrada, Destino, Dividir, etc.).
- **Parámetros.** Enumera los parámetros relativos a la rama de modelado de la ruta que pueden visualizarse o editarse cada vez que se actualiza el modelo. Estos parámetros se identifican cuando pulsa en el botón **Parámetros del generador de modelos** de la pestaña **Despliegue** del cuadro de diálogo Propiedades de ruta.
- **Nodo Modelado.** Muestra el nombre y el tipo del nodo de modelado utilizado para generar o actualizar el modelo.

Presentación preliminar de descripciones de rutas

Puede ver el contenido de una descripción de ruta en un explorador Web pulsando en una opción del cuadro de diálogo Propiedades de ruta. El contenido de la descripción depende de las opciones que especifique en la pestaña Despliegue del cuadro de diálogo. Consulte el tema “Opciones de despliegue de rutas” en la página 133 para obtener más información.

Para ver una descripción de ruta:

1. En el menú principal de IBM SPSS Modeler, pulse en:
Herramientas > Propiedades de ruta > Despliegue
2. Establezca el tipo de despliegue, el nodo de puntuación designada y los parámetros de puntuación.
3. Si el tipo de despliegue es Actualización de modelos, puede seleccionar de forma opcional un:
 - Modo de modelado y cualquier parámetro del generador de modelos
 - Nugget de modelo en la rama de puntuación de la ruta
4. Pulse en el botón **Presentación preliminar de descripción de ruta**.

Exportación de descripciones de ruta

Puede exportar el contenido de la descripción de ruta a un archivo HTML.

Para exportar una descripción de ruta:

1. En el menú principal, pulse en:
Archivo > Exportar descripción de ruta
2. Introduzca un nombre para el archivo HTML y pulse en **Guardar**.

Ejecución de rutas

Una vez que se han especificado las opciones de las rutas y se han conectado los nodos necesarios, se puede ejecutar la ruta ejecutando los datos a través de los nodos de la ruta. Hay varias formas de ejecutar una ruta en IBM SPSS Modeler. Tiene la posibilidad de:

- Pulsar en **Ejecutar** en el menú Herramientas.
- Pulsar en uno de los botones **Ejecutar...** de la barra de herramientas. Estos botones permiten ejecutar toda la ruta o tan solamente el nodo terminal seleccionado. Consulte el tema “Barra de herramientas de IBM SPSS Modeler” en la página 19 para obtener más información.
- Ejecute una sola ruta de datos pulsando con el botón derecho en un nodo terminal y seleccionando **Ejecutar** en el menú emergente.

- Ejecute parte de una ruta de datos pulsando con el botón derecho en cualquier nodo no terminal y seleccionando **Ejecutar desde aquí** en el menú emergente. Al realizar esta acción, solamente se realizarán aquellas operaciones después del nodo seleccionado.

Para detener la ejecución de una ruta en curso, se puede pulsar en el botón rojo Detener de la barra de herramientas o seleccionar **Detener ejecución** en el menú Herramientas.

Si la ruta tarda más de tres segundos en ejecutarse, se muestra el cuadro de diálogo Comentarios de la ejecución para indicar el progreso.

Algunos nodos muestran información adicional sobre la ejecución de la ruta. Esta información aparece al seleccionar la fila correspondiente en el cuadro de diálogo. La primera fila se selecciona de forma automática.

Trabajo con modelos

Si una ruta incluye un nodo de modelado (o sea, un nodo de las pestañas Modelado o Modelado de bases de datos de la paleta de nodos), se creará un **nugget de modelo** cuando se ejecute la ruta. Un nugget de modelo es un contenedor para un **modelo**, o sea, un conjunto de reglas, fórmulas o ecuaciones que le permiten generar predicciones con sus datos de origen, y que está en el centro del análisis predictivo.



Figura 12. Nugget de modelo

Cuando se ejecuta correctamente un nodo de modelado, el nugget de modelo correspondiente se coloca en el lienzo de rutas, donde se representa por medio de un icono con forma de diamante dorado (de aquí su nombre). Puede abrir el nugget y explorar su contenido para ver los detalles sobre el modelo. Para ver las predicciones, se adjuntan y se ejecutan uno o más nodos terminales, y cuyo resultado presentará las predicciones de forma legible.

Una ruta de modelado típica consta de dos ramas. La **rama de modelado** contiene el nodo de modelado, junto con los nodos de origen y de procesamiento que la preceden. La **rama de puntuación** se crea cuando se ejecuta el nodo de modelado, y contiene el nugget de modelo y el nodo de terminal o los nodos que se utilizan para ver las predicciones.

Si desea obtener más información, consulte el manual *Nodos de modelado de IBM SPSS Modeler*.

Adición de comentarios y anotaciones a nodos y rutas

Es posible que tenga que describir una ruta para otras personas de su organización. Para ayudarle, puede añadir comentarios explicativos a rutas, nodos y nuggets de modelo.

Otros usuarios pueden visualizar estos comentarios en la pantalla o pueden imprimir la imagen de la ruta que incluya los comentarios.

Puede elaborar una lista con todos los comentarios de una ruta o supernodo, cambiar el orden de los comentarios en la lista, editar el texto de comentarios y cambiar el color de fondo o de máscara de un comentario. Consulte el tema “Lista de comentarios de ruta” en la página 59 para obtener más información.

También puede añadir notas en forma de anotaciones de texto a las rutas, nodos y nuggets mediante la pestaña Anotaciones del cuadro de diálogo de propiedades de ruta, un cuadro de diálogo de nodo o una ventana de nugget de modelo. Estas notas solamente son visibles si la pestaña Anotaciones está abierta,

salvo que las anotaciones de la ruta también se pueden mostrar como comentarios en pantalla. Consulte el tema “Anotaciones” en la página 60 para obtener más información.

Comentarios

Los comentarios toman la forma de cuadros de texto en los que se puede introducir cualquier cantidad de texto, y puede añadir todos los comentarios que desee. Un comentario puede estar libre (sin vinculación a ningún objeto de ruta) o puede estar conectado a uno o más nodos o nuggets de modelo de la ruta. Los comentarios libres se suelen utilizar para describir el propósito general de la ruta, mientras que los comentarios conectados describen el nodo o nugget al que están vinculados. Los nodos y nuggets pueden tener más de un comentario vinculado y la ruta puede tener cualquier número de comentarios libres.

Note: También se pueden mostrar anotaciones como comentarios en pantalla, aunque no se pueden añadir a nodos o nuggets. Consulte el tema “Conversión de anotaciones en comentarios” en la página 59 para obtener más información.

La apariencia del cuadro de texto cambia para indicar el modo actual del comentario (o anotación mostrada como un comentario), como muestra la siguiente tabla.

Tabla 3. Modos de cuadros de texto de comentarios y anotación

Cuadro de texto de comentarios	Cuadro de texto de anotación	Modo	Indica	Obtenido por...
		Editar	El comentario está abierto para su edición.	Creación de un nuevo comentario o anotación o selección de uno existente.
		Última selección	El comentario se puede mover, cambiar su tamaño o eliminarse.	Pulse en el fondo de una ruta tras la edición o pulse una vez en un comentario o anotación existente.
		Ver	Ha finalizado la edición.	Selección de otro nodo, comentario o anotación tras la edición.

Cuando crea un nuevo comentario libre, se mostrará inicialmente en la esquina superior izquierda del lienzo de rutas.

Si vincula un comentario a un nodo o nugget, el comentario se muestra inicialmente por encima del objeto de ruta al que está vinculado.

El cuadro de texto aparece en color blanco mostrando que puede introducir el texto. Cuando haya introducido el texto, pulse fuera del cuadro de texto. El fondo cambiará a amarillo para mostrar que ha completado el comentario. El comentario permanece seleccionado y podrá moverlo, cambiar su tamaño o eliminarlo.

Cuando pulse de nuevo, el borde cambia a líneas sólidas para mostrar que ha finalizado la edición.

Si pulsa dos veces en un comentario, cambiará el cuadro de texto al modo de edición, el fondo cambiará a blanco y podrá editar el texto del comentario.

También puede vincular comentarios en Supernodos.

Operaciones que implican comentarios: Puede realizar diferentes operaciones en los comentarios. Tiene la posibilidad de:

- Añadir un comentario libre

- Vincular un comentario a un nodo o nugget
- Editar un comentario
- Redimensionar un comentario
- Mover un comentario
- Desconectar un comentario
- Eliminar un comentario
- Mostrar u ocultar todos los comentarios de una ruta

Añadir comentario libre

1. Asegúrese que no ha seleccionado ningún elemento en la ruta.
2. Realice una de las siguientes acciones:
 - En el menú principal, pulse en:
 - Insertar > Nuevo comentario**
 - Pulse con el botón derecho en el fondo de la ruta y pulse en **Nuevo comentario** en el menú emergente.
 - Pulse en el botón **Nuevo comentario** de la barra de herramientas.
3. Introduzca el texto de comentario (o pegue el texto desde el portapapeles).
4. Pulse en un nodo de la ruta para guardar el comentario.

Vincular un comentario a un nodo o nugget

1. Seleccione uno o más nodos o nuggets en el lienzo de rutas.
2. Realice una de las siguientes acciones:
 - En el menú principal, pulse en:
 - Insertar > Nuevo comentario**
 - Pulse con el botón derecho en el fondo de la ruta y pulse en **Nuevo comentario** en el menú emergente.
 - Pulse en el botón **Nuevo comentario** de la barra de herramientas.
3. Escriba el texto de comentario.
4. Pulse en otro nodo de la ruta para guardar el comentario.
Si lo desea, puede:
5. Introducir un comentario libre (consulte la sección anterior).
6. Realice una de las siguientes acciones:
 - Seleccione el comentario, pulse F2 y seleccione el nodo o nugget.
 - Seleccione el nodo o nugget, pulse F2 y seleccione el comentario.
 - (Ratón con tres botones únicamente) Mueva el puntero del ratón por el comentario, mantenga pulsado el botón central, arrastre el puntero por el nodo o nugget y suelte el botón del ratón.

Vincular un comentario adicional a un nodo o nugget

Si ya ha vinculado un comentario a un nodo o nugget, o si ya está a nivel de ruta y desea vincularlo a un nodo o nugget adicional, realice una de las siguientes acciones:

- Seleccione el comentario, pulse F2 y seleccione el nodo o nugget.
- Seleccione el nodo o nugget, pulse F2 y seleccione el comentario.
- (Ratón con tres botones únicamente) Mueva el puntero del ratón por el comentario, mantenga pulsado el botón central, arrastre el puntero por el nodo o nugget y suelte el botón del ratón.

Editar un comentario existente

1. Realice una de las siguientes acciones:

- Pulse dos veces en el cuadro de texto del comentario.
 - Seleccione el cuadro de texto y pulse Intro.
 - Pulse con el botón derecho en el cuadro de texto para mostrar su menú; a continuación, pulse en Edición.
2. Edite el texto del comentario. Puede utilizar las teclas de acceso directo estándar de Windows cuando edite; por ejemplo, Ctrl+C para copiar el texto. Se enumerarán otras opciones durante la edición en el menú emergente del comentario.
 3. Pulse fuera del cuadro de texto una vez para ver los controles de redimensionado y pulse de nuevo para completar el comentario.

Redimensionar un cuadro de texto

1. Seleccione el comentario para ver los controles de redimensionado.
2. Pulse y arrastre un control para redimensionar el cuadro.
3. Pulse fuera del cuadro de texto para guardar el cambio.

Mover un comentario existente

Si desea mover un comentario, pero no sus objetos vinculados (si los tiene), realice una de las siguientes acciones:

- Mueva el puntero por encima del comentario, mantenga pulsado el botón izquierdo del ratón y arrastre el comentario a su nueva ubicación.
- Seleccione el comentario, mantenga pulsada la tecla Alt y mueva el comentario con los cursores.

Si desea mover un comentario junto con los nodos o nuggets a los que está vinculado:

1. Seleccione todos los objetos que desee mover.
2. Realice una de las siguientes acciones:
 - Mueva el puntero por encima de uno de los objetos, mantenga pulsado el botón izquierdo del ratón y arrastre los objetos a su nueva ubicación.
 - Seleccione uno de los objetos, mantenga pulsada la tecla Alt y mueva los objetos con los cursores.

Desconectar un comentario a un nodo o nugget

1. Seleccione uno o más comentarios para desconectar.
2. Realice una de las siguientes acciones:
 - Pulse F3.
 - Pulse con el botón derecho en un comentario seleccionado y, a continuación, en Desconectar en su menú.

Eliminar un comentario

1. Seleccione uno o más comentarios para eliminar.
2. Realice una de las siguientes acciones:
 - Pulse la tecla Supr.
 - Pulse con el botón derecho en un comentario seleccionado y, a continuación, en Eliminar en su menú.

Si el comentario se ha añadido a un nodo o nugget, la línea de conexión también se ha eliminado.

Si el comentario era originalmente una anotación de ruta o supernodo que se había convertido en comentario libre, se eliminará el comentario del lienzo pero se retendrá el texto en la pestaña Anotaciones de la ruta o el supernodo.

Para mostrar u ocultar todos los comentarios de una ruta

1. Realice una de las siguientes acciones:

- En el menú principal, pulse en:
Ver > Comentarios
- Pulse en el botón **Mostrar/ocultar comentarios** de la barra de herramientas.

Lista de comentarios de ruta: Puede ver una lista de todos los comentarios realizados en una ruta concreta o supernodo.

En esta lista, podrá

- Cambiar el orden de comentarios
- Editar el texto del comentario
- Cambiar el color de máscara o fondo de un comentario

Lista de comentarios

Para crear una lista de los comentarios de una ruta, realice una de las siguientes acciones:

- En el menú principal, pulse en:
Herramientas > Propiedades de ruta > Comentarios
- Pulse con el botón derecho del ratón en el panel de gestores y en **Propiedades de ruta** y, a continuación, en **Comentarios**.
- Pulse con el botón derecho del ratón en el fondo de un comentario en el lienzo de rutas y seleccione **Propiedades de ruta y Comentarios**.

Texto. El texto del comentario. Pulse dos veces en el texto para cambiar el campo a un cuadro de texto editable.

Enlaces. El nombre del nodo al que se vincula el comentario. Si el campo está vacío, el comentario se aplica a la ruta.

Botones de posicionamiento. Suben o bajan un comentario en la lista.

Colores de comentario. Para cambiar el color de máscara o de fondo de un comentario, seleccione el comentario, seleccione la casilla de verificación **Colores personalizados**, seleccione un color de la lista **Fondo** o **Máscara** (o ambas). Pulse en **Aplicar** y, a continuación, en el fondo de la ruta para ver el efecto del cambio. Pulse en **Aceptar** para guardar el cambio.

Conversión de anotaciones en comentarios: Las anotaciones realizadas en rutas o supernodos se pueden convertir en comentarios.

En el caso de rutas, la anotación se convierte en un comentario libre (es decir, no vinculado a ningún nodo) en el lienzo de rutas.

Si una anotación de supernodo se convierte en un comentario, el comentario no se vincula al supernodo en el lienzo de rutas, sino que será visible cuando haga zoom en el supernodo.

Para convertir una anotación de ruta en un comentario

1. Pulse en **Propiedades de ruta** en el menú Herramientas. (También puede pulsar con el botón derecho del ratón en una ruta del panel de gestores y seleccionar **Propiedades de ruta**.)
2. Pulse en la pestaña **Anotaciones**.
3. Seleccione la casilla de verificación **Mostrar anotación como comentario**.
4. Pulse en **Aceptar**.

Para convertir una anotación de supernodo en un comentario

1. Pulse dos veces en el icono del supernodo en el lienzo de rutas.
2. Pulse en la pestaña **Anotaciones**.
3. Seleccione la casilla de verificación **Mostrar anotación como comentario**.
4. Pulse en **Aceptar**.

Anotaciones

Los nodos, las rutas y los modelos se pueden anotar de distintas formas. Se pueden añadir anotaciones descriptivas y especificar un nombre personalizado. Estas opciones resultan especialmente útiles cuando se generan informes para rutas añadidas al panel de proyectos. En nodos y nuggets de modelo puede añadir texto de información sobre herramientas para ayudar a distinguir nodos que sean parecidos en el lienzo de rutas.

Adición de anotaciones

Cuando se edita un nodo o un nugget de modelo, se abre un cuadro de diálogo con pestañas. Una de ellas es Anotaciones, que permite configurar distintas opciones de anotación. También se puede abrir la pestaña Anotaciones directamente.

1. Para anotar un nodo o un nugget de modelo, pulse con el botón derecho del ratón en el nodo o nugget del lienzo de rutas y seleccione **Cambiar nombre y anotar**. Se abrirá el cuadro de diálogo de edición con la pestaña Anotaciones visible.
2. Para anotar una ruta, pulse en **Propiedades de ruta** en el menú Herramientas. (También puede pulsar con el botón derecho del ratón en una ruta del panel de gestores y seleccionar **Propiedades de ruta**.) Pulse en la pestaña Anotaciones.

Nombre. Seleccione **Personalizado** para ajustar el nombre generado automáticamente o asignar un nombre exclusivo al nodo, que se mostrará en el lienzo de rutas.

Texto de información sobre herramientas. (Para nodos y o un nugget de modelo únicamente) Introduzca el texto utilizado como información sobre herramientas en el lienzo de rutas. Esto resulta particularmente útil cuando se trabaja con muchos nodos parecidos.

Palabras clave. Especifique las palabras clave que desea utilizar en los informes del proyecto y cuando realice búsquedas de nodos en una ruta o seguimientos de objetos almacenados en el repositorio (consulte “Acerca de IBM SPSS Collaboration and Deployment Services Repository” en la página 119). Se pueden especificar varias palabras clave separadas por punto y coma (por ejemplo, ingresos; tipo recorte; valor de reclamación). Los espacios en blanco al comienzo y final de cada palabra clave se eliminan (por ejemplo, ingresos; tipo recorte producirá los mismos resultados que ingresos; tipo recorte). (Los espacios en blanco dentro de las palabras clave no se eliminan. Por ejemplo, tipo recorte, con un espacio, y tipo recorte, con dos espacios, no es lo mismo.)

El área de texto principal se puede utilizar para introducir anotaciones extensas relacionadas con las operaciones del nodo o de las decisiones realizadas en el mismo. Por ejemplo, cuando se comparten o se vuelven a utilizar rutas, resulta útil realizar anotaciones sobre decisiones como puede ser el descarte de un campo con muchos espacios vacíos que utilizan un nodo Filtrar. Al anotar el nodo se almacena esta información con el propio nodo. También se puede optar por incluir estas anotaciones en un informe del proyecto creado a partir del panel de proyectos. Consulte el tema “Introducción a los proyectos” en la página 143 para obtener más información.

Mostrar anotación como comentario. (Para anotaciones de ruta y supernodo únicamente) Seleccione esta casilla de verificación para convertir la anotación en un comentario libre que se verá en el lienzo de rutas. Consulte el tema “Adición de comentarios y anotaciones a nodos y rutas” en la página 55 para obtener más información.

ID. Muestra un ID exclusivo que se puede utilizar para hacer referencia al nodo con fines de creación de scripts o automatización. Este valor se genera automáticamente al crear el nodo y no cambiará. Además,

tenga en cuenta que para evitar confusiones con la letra "O", no se utilizan ceros en los identificadores de los nodos. Utilice el botón Copiar situado a la derecha para copiar y pegar el ID en los scripts o en los lugares que sea necesario.

Almacenamiento de rutas de datos

Una vez creada una ruta, se puede guardar para reutilizarla más adelante.

Para guardar una ruta

1. En el menú Archivo, pulse en **Guardar ruta o Guardar ruta como**.
2. En el cuadro de diálogo Guardar, busque la carpeta en la que desea guardar el archivo de ruta.
3. Introduzca un nombre para la ruta en el cuadro de texto Nombre de archivo.
4. Seleccione **Añadir al proyecto** si desea añadir la ruta guardada en el proyecto actual.

Si pulsa en **Guardar**, la ruta se almacenará con la extensión **.str* en el directorio especificado.

Archivos de copia de seguridad automática. Cada vez que se guarda una ruta, la versión del archivo guardada anteriormente se conserva automáticamente como una copia de seguridad, con un guión añadido al nombre del archivo (por ejemplo, *mystream.str-*). Para restaurar la versión de la copia de seguridad, sólo hay que eliminar el guión y volver a abrir el archivo.

Almacenamiento de estados

Además de rutas, también se pueden guardar **estados**, que incluyen el diagrama de ruta que se muestra y cualquier nugget de modelo que se haya creado (enumerado en la pestaña Modelos del panel de gestores).

Para guardar un estado

1. En el menú Archivo, pulse en:
Estado > Guardar estado o Guardar estado como
2. En el cuadro de diálogo Guardar, busque la carpeta en la que desea guardar el archivo de estado.

Si pulsa en **Guardar** se almacena el estado con la extensión **.cst* en el directorio especificado.

Almacenamiento de nodos

También puede guardar un único nodo si pulsa con el botón derecho en el nodo del lienzo de rutas y selecciona **Guardar nodo** en el menú emergente. Utilice la extensión de archivo **.nod*.

Almacenamiento de varios objetos de ruta

Cuando se elige salir de IBM SPSS Modeler sin haber guardado varios objetos (como rutas, proyectos o nuggets de modelos) se le solicitará que se guarde todo antes de cerrar por completo el software. Si opta por guardar los elementos, aparecerá un cuadro de diálogo con opciones para guardar cada objeto.

1. Seleccione las casillas de verificación de los objetos que desee guardar.
2. Pulse en **Aceptar** para guardar cada objeto en el lugar deseado.

A continuación, aparecerá un cuadro de diálogo Guardar estándar para cada objeto. Una vez que haya terminado de guardar, la aplicación se cerrará como se indicó en un principio.

Almacenamiento de resultados

Ahora puede guardar las tablas, gráficos e informes generados de los nodos de resultados de IBM SPSS Modeler en formato de objetos de resultados (**.cou*).

1. Cuando se muestran los resultados que desea guardar, en los menús de ventana de resultados pulse:
Archivo > Guardar
2. Especifique un nombre y una ubicación para el archivo de resultados.

3. De forma opcional, seleccione **Añadir archivo al proyecto** en el cuadro de diálogo Guardar para incluir el archivo en el proyecto actual. Consulte el tema “Introducción a los proyectos” en la página 143 para obtener más información.

De forma alternativa, puede pulsar con el botón derecho en cualquier objeto de resultados que se enumera en el panel de gestores y seleccionar **Guardar** en el menú emergente.

Cifrado y descifrado de información

Al guardar una ruta, un nodo, un proyecto, un archivo de resultados o un nugget de modelo, puede cifrarlos para evitar su uso no autorizado. Para ello, seleccione una opción adicional al guardar y añada una contraseña al elemento en cuestión. Este cifrado puede definirse para cualquier elemento que desee guardar y ofrece seguridad adicional; no es igual que el cifrado SSL, que se utiliza al pasar archivos de IBM SPSS Modeler a IBM SPSS Modeler Server.

Al intentar abrir un elemento cifrado, se le pedirá que introduzca la contraseña. Tras introducir la contraseña correcta, el elemento se descifrá automáticamente y se abrirá de la forma habitual.

Para cifrar un elemento

1. En el cuadro de diálogo Guardar del elemento cifrado, pulse en **Opciones**. Aparecerá el cuadro de diálogo Opciones de cifrado.
2. Seleccione **Cifrar este archivo**.
3. De manera opcional, para mayor seguridad, seleccione **Enmascarar contraseña**. Al hacerlo, todos los datos introducidos se muestran como una serie de puntos.
4. Introduzca la contraseña. *Advertencia:* si olvida la contraseña, no podrá abrir el archivo o modelo.
5. Si ha seleccionado **Enmascarar contraseña**, vuelva a introducir la contraseña para confirmar que la ha escrito correctamente.
6. Pulse en **Aceptar** para volver al cuadro de diálogo Guardar.

Note: si guarda una copia de un elemento protegido mediante cifrado, el nuevo elemento se guardará automáticamente en formato cifrado y utilizará la contraseña original, a no ser que se modifique la configuración en el cuadro de diálogo Opciones de cifrado.

Carga de archivos

Puede volver a cargar un número de objetos guardados en IBM SPSS Modeler:

- Rutas (.str)
- Estados (.cst)
- Modelos (.gm)
- Paleta de modelos (.gen)
- Nodos (.nod)
- Resultado (.cou)
- Proyectos (.cpj)

Apertura de archivos nuevos

Las rutas se pueden cargar directamente en el menú Archivo.

1. Para ello, en el menú Archivo, pulse en **Abrir ruta**.

El resto de tipos de archivo se puede abrir mediante los elementos del submenú disponibles en el menú Archivo. Por ejemplo, para cargar un modelo, en el menú Archivo pulse:

Modelos > Abrir modelo o Cargar paleta de modelos

Apertura de archivos usados recientemente

Para cargar de forma rápida los archivos utilizados recientemente, se pueden utilizar las opciones que se encuentran en la parte inferior del menú Archivo.

Seleccione **Rutas recientes**, **Proyectos recientes** o **Estados recientes** para desplegar una lista de los archivos utilizados recientemente.

Correlacionar rutas de datos

Mediante la herramienta de correlación se puede conectar un nuevo origen de datos con una ruta ya existente. La herramienta de correlación no solamente establecerá la conexión, sino que además permite especificar el modo en que los campos del nuevo origen van a sustituir a los de la ruta existente. En lugar volver a crear una ruta de datos entera para un nuevo origen de datos, se puede conectar con una ruta existente.

La herramienta de correlación de datos permite unir dos fragmentos de rutas y garantizar que todos los nombres de los campos (esenciales) coincidan correctamente. En resumen, la correlación de datos implica la creación de un nuevo nodo Filtrar que establece la correspondencia entre los campos adecuados cambiándoles el nombre.

Existen dos modos de correlacionar datos:

Seleccionar nodo de sustitución. Este método comienza con el nodo que se va a sustituir. En primer lugar, pulse con el botón derecho en el nodo que va a sustituir; a continuación, utilizando la opción **Correlación de datos > Seleccionar nodo de sustitución** del menú emergente, seleccione el nodo con el que desea sustituirlo.

Correlacionar a. Este método comienza con el nodo que se va a introducir en la ruta. En primer lugar, pulse con el botón derecho en el nodo que va a introducir; a continuación, utilizando la opción **Correlación de datos > Correlacionar a** del menú emergente, seleccione el nodo al que debería unirse. Este método resulta especialmente adecuado para correlacionar a un nodo terminal. *Note:* No se puede correlacionar a nodos Fundir o Añadir. En su lugar, hay que conectar la ruta con el nodo Fundir de la forma habitual.

La correlación de datos está estrechamente integrada en la construcción de rutas. Si intenta conectar con un nodo que ya tiene una conexión, se le ofrecerá la opción de sustituir la conexión o realizar una correlación a ese nodo.

Correlación de datos a una plantilla

Para sustituir el origen de datos de una ruta de plantilla por un nodo de origen nuevo introduciendo los datos del usuario en IBM SPSS Modeler, es necesario utilizar la opción **Seleccionar nodo de sustitución** en el menú emergente Correlación de datos. Esta opción está disponible para todos los nodos a excepción de los nodos terminales y los nodos Fundir y Agregar. Cuando se utiliza la herramienta de correlación de datos para llevar a cabo esta acción, se puede garantizar que se establece la correspondencia de los campos correctamente entre las operaciones de ruta existentes y el nuevo origen de datos. Los siguientes pasos proporcionan conceptos básicos sobre el proceso de correlación de datos.

Paso 1: especificar los campos esenciales en el nodo de origen. Para que las operaciones de ruta se ejecuten correctamente, es necesario especificar campos esenciales. Consulte el tema “Especificación de campos esenciales” en la página 64 para obtener más información.

Paso 2: agregar un nuevo origen de datos al lienzo de rutas. Si se utiliza uno de los nodos de origen, se obtienen los nuevos datos de sustitución.

Paso 3: sustituir el nodo de origen de la plantilla. Mediante la opción Correlación de datos del menú emergente para el nodo de origen de la plantilla, pulse en **Seleccionar nodo de sustitución** y, a continuación, seleccione el nodo de origen para los datos de sustitución.

Paso 4: comprobar los campos correlacionados. En el cuadro de diálogo que se abre, compruebe que el software correlaciona correctamente los campos desde el origen de datos de sustitución a la ruta. Los campos esenciales sin correlacionar aparecen en rojo. Estos campos se utilizan en las operaciones de ruta y se deben sustituir por un campo similar en el nuevo origen de datos para que las operaciones siguientes funcionen correctamente. Consulte el tema “Análisis de campos correlacionados” en la página 65 para obtener más información.

Una vez utilizado el cuadro de diálogo para comprobar que se han correlacionado correctamente todos los campos esenciales, el origen de datos antigua se desconecta de la ruta con un nodo Filtrar denominado *Correlacionar*. Este nodo Filtrar dirige la correlación de campos real en la ruta. También se incluye un nodo Filtrar *Cancelar correlación* en el lienzo de rutas. Si se añade el nodo Filtrar *Cancelar correlación*, se puede utilizar para invertir la correlación de los nombres de los campos. Deshará la correlación entre los campos, pero será necesario editar cualquier nodo terminal que le suceda para volver a seleccionar los campos y las superposiciones.

Correlación entre rutas

Al igual que en la conexión de nodos, este método de correlación de datos no exige que se configuren previamente los campos esenciales. Con este método se establece la conexión de una ruta a otra de forma sencilla utilizando la opción **Correlacionar a** del menú emergente Correlación de datos. Este tipo de correlación de datos es útil para correlacionar a nodos terminales y para realizar operaciones de copiar y pegar entre rutas. *Note:* con la opción **Correlacionar a** no se pueden correlacionar a nodos Fundir, Añadir ni con ningún nodo de origen.

Para correlacionar datos entre rutas

1. Pulse con el botón derecho del ratón en el nodo que desee utilizar para conectarlo con la ruta nueva.
2. En el menú, pulse en:
Correlación de datos > Correlacionar a
3. Utilice el cursor para seleccionar un nodo de destino en la ruta objetivo.
4. En el cuadro de diálogo que se abre, asegúrese de que los campos se corresponden de forma adecuada y pulse en **Aceptar**.

Especificación de campos esenciales

Al correlacionar a una ruta existente, normalmente el autor de la ruta especifica los campos esenciales. Estos campos esenciales indican si un campo determinado se utiliza en las operaciones siguientes de la ruta. Por ejemplo, la ruta actual puede generar otro modelo que utiliza un campo denominado *Churn*. En esta ruta, *Churn* es un campo esencial porque no se podría generar el modelo sin él. Del mismo modo, los campos utilizados en los nodos de manipulación, como es el nodo Derivar, son necesarios para derivar el nuevo campo. La configuración explícita de estos campos como esenciales ayuda a garantizar que éstos se correlacionan a los campos adecuados del nuevo nodo de origen. Si hay campos obligatorios sin correlacionar, aparecerá un mensaje de error. En caso de que considere que determinadas manipulaciones o nodos de resultados no son necesarios, se pueden eliminar los nodos de la ruta y eliminar los campos correspondientes de la lista Campos esenciales.

Para configurar los campos esenciales

1. Pulse con el botón derecho del ratón en el nodo de origen de la ruta de plantilla que va a sustituirse.
2. En el menú, pulse en:
Correlación de datos > Especificar campos esenciales
3. El selector de campos permite añadir o eliminar campos de la lista. Para abrir el selector de campos, pulse en el icono que se encuentra a la derecha de la lista de campos.

Análisis de campos correlacionados

Una vez seleccionado el punto en el que una ruta o origen de datos se va a correlacionar a otra, se muestra un cuadro de diálogo para seleccionar los campos a correlacionar o para asegurarse de que la correlación predeterminada del sistema es correcta. Si los campos esenciales se han configurado para la ruta o para el origen de datos, estos campos se resaltan en rojo. Los campos sin correlacionar del origen de datos pasarán por el nodo Filtrar sin sufrir modificaciones, aunque también se pueden correlacionar campos no esenciales.

Original. Enumera todos los campos de la plantilla o de la ruta existente: todos los campos que se encuentran por debajo. Los campos del nuevo origen de datos se correlacionarán a estos campos.

Correlacionado. Indica los campos seleccionados para establecer la correlación a los campos de plantilla. Puede que sea necesario modificar los nombres de estos campos para que se correspondan con los campos originales utilizados en las operaciones de ruta. Pulse en la casilla de un campo de la tabla para activar la lista de campos disponibles.

Si no está seguro de los campos a correlacionar, puede resultar útil analizar los datos de origen antes de llevar a cabo la correlación. Por ejemplo, se puede utilizar la pestaña Tipos del nodo de origen para obtener un resumen de los datos de origen.

Sugerencias y atajos

Familiarícese con las siguientes sugerencias y atajos para trabajar de forma rápida y sencilla:

- **Genere rutas de forma rápida pulsando dos veces en el ratón.** Pulse dos veces con el ratón en un nodo de la paleta para añadirlo y conectarlo a la ruta actual.
- **Utilice combinaciones de teclas para seleccionar los nodos posteriores en la ruta.** Pulse Ctrl+Q y Ctrl+W para cambiar la selección de todos los nodos por debajo de la ruta.
- **Utilice las teclas de acceso directo para conectar y desconectar nodos.** Cuando se selecciona un nodo en el lienzo, pulse F2 para comenzar una conexión, el tabulador para desplazarse hasta el nodo deseado y Mayús+Barra espaciadora para finalizar la conexión. Pulse F3 para desconectar todas las entradas y salidas del nodo seleccionado.
- **Personalice la pestaña Paleta de nodos con los nodos que utilice con más frecuencia.** En el menú Herramientas pulse en **Administrar paletas** para abrir un cuadro de diálogo desde el que podrá añadir, eliminar o mover los nodos mostrados en la pestaña Paleta de nodos.
- **Cambie el nombre de los nodos y añada información sobre herramientas.** Los cuadros de diálogo de cada nodo incluyen una pestaña Anotaciones en la que se puede especificar un nombre personalizado para los nodos del lienzo y añadir información sobre herramientas para facilitar la organización de la ruta. También se pueden incluir anotaciones largas para realizar un seguimiento del proceso, guardar los detalles del mismo y denotar cualquier decisión empresarial necesaria o tomada previamente.
- **Inserte valores automáticamente en una expresión CLEM.** A través del generador de expresiones, al que se puede acceder desde distintos cuadros de diálogo (como los de los nodos Derivar y Filtro), se pueden insertar automáticamente valores de campo en una expresión CLEM. Pulse en el botón de valores del generador de expresiones para seleccionar uno de los valores de campo existentes.



Figura 13. Botón de valores

- **Busque archivos con rapidez.** Cuando se buscan archivos en un cuadro de diálogo Abrir, utilice la lista Archivo (pulse en el botón de diamante amarillo en la parte superior del cuadro de diálogo, junto al campo Buscar en) para acceder a los directorios utilizados anteriormente así como los directorios por omisión de IBM SPSS Modeler. Utilice los botones Adelante y Atrás para desplazarse por los directorios a los que se ha accedido.

- **Minimice las ventanas de resultados abarrotadas.** Se puede cerrar o eliminar la ventana de resultados muy rápidamente con el botón X rojo que se encuentra en la esquina superior derecha de todas las ventanas de resultados. Esto permite conservar tan solamente los resultados útiles o interesantes en la pestaña Resultados del panel de gestores.

Está disponible una gama completa de atajos de teclado para el software. Consulte el tema “Accesibilidad desde el teclado” en la página 168 para obtener más información.

¿Sabía que puede...?

- Arrastrar y seleccionar un grupo de nodos del lienzo de rutas con el ratón.
- Copiar y pegar nodos de una ruta a otra.
- Acceder a la Ayuda desde todos los cuadros de diálogo y ventanas de resultados.
- Obtener ayuda sobre CRISP-DM (Cross-Industry Standard Process for Data Mining). (En el menú Ayuda, pulse en **Ayuda de CRISP-DM.**)

Capítulo 6. Cómo gestionar valores perdidos

Conceptos básicos de valores perdidos

Durante la fase Preparación de los datos de la minería de datos, con frecuencia deseará sustituir los valores perdidos de los datos. Los **valores perdidos** son valores del conjunto de datos desconocidos, sin recopilar o incorrectamente introducidos. Por lo general, estos valores no son válidos en sus campos. Por ejemplo, el campo *Sexo* debe contener los valores *M* y *F*. Si descubre los valores *Y* o *Z* en el campo, puede asumir con seguridad que esos valores no son válidos y que se deben interpretar por lo tanto como espacios en blanco. and should therefore be interpreted as blanks. Del mismo modo, un valor negativo para el campo *Edad* no tendría sentido y, por tanto, también debería interpretarse como un valor vacío. En muchas ocasiones, estos valores obviamente erróneos se han introducido deliberadamente o se han dejado los campos vacíos durante un cuestionario para indicar la omisión de una respuesta. En ocasiones deseará examinar estos elementos vacíos con mayor detenimiento para determinar si una respuesta omitida, como la negativa a proporcionar la edad de una persona, es un factor para predecir un resultado específico.

Algunas técnicas de modelado gestionan ciertos datos perdidos mejor que otros. Por ejemplo, C5.0 y Apriori gestionan correctamente los valores declarados explícitamente "perdidos" en un nodo Tipo. Otras técnicas de modelado presentan problemas al manipular valores perdidos y precisan tiempos de entrenamiento más largos, por lo que se obtienen modelos menos precisos.

Hay varios tipos de valores perdidos reconocidos por IBM SPSS Modeler:

- **Valores nulos o perdidos por el sistema.** Son valores que no son de cadena y que se han dejado en blanco en la base de datos o el archivo de origen y que no se han definido específicamente como "perdidos" en un nodo Tipo o de origen. Los valores perdidos del sistema se muestran como **\$null\$**. Tenga en cuenta que las cadenas vacías no se consideran nulas en IBM SPSS Modeler, aunque se pueden tratar como tales en algunas bases de datos.
- **Cadenas vacías y espacios en blanco.** Los valores de cadenas vacías y los espacios en blanco (cadenas con caracteres no visibles) se tratan como distintos de los valores nulos. Las cadenas vacías se tratan como equivalentes al espacio en blanco en la mayoría de los casos. Por ejemplo, si selecciona la opción de tratar los espacios en blanco como vacíos en un nodo Tipo o de origen, esta configuración se aplica a las cadenas vacías también.
- **Valores vacíos o perdidos definidos por el usuario.** Estos valores, como desconocido, 99, o -1, que se definen explícitamente como que faltan en el nodo Tipo o de origen. Si lo prefiere, también puede elegir tratar los valores nulos o espacios en blanco como vacíos, lo que permite que se les marque para un tratamiento especial y que se excluyan de la mayoría de los cálculos. Por ejemplo, puede utilizar la función @BLANK para tratar estos valores junto con otros tipos de valores perdidos, como vacíos.

Lectura de datos mezclados. Tenga en cuenta que al leer los campos con almacenamiento numérico (entero, real, hora, marca de tiempo o fecha), cualquier valor no numérico se establece como *nulo* o *perdido por el sistema*. Esto ocurre porque, a diferencia de otras aplicaciones, no permite la combinación de tipos de almacenamiento dentro de un campo. Para evitarlo, cualquier campo con datos mezclados debe leerse como cadenas, ya sea cambiando el tipo de almacenamiento en el nodo de origen o en la aplicación externa, según sea necesario.

Lectura de cadenas vacías desde Oracle. Al leer de o escribir en una base de datos de Oracle, debe tener en cuenta que, a diferencia de IBM SPSS Modeler y la mayoría del resto de base de datos, Oracle trata y almacena valores de cadenas vacíos tanto como valores equivalentes hasta nulos. Esto significa que los mismos datos extraídos desde una base de datos de Oracle pueden comportarse de manera diferente a cuando se extraen desde otra base de datos o de un archivo, y pueden devolver resultados diferentes.

Cómo gestionar valores perdidos

En lo que respecta a los conocimientos empresariales o de dominio, resultaría conveniente decidir cómo desea tratar los valores perdidos. Para reducir el tiempo de formación y aumentar la precisión, posiblemente deseará eliminar los espacios vacíos del conjunto de datos. Por otro lado, la presencia de valores vacíos puede aportar una mayor comprensión y nuevas oportunidades empresariales. A la hora de elegir la mejor técnica, debe tener en cuenta los siguientes aspectos de los datos:

- Tamaño del conjunto de datos
- Número de campos que contienen espacios vacíos
- Cantidad de información perdida

Por lo general, se pueden seguir dos métodos:

- Excluir los campos o registros con valores perdidos
- Imputar, sustituir o forzar los valores perdidos mediante varios métodos

Ambos métodos se pueden automatizar por completo mediante el nodo Auditoría de datos. Por ejemplo, puede generar un nodo Filtrar que excluya los campos con demasiados valores perdidos para que sea útil para el modelado, y generar un Supernodo que impute los valores perdidos para uno de los campos que permanecen, o para todos ellos. Aquí es donde se demuestra la potencia real de la auditoría, que permite no sólo evaluar el estado actual de los datos, sino también realizar acciones en función de la evaluación.

Gestión de registros con valores perdidos

Si la mayoría de los valores perdidos se concentra en un pequeño número de registros, puede excluir esos registros. Por ejemplo, los bancos suelen conservar registros completos y detallados de sus clientes con préstamo. No obstante, si el banco siguiera una política menos restrictiva en cuanto a concesiones de préstamos para su propio personal, probablemente los datos recopilados para estos préstamos del personal contarían con varios campos en blanco. En este caso, podría recurrir a dos opciones para gestionar estos valores perdidos:

- Puede utilizar un nodo Seleccionar para eliminar los registros del personal.
- Si el conjunto de datos es grande, puede descartar todos los registros con espacios vacíos.

Gestión de campos con valores perdidos

Si la mayoría de valores perdidos se concentra en un número pequeño de campo, puede dirigirlos al nivel de campo en lugar de al nivel de registro. Este método también permite experimentar con la importancia relativa de campos determinados, antes de decidir un método de gestión de valores perdidos. Si en el modelado un campo resulta sin importancia, probablemente no interese conservarlo, independientemente de la cantidad de valores perdidos que tenga.

Por ejemplo, una empresa de investigación de mercado puede recopilar datos de un cuestionario general que contiene 50 preguntas. Dos de estas preguntas se refieren a persuasión política y edad, información que muchas personas son reticentes a aportar. En este caso, es posible que *Age* y *Political_persuasion* cuenten con varios valores perdidos.

Nivel de medición de campos

Cuando considere qué método le conviene utilizar, no debería olvidar el nivel de medición de campos con valores perdidos.

Campos numéricos. Para los tipos de campos numéricos, como los *Continuos*, siempre debería eliminar todos los valores no numéricos antes de generar un modelo, ya que muchos modelos no funcionarían si hay algún espacio vacío incluido en los campos numéricos.

Campos categóricos. Para campos categóricos, como *Nominal* y *Marca*, no es necesaria la modificación de los valores perdidos, pero aumenta la precisión del modelo. Por ejemplo, un modelo que utiliza el campo *Sexo* seguirá en funcionamiento con valores sin sentido, como *Y* y *Z*. Sin embargo, si se eliminan todos los valores distintos de *M* y *F*, se aumentará la precisión del modelo.

Cribado o eliminación de campos

Para filtrar los campos con demasiados valores perdidos, tiene varias opciones:

- Puede utilizar el nodo Auditoría de datos para filtrar los campos según la calidad.
- Puede utilizar un nodo Selección de características para filtrar campos con más de un porcentaje especificado de valores perdidos y para clasificar campos según la importancia relativa para un objetivo especificado.
- En lugar de eliminar los campos, puede utilizar un nodo Tipo para definir el rol de los campos en **Ninguna**. De esta forma mantendrá los campos de la base de datos pero los excluirá de los procesos de modelado.

Imputación o relleno de valores perdidos

En casos en los que únicamente hay algunos valores perdidos, puede resultar útil insertar valores para sustituir los espacios vacíos. Puede hacer esto desde el informe de auditoría de datos, que permite especificar las opciones que considere oportunas para campos específicos y, a continuación, genera un Supernodo que imputa valores mediante varios métodos. Este método es el más flexible y, además, permite especificar la gestión de un gran número de campos en un solo nodo.

Los siguientes métodos están disponibles para introducir valores perdidos:

Fija. Sustituye un valor fijo (ya sea la media de campo, el punto medio del rango o un constante que especifique).

Aleatorios. Sustituye un valor aleatorio basado en una distribución uniforme o normal.

Expresión. Permite especificar una expresión personalizada. Por ejemplo, podría sustituir los valores con una variable global creada por el nodo Val. globales.

Algoritmo. Sustituye un valor predicho por un modelo basado en el algoritmo C&RT. En cada campo imputado con este método, habrá un modelo C&RT independiente, junto con un nodo Rellenar que sustituye valores vacíos y nulos con el valor que predice el modelo. A continuación, se utiliza un nodo Filtrar para eliminar los campos de predicción generados por el modelo.

Si lo prefiere, para forzar valores para campos específicos, puede utilizar un nodo Tipo para garantizar que los tipos de campo únicamente incluyen valores legales y, a continuación, establezca la columna *Comprobar* en **Forzar** para los campos en los que necesite sustituir los espacios vacíos.

Funciones CLEM para valores perdidos

Existen varias funciones para gestionar valores perdidos. Las siguientes funciones suelen utilizarse en los nodos Seleccionar y Rellenar para descartar o rellenar valores perdidos:

- `count_nulls(LISTA)`
- `@BLANK(CAMPO)`
- `@NULL(CAMPO)`
- `undef`

Las funciones @ pueden utilizarse conjuntamente con la función @FIELD para identificar la presencia de valores nulos o vacíos en uno o varios campos. Los campos pueden marcarse sencillamente cuando contengan valores nulos o vacíos, o bien, pueden rellenarse con valores de sustitución o utilizarse en distintas operaciones.

Puede contar nulos en una lista de campos, de la siguiente manera:

```
count_nulls(['cardtenure' 'card2tenure' 'card3tenure'])
```

Al utilizar cualquiera de las funciones que aceptan una lista de campos como entrada, las funciones especiales @FIELDS_BETWEEN y @FIELDS_MATCHING se pueden utilizar, como se muestra en el siguiente ejemplo:

```
count_nulls(@FIELDS_MATCHING('tarjeta*'))
```

Puede utilizar la función undef para rellenar campos con el valor perdido del sistema, que se muestra como \$null\$. Por ejemplo, podría utilizar una instrucción condicional para sustituir cualquier valor numérico, como en:

```
if not(Edad > 17) or not(Edad > 66) then undef else Edad endif
```

Así podrá sustituir cualquier elemento no comprendido en el rango con un valor perdido del sistema, que se muestra como \$null\$. Mediante la función not(), puede representar todos los valores numéricos restantes, incluidos los negativos. Consulte el tema “Funciones para gestionar los valores vacíos y nulos” en la página 115 para obtener más información.

Nota acerca del descarte de registros

Tenga en cuenta que, al utilizar el nodo Seleccionar para descartar registros, la sintaxis utiliza una lógica de tres valores e incluye automáticamente los valores nulos en las instrucciones de selección. Para excluir valores nulos (perdidos por el sistema) de una expresión de selección, deberá especificarlo expresamente, utilizando and not en la expresión. Por ejemplo, para seleccionar e incluir todos los registros donde el tipo de medicamento prescrito es Drug C, deberá utilizar la siguiente instrucción de selección:

```
Drug = 'drugC' and not(@NULL(Drug))
```

En las versiones anteriores, en este tipo de situaciones se excluían los valores nulos.

Capítulo 7. Generación de expresiones CLEM

Acerca de CLEM

El lenguaje para la manipulación de expresiones de control (CLEM) es un lenguaje para analizar y manipular los datos que fluyen en las rutas de IBM SPSS Modeler. Los analistas de datos suelen utilizar CLEM en las operaciones de rutas para realizar tareas tan simples como derivar beneficios de datos de costes e ingresos, o tan complejas como transformar datos del registro Web en un conjunto de campos y registros con información útil.

CLEM se utiliza en IBM SPSS Modeler para:

- Comparar y evaluar condiciones en campos de registros.
- Derivar valores para campos nuevos.
- Derivar valores nuevos para campos existentes.
- Especular sobre la secuencia de registros.
- Insertar datos de registros en informes.

Las expresiones CLEM son indispensables para la preparación de datos en IBM SPSS Modeler y se puede utilizar en una amplia gama de nodos desde operaciones con campo y registros (Seleccionar, Equilibrar, Rellenar) a gráficos y resultados (Análisis, Informe, Tabla). Por ejemplo, puede utilizar CLEM en un nodo Derivar para crear un nuevo campo según una fórmula como, por ejemplo, cociente.

Las expresiones CLEM también se pueden utilizar para operaciones globales de búsqueda y sustitución. Por ejemplo, la expresión @NULL(@FIELD) se puede utilizar en un nodo de relleno para sustituir los **valores perdidos por el sistema** con el valor entero 0. (Para sustituir **valores perdidos por el usuario**, también llamados valores vacíos, utilice la función @BLANK.)

También se pueden crear expresiones CLEM más complejas. Por ejemplo, puede derivar nuevos campos basándose en un conjunto de reglas condicionales, como una nueva categoría de valores creada mediante estas expresiones: If: CardID = @OFFSET(CardID,1), Then: @OFFSET(ValueCategory,1), Else: 'exclude'.

Este ejemplo utiliza la función @OFFSET para decir, "Si el valor del campo *CardID* para un determinado registro es el mismo que para el registro anterior, devolver el valor del campo denominado *ValueCategory* para el registro anterior. De lo contrario, asignar la cadena "exclude". Es decir, si los identificadores *CardID* para registros adyacentes son iguales, se deberían asignar a la misma categoría de valores. (Los registros con la cadena exclude pueden seleccionarse posteriormente mediante un nodo Seleccionar.)

CLEMEjemplos

Para ilustrar la sintaxis correcta además de los tipos de expresiones posibles con CLEM, se muestran a continuación algunas expresiones de ejemplo.

Expresiones simples

Las fórmulas pueden ser tan simples como ésta, que deriva de un nuevo campo basado en los valores de los campos *Después de* y *Antes de*:

(Después de - Antes de) / Antes de * 100,0

Tenga en cuenta que los nombres de campos no se entrecorillan cuando hacen referencia a los valores de los mismos.

De igual modo, la siguiente expresión devuelve simplemente el logaritmo de cada valor para el campo *salario*.

```
log(salario)
```

Expresiones complejas

Las expresiones también pueden ser más largas y complejas. La siguiente expresión devuelve *verdadero* si el valor de dos campos (*\$KX-Kohonen* y *\$KY-Kohonen*) se encuentra dentro de los rangos especificados. Tenga en cuenta que aquí los nombres de campos tienen comillas simples debido a que contienen caracteres especiales.

```
(' $KX-Kohonen' >= -0.2635771036148072 and ' $KX-Kohonen' <= 0.3146203637123107  
and ' $KY-Kohonen' >= -0.18975617885589602 and  
' $KY-Kohonen' <= 0.17674794197082522) -> T
```

Algunas funciones, como las de cadenas, requieren que el usuario introduzca varios parámetros utilizando una sintaxis correcta. En el ejemplo siguiente, la función *subscrs* se utiliza para devolver el primer carácter de un campo *producir_ID*, indicando si un elemento es orgánico, modificado genéticamente o convencional. Los resultados de una expresión se describen con *-> "Resultado"*.

```
subscrs(1,producir_ID) -> `c`
```

De igual modo, la siguiente expresión es:

```
stripchar(`3`,`123`) -> `12`
```

Se debe tener en cuenta que los caracteres siempre se encapsulan dentro de comillas inversas simples.

Combinación de funciones en una expresión

Con frecuencia, las expresiones CLEM consisten en una combinación de funciones. La función siguiente combina *subscr* y *lowertoupper* para devolver el primer carácter de *producir_ID* y convertirlo en mayúscula.

```
lowertoupper(subscr(1,producir_ID)) -> `C`
```

Esta misma expresión se puede escribir de forma abreviada como:

```
lowertoupper(producir_ID(1)) -> `C`
```

Otra combinación de funciones utilizada normalmente es:

```
locchar_back(`n`, (length(página_web)), página_web)
```

Esta expresión coloca el carácter *`n`* dentro de los valores del campo *página_web* leyendo hacia atrás desde el último carácter del valor del campo. Incluyendo también la función *length*, la expresión calculará dinámicamente la longitud del valor actual en lugar de utilizar un número estático como 7, que no será válido para los valores con menos de siete caracteres.

Funciones especiales

Hay numerosas funciones especiales disponibles (precedidas por un símbolo @). Algunas de las más utilizadas son:

```
@BLANK('ID de referencia') -> T
```

Las funciones especiales se utilizan con frecuencia en combinación, un método muy común de marcar elementos vacíos en más de un campo al mismo tiempo.

```
@BLANK(@FIELD)-> T
```

En la documentación de CLEM se ofrecen más ejemplos. Consulte el tema “Conceptos básicos de la referencia de CLEM” en la página 85 para obtener más información.

Valores y tipos de datos

Las expresiones CLEM son similares a las fórmulas construidas a partir de valores, nombres de campos, operadores y funciones. La expresión CLEM válida más sencilla es un valor o nombre de campo. Algunos ejemplos de valores válidos son:

```
3
1.79
'plátano'
```

Algunos ejemplos de nombres de campos son:

```
ID_Producto
'$P-NextField'
```

donde *Producto* es el nombre de un campo de un conjunto de datos de la cesta de la compra, '*\$P-NextField*' es el nombre de un parámetro, y el valor de la expresión es el valor del campo correspondiente. Normalmente los nombres de campos empiezan por una letra y pueden contener dígitos y signos de subrayado (_). Pero se pueden utilizar nombres que no cumplan estas reglas si se delimitan con comillas. Los valores de CLEM pueden ser:

- Cadenas: por ejemplo, "c1", "Tipo 2", "un texto libre"
- Enteros: por ejemplo, 12, 0, -189
- Números reales: por ejemplo, 12,34, 0,0, -0,0045
- Campos de fecha/hora: por ejemplo, 05/12/2002, 12/05/2002, 12/05/02

También se pueden utilizar los siguientes elementos:

- Códigos de caracteres: por ejemplo, `a` o 3
- Listas de elementos: por ejemplo, [1 2 3], ['Tipo 1' 'Tipo 2']

Las listas y los códigos de caracteres no suelen aparecer como valores de campos. Normalmente se utilizan como argumentos de funciones de CLEM.

Reglas de entrecomillado

Aunque el software es flexible a la hora de determinar los campos, los valores, los parámetros y las cadenas que se utilizan en una expresión CLEM, las siguientes reglas generales ofrecen una lista de "procedimientos recomendables" que seguir cuando se creen expresiones:

- **Cadenas:** utilice siempre comillas dobles cuando escriba cadenas ("Tipo 2" o "valor"). También se pueden utilizar las comillas simples, pero con el riesgo de crear confusión con los campos entrecomillados.
- **Caracteres:** utilice siempre comillas inversas simples como esta ``. Por ejemplo, observe el carácter d en la siguiente función `stripchar(`d`,`drugA`)`. La única excepción se da cuando se utiliza un entero para hacer referencia a un carácter específico de una cadena. Por ejemplo, observe el carácter 5 en la función `lowertoupper("medicamento"(5))` → "A". *Note:* En un teclado estándar de R.U. o EE.UU., la tecla para el carácter de comilla inversa (acento grave, Unicode 0060) puede estar justo debajo de la tecla Esc.
- **Campos:** los campos no suelen entrecomillarse cuando se utilizan en expresiones CLEM (`subscr(2,IDmatriz)`) → CAR). Pero se pueden utilizar comillas simples, cuando sea necesario, para poner espacios u otros caracteres especiales ('Número de orden'). los campos entrecomillados pero no definidos en el conjunto de datos se confundirán con cadenas.
- **Parámetros:** utilice siempre comillas simples ('\$P-umbra1').

Expresiones y condiciones

Las expresiones CLEM pueden devolverse como resultados (al derivar nuevos valores); por ejemplo:

```
Weight * 2.2
Age + 1
sqrt(Signal-Echo)
```

O pueden evaluar como *verdadero* o *falso* (al seleccionar en función de una condición); por ejemplo:

```
Drug = "medicamentoA"
Age < 16
not(PowerFlux) and Power > 2000
```

Puede combinar operadores y funciones arbitrariamente en las expresiones CLEM; por ejemplo:

```
sqrt(abs(Signal)) * max(T1, T2) + Baseline
```

La prioridad del operador y los paréntesis determinan el orden en el que se evalúa la expresión. En este ejemplo, el orden de evaluación es el siguiente:

- Se evalúa `abs(Signal)` y `sqrt` se aplica a su resultado.
- Se evalúa `max(T1, T2)`.
- Los dos resultados se multiplican: `*` tiene mayor prioridad que `+`.
- Por último, `Baseline` se añade al resultado.

El orden descendente de la prioridad (es decir, de las operaciones que se realizan primero a las que se realizan en último lugar) es el siguiente:

- Argumentos de funciones
- Llamadas de funciones
- **xx**
- **x / mod div rem**
- + -
- > < >= <= /== == = /=

Si desea omitir la prioridad, o si tiene alguna duda sobre el orden de evaluación, puede utilizar paréntesis para hacerlo más explícito; por ejemplo,

```
sqrt(abs(Signal)) * (max(T1, T2) + Baseline)
```

Parámetros de ruta, sesión y Supernodo

Se pueden definir los parámetros para utilizarlos en scripts y expresiones CLEM. Son, de hecho, variables definidas por el usuario que se guardan y conservan con la ruta actual, sesión, o Supernodo, y a los que se puede acceder tanto desde la interfaz de usuario como a través de scripts. Si, por ejemplo, se guarda una ruta, cualquier conjunto de parámetros para esa ruta también se guarda. (Así se distinguen de las variables de script local, que sólo se pueden utilizar en el script en que se declaran.) Con frecuencia los parámetros se utilizan en los scripts para controlar el comportamiento del script, proporcionando información sobre los campos y valores que no necesitan estar codificados internamente en el script.

El ámbito de un parámetro depende de dónde se establezca:

- los parámetros de ruta se pueden establecer en un script de ruta o en el cuadro de diálogo de propiedades de ruta, y están disponibles para todos los nodos de la ruta. Se muestran en la lista Parámetros del generador de expresiones.
- Los parámetros de sesión se pueden establecer en un script autónomo o en el cuadro de diálogo de parámetros de sesión. Están disponibles para todas las rutas utilizadas en la sesión actual (todas las rutas enumeradas en la pestaña Rutas del panel de gestores).

También se pueden configurar los parámetros para Supernodos, donde sólo estarán visibles para los nodos encapsulados dentro del Supernodo.

Uso de parámetros en expresiones CLEM

Los parámetros se representan en las expresiones CLEM como \$P-nombrep, donde nombrep es el nombre del parámetro. Cuando se utilizan en expresiones CLEM, los parámetros se deben poner entre comillas simples; por ejemplo, '\$P-escala'.

Los parámetros disponibles se pueden ver fácilmente mediante el generador de expresiones. Para ver los parámetros actuales:

1. En cualquier cuadro de diálogo que acepte expresiones CLEM, pulse en el botón Generador de expresiones.
2. En la lista Campos, seleccione **Parámetros**.

Puede seleccionar parámetros de la lista para insertarlos en la expresión CLEM. Consulte el tema “Selección de campos, parámetros y variables globales” en la página 80 para obtener más información.

Cómo trabajar con cadenas

Se pueden realizar varias operaciones con las cadenas, como:

- Convertir una cadena a mayúsculas o minúsculas: `uppertolower(CAR)`.
- Eliminar caracteres especificados, como ``ID_`` o ``$``, de una variable de cadena: `stripchar(CAR,CADENA)`.
- Determinar la longitud (número de caracteres) de una variable de cadena: `length(CADENA)`.
- Comprobar el orden alfabético de valores de cadena: `alphabefore(CADENA1, CADENA2)`.
- Eliminar los espacios en blanco iniciales o finales de los valores: `trim(CADENA)`, `trim_start(CADENA)` o `trimend(CADENA)`.
- Extraer los primeros o últimos *n* caracteres de una cadena: `startstring(LONGITUD, CADENA)` o `endstring(LONGITUD, CADENA)`. Por ejemplo, supongamos que tiene un campo denominado *elemento* que combina un nombre de producto con un código de ID de cuatro dígitos (ACME CAMERA-D109). Para crear un nuevo campo que contiene únicamente el código de cuatro dígitos, especifique la siguiente fórmula en un nodo Derivar:
`endstring(4, elemento)`
- Coincidencia de un patrón específico: `CADENA matches PATRÓN`. Por ejemplo, para seleccionar personas con "mercado" en cualquier parte de su cargo, puede especificar los siguientes elementos en un nodo Seleccionar:
`puesto_trabajo matches "*mercado*"`
- Sustituir todas las instancias de una subcadena dentro de una cadena: `replace(SUBCADENA, NUEVASUBCADENA, CADENA)`. Por ejemplo, para sustituir todas las instancias de un carácter no admitido, como una barra vertical (|), con un punto y coma antes de la minería de texto, utilice la función `replace` en un nodo Rellenar. En **Rellenar campos**, seleccione todos los campos donde pueda estar el carácter. En la condición **Reemplazar**, seleccione **Siempre** y especifique las siguientes condiciones en **Reemplazar con**:
`replace('|',';',@FIELD)`
- Derivación de un campo de marca en función de la presencia de una subcadena específica. Por ejemplo, puede utilizar una función de cadena en un nodo Derivar para generar un campo de marcas independiente para cada respuesta con una expresión como:
`hassubstring(museums,"museo_de_diseño")`

Consulte el tema “Funciones de cadena” en la página 100 para obtener más información.

Gestión de elementos vacíos y valores perdidos

La sustitución de elementos vacíos o valores perdidos es una tarea de preparación de datos común para los analistas de datos. CLEM ofrece varias herramientas para automatizar la gestión de elementos vacíos. El nodo Rellenar es el lugar más común para trabajar con elementos vacíos; sin embargo, las siguientes funciones se pueden utilizar en cualquier nodo que acepte expresiones CLEM.

- @BLANK(CAMPO) se puede utilizar para determinar registros cuyos valores están vacíos en un campo en particular como, por ejemplo, *Age* (edad).
- @NULL(CAMPO) se puede utilizar para determinar registros cuyos valores ha perdido el sistema para el campo o campos especificados. En IBM SPSS Modeler, los valores perdidos por el sistema se muestran como valores \$null\$.

Consulte el tema “Funciones para gestionar los valores vacíos y nulos” en la página 115 para obtener más información.

Cómo trabajar con números

En IBM SPSS Modeler hay numerosas operaciones estándar con valores numéricos disponibles como, por ejemplo:

- Calcular el seno del ángulo especificado: $\sin(\text{NÚM})$
- Calcular el logaritmo natural de campos numéricos: $\log(\text{NÚM})$
- Calcular la suma de dos números: $\text{NÚM1} + \text{NÚM2}$

Consulte el tema “Funciones numéricas” en la página 97 para obtener más información.

Cómo trabajar con fechas y horas

Los formatos de fecha y hora pueden variar dependiendo de la configuración regional y el origen de datos. Los formatos de fecha y hora son específicos de cada ruta y se definen en el cuadro de diálogo de propiedades de la ruta. Los siguientes ejemplos son funciones utilizadas normalmente para trabajar con campos de fecha/hora.

Cálculo del tiempo transcurrido

Puede calcular fácilmente el tiempo transcurrido desde una fecha de línea base utilizando un grupo de funciones similares a la mostrada a continuación. Esta función devuelve el tiempo en meses desde la fecha de línea base hasta la fecha representada por la cadena de fecha FECHA como un número real. Esta cifra es aproximada; se basa en un mes de 30,0 días.

`date_in_months(Fecha)`

Comparación de valores de fecha/hora

Los valores de campos de fecha/hora se pueden comparar en los registros utilizando funciones similares a la mostrada a continuación. Esta función devuelve un valor de *verdadero* si la cadena de fecha FECHA1 representa una fecha anterior a la representada por la cadena de fecha FECHA2. En el resto de los casos, esta función devuelve 0.

`date_before(Fecha1, Fecha2)`

Cálculo de diferencias

También puede calcular la diferencia entre dos horas y dos fechas utilizando funciones tales como:

`date_weeks_difference(Fecha1, Fecha2)`

Esta función devuelve el tiempo en semanas desde la fecha representada por la cadena de fecha FECHA1 hasta la fecha representada por la cadena de fecha FECHA2 como un número real. Éste se basa en una semana de 7,0 días. Si FECHA2 es anterior a FECHA1, esta función devuelve un número negativo.

Fecha actual

La fecha actual se puede añadir al conjunto de datos utilizando la función @TODAY. La fecha actual se añade como una cadena al campo especificado o a un campo nuevo utilizando el formato de fecha seleccionado en el cuadro de diálogo de propiedades de la ruta. Consulte el tema “Funciones de fecha y hora” en la página 105 para obtener más información.

Resumen de varios campos

El lenguaje CLEM incluye un número de funciones que devuelven estadísticos de resumen de varios campos. Estas funciones pueden ser especialmente útiles al analizar datos de encuestas, donde se pueden almacenar varias respuestas a una pregunta en varios campos. Consulte el tema “Trabajo con datos de respuestas múltiples” en la página 78 para obtener más información.

Funciones de comparación

Se pueden comparar los valores de varios campos mediante las funciones `min_n` y `max_n`, por ejemplo:

```
max_n(['card1fee' 'card2fee' 'card3fee' 'card4fee'])
```

También se pueden utilizar varias funciones de recuento para obtener recuentos de valores que cumplen criterios específicos, incluso cuando esos valores se almacenan en varios campos. Por ejemplo, el recuento del número de tarjetas que se han guardado durante más de cinco años:

```
count_greater_than(5, ['cardtenure' 'card2tenure' 'card3tenure'])
```

Para contar los valores nulos en el mismo conjunto de campos:

```
count_nulls(['cardtenure' 'card2tenure' 'card3tenure'])
```

Recuerde que este ejemplo cuenta el número de cartas que se guardan, no el número de personas que las guardan. Consulte el tema “Funciones de comparación” en la página 93 para obtener más información.

Para contar el número de veces que un valor concreto se repite en varios campos, puede utilizar la función `count_equal`. El siguiente ejemplo cuenta el número de campos de la lista que contienen el valor Y.

```
count_equal("Y", [respuesta1, respuesta2, respuesta3])
```

Teniendo en cuenta los siguientes valores de los campos de la lista, la función devuelve los resultados del valor Y, tal y como se muestra.

Tabla 4. Valores de función.

Respuesta1	Respuesta2	Respuesta3	Recuento
Y	N	Y	2
Y	N	N	1

Funciones numéricas

Se pueden obtener estadísticas de varios campos mediante las funciones `sum_n`, `mean_n` y `sdev_n`, por ejemplo:

```
sum_n(['card1ba1' 'card2ba1' 'card3ba1'])
```

```
mean_n(['card1ba1' 'card2ba1' 'card3ba1'])
```

Consulte el tema “Funciones numéricas” en la página 97 para obtener más información.

Generación de listas de campos

Al utilizar cualquiera de las funciones que aceptan una lista de campos como entrada, las funciones especiales @FIELDS_BETWEEN(inicio, fin) y @FIELDS_MATCHING(patcón) se pueden utilizar como entrada. Por ejemplo, suponiendo que el orden de los campos es el que se muestra en el ejemplo sum_n anterior, lo siguiente sería equivalente:

```
sum_n(@FIELDS_BETWEEN(card1ba1, card3ba1))
```

También, para contar el número de valores nulos en todos los campos que empiezan con "tarjeta":

```
count_nulls(@FIELDS_MATCHING('tarjeta*'))
```

Consulte el tema “Campos especiales” en la página 115 para obtener más información.

Trabajo con datos de respuestas múltiples

Se pueden utilizar diferentes funciones de comparación para analizar datos de respuestas múltiples, incluyendo:

- value_at
- first_index / last_index
- first_non_null / last_non_null
- first_non_null_index / last_non_null_index
- min_index / max_index

Por ejemplo, suponga una pregunta de respuesta múltiple planteada para la primera, segunda y tercera razón más importante para decidir una compra concreta (por ejemplo, precio, recomendaciones personales, revisión, proveedor local y otros). En este caso, puede determinar la importancia del precio derivando el índice del campo en el que se incluyó por primera vez:

```
first_index("precio", [Razón1 Razón2 Razón3])
```

Del mismo modo, imagine que ha solicitado a los clientes que evalúen 3 coches en función de sus posibilidades de compra y ha codificado las respuestas en tres campos diferentes, de la siguiente forma:

Tabla 5. Ejemplo de clasificación de coches.

ID de cliente	coche1	coche2	coche3
101	1	3	2
102	3	2	1
103	2	3	1

En este caso, puede determinar el índice del campo del coche que guste más (con el mayor o el menor rango) con la función min_index:

```
min_index(['coche1' 'coche2' 'coche3'])
```

Consulte el tema “Funciones de comparación” en la página 93 para obtener más información.

Referencia de conjuntos de respuestas múltiples

La función @MULTI_RESPONSE_SET especial se puede utilizar para hacer referencia a todos los campos de un conjunto de respuestas múltiples. Por ejemplo, si los tres campos *coche* del ejemplo anterior se incluyen en un conjunto de respuesta múltiple denominado *clasificaciones_coche*, la siguiente función devolvería el mismo resultado:

```
max_index(@MULTI_RESPONSE_SET("clasificaciones_coche"))
```

Generador de expresiones

Puede escribir expresiones CLEM manualmente o utilizar el generador de expresiones, que muestra una lista completa de los operadores y funciones de CLEM, así como campos de datos de la ruta actual, lo que permite crear expresiones rápidamente sin tener que memorizar los nombres exactos de los campos o funciones. Además, los controles del generador añaden automáticamente las comillas adecuadas a los campos y valores, lo que facilita la creación de expresiones correctas sintácticamente.

Note: El generador de expresiones no se admite en los scripts ni en la configuración de parámetros.

Acceso al generador de expresiones

El generador de expresiones está disponible en todos los nodos en los que se utilizan expresiones CLEM, incluidos los nodos Seleccionar, Equilibrar, Derivar, Rellenar, Análisis, Informe y Tabla. Para abrirlo, basta con pulsar en el botón de calculadora situado justo a la derecha del campo de fórmulas.

Creación de expresiones

El generador de expresiones ofrece no sólo listas completas de campos, funciones y operadores, sino también acceso a los valores de datos si sus datos están instanciados.

Para crear una expresión utilizando el generador de expresiones:

1. Escríbala en el campo de expresiones, utilizando las listas de campos y funciones como referencias.
o
2. Seleccione las funciones y los campos que desee en las listas de desplazamiento.
3. Pulse dos veces o pulse en el botón con la flecha amarilla para añadir el campo o la función al campo de expresiones.
4. Utilice los botones de operaciones del centro del cuadro de diálogo para insertar las operaciones en la expresión.

Selección de funciones

La lista de funciones muestra todos los operadores y las funciones de CLEM disponibles. Desplácese para seleccionar una función de la lista o, para facilitar su búsqueda, utilice la lista desplegable para que se muestre un subconjunto de funciones u operadores. Las funciones disponibles se agrupan en categorías para facilitar su búsqueda.

Gran parte de estas categorías están descritas en la sección de referencia de la descripción del lenguaje de CLEM. Consulte el tema “Referencia de funciones” en la página 90 para obtener más información.

Las otras categorías son las siguientes:

- **Funciones generales** contiene una selección de algunas de las funciones usadas con mayor frecuencia.
- **Usados recientemente** contiene una lista de las funciones de CLEM utilizadas en la sesión actual.
- **Funciones @** contiene una lista de todas las funciones especiales, cuyos nombres van precedidos del símbolo "@".
- **Funciones de base de datos.** Si la ruta incluye una conexión a la base de datos (a través de un nodo de origen Base de datos), esta selección enumera las funciones disponibles desde la base de datos, incluidas las funciones definidas por el usuario (UDF).
- **Agregados de base de datos.** Si la ruta incluye una conexión a la base de datos (a través de un nodo de origen Base de datos), esta selección enumera las opciones de agregación disponibles desde la base de datos.

- **Agregados de ventana de base de datos.** Si la ruta incluye una conexión a la base de datos (a través de un nodo de origen Base de datos), esta selección enumera las opciones de agregación de ventana que puede utilizar en dicha base de datos.
Aunque se les llama "agregados" estas opciones no están diseñadas para utilizarlas en el nodo Agregar. Resultan más aplicables para nodos, tales como Derivar o Seleccionar. Esto es debido a que su salida es escalar y no una agregación verdadera. Es decir, no disminuyen la cantidad de datos que se muestra en los resultados del mismo que lo hace el nodo Agregar. Por ejemplo, puede utilizar este tipo de agregación para proporcionar una media móvil que descienda por las filas de datos, tales como la "media de la fila actual y de todas las filas".
- **Agregados incorporados.** Contiene una lista de las modalidades de agregación posibles que se pueden utilizar.
- **Operadores** enumera todos los operadores que puede utilizar durante la generación de expresiones. Los operadores también están disponibles en los botones del centro del cuadro de diálogo.
- **Todas las funciones** contiene una lista completa de las funciones de CLEM disponibles.

Una vez que haya seleccionado un grupo de funciones, pulse dos veces para insertar las funciones en el campo de expresiones en el punto indicado por la posición del cursor.

Selección de campos, parámetros y variables globales

La lista de campos muestra todos los campos disponibles en este punto de la ruta de datos. Desplácese para seleccionar un campo de la lista. Pulse dos veces o utilice la tecla de dirección amarilla para añadir un campo a la expresión.

Consulte el tema "Parámetros de ruta, sesión y Supernodo" en la página 74 para obtener más información.

Además de los campos, también puede seleccionar uno de los siguientes elementos:

Conjuntos de respuestas múltiples. Para obtener más información, consulte la guía *Nodos de origen, proceso y resultado de IBM SPSS Modeler*.

Utilizados recientemente contiene una lista de campos, conjuntos de respuestas múltiples, parámetros y valores globales utilizados en la sesión actual.

Parámetros. Consulte el tema "Parámetros de ruta, sesión y Supernodo" en la página 74 para obtener más información.

Valores globales. Para obtener más información, consulte la guía *Nodos de origen, proceso y resultado de IBM SPSS Modeler*.

Presentación o selección de valores

Los valores de campo se pueden visualizar desde diferentes ubicaciones del sistema, incluyendo el generador de expresiones, informes de auditoría de datos y cuando edita valores futuros en un nodo Intervalos de tiempo. Tenga en cuenta que los datos deben estar completamente instanciados en un nodo de origen o Tipo para utilizar esta característica, de forma que se conozcan los elementos de almacenamiento, tipos y valores.

Para visualizar los valores de un campo desde el nodo Generador de expresiones o Intervalos de tiempo, seleccione el campo que desee y pulse en el botón del selector de valores para abrir un cuadro de diálogo con los valores del campo seleccionado. Puede seleccionar un valor y pulsar en **Insertar** para pegar el valor en la expresión o lista actual.



Figura 14. Botón Selector de valores

En los campos nominales y de marcas, se muestran todos los valores. En los campos continuos (rangos numéricos), se muestran los valores máximos y mínimos.

Comprobación de expresiones CLEM

Pulse en **Comprobar** en el generador de expresiones (esquina inferior derecha) para validar la expresión. Las expresiones que no se han comprobado se muestran en rojo. Si se detectan errores, aparece un mensaje indicando la causa.

Se comprueban los siguientes elementos:

- Entrecorillado correcto de valores y nombres de campos
- Uso correcto de parámetros y variables globales
- Uso válido de operadores
- Existencia de campos de referencia
- Existencia y definición de valores globales de referencia

Si se han encontrado errores en la sintaxis, pruebe a crear la expresión utilizando las listas y los botones de operadores en lugar de escribir la expresión manualmente. Con este método se añaden automáticamente las comillas adecuadas para los campos y valores.

Buscar y reemplazar

El cuadro de diálogo Buscar/reemplazar está disponible en lugares donde edita texto de script o de expresión, incluido el editor de scripts, el generador de expresiones CLEM o cuando define una plantilla en el nodo Informe. Al editar texto en cualquiera de estas áreas, pulse Ctrl+F para acceder al cuadro de diálogo, asegurándose de que el cursor está centrado en un área de texto. Por ejemplo, si trabaja en un nodo Rellenar, puede acceder al cuadro de diálogo desde cualquiera de las áreas de texto de la pestaña Configuración o desde el campo de texto del generador de expresiones.

1. Con el cursor en un área de texto, pulse Ctrl+F para acceder al cuadro de diálogo Buscar/reemplazar.
2. Introduzca el texto que desee buscar o selecciónelo de la lista desplegable de elementos buscados recientemente.
3. Introduzca el texto de reemplazo, si lo tiene.
4. Pulse en **Buscar siguiente** para iniciar la búsqueda.
5. Pulse en **Reemplazar** para reemplazar la sección actual o en **Reemplazar todos** para actualizar todas las instancias o sólo las seleccionadas.
6. El cuadro de diálogo se cierra después de cada operación. Pulse F3 desde cualquier área de texto para repetir la operación de búsqueda más reciente o pulse Ctrl+F para volver a acceder al cuadro de diálogo.

Opciones de búsqueda

Coincidir mayúsculas y minúsculas. Especifica si la operación de búsqueda hace distinción entre mayúsculas y minúsculas; por ejemplo, si *miVar* es igual que *miVar*. El texto de reemplazo siempre se introduce exactamente como se ha introducido, independientemente de este ajuste.

Sólo palabras completas. Especifica si la operación de búsqueda tiene en cuenta el texto incluido dentro de las palabras. Por ejemplo, si se selecciona, la búsqueda de *fuego* no será igual que *cortafuegos* ni que *corta-fuegos*.

Expresiones regulares. Especifica si se utiliza la sintaxis de expresiones regulares (consulte la sección siguiente). Si está seleccionado, la opción **Sólo palabras completas** está desactivada y su valor se ignora.

Sólo texto seleccionado. Controla el ámbito de la búsqueda al utilizar la opción **Reemplazar todos**.

Sintaxis de expresiones regulares

Las expresiones regulares le permiten buscar caracteres especiales como caracteres de tabulador o de nueva línea, clases o rangos de caracteres como de la *a* a la *d*, cualquier dígito o no dígito y límites como el principio o el final de una línea. Se admiten los siguientes tipos de expresiones.

Tabla 6. Coincidencias de caracteres.

Caracteres	Coincidencias
x	El carácter x
\\	El carácter de barra inclinada invertida
\\0n	El carácter con valor octal 0n (0 <= n <= 7)
\\0nn	El carácter con valor octal 0nn (0 <= n <= 7)
\\0mnn	El carácter con valor octal 0mnn (0 <= m <= 3, 0 <= n <= 7)
\\xhh	El carácter con valor hexadecimal 0xhh
\\uhhhh	El carácter con valor hexadecimal 0xhhhh
\\t	El carácter de tabulador ('\\u0009')
\\n	El carácter de nueva línea (avance de línea) ('\\u000A')
\\r	El carácter de retorno de carro ('\\u000D')
\\f	El carácter de avance de página ('\\u000C')
\\a	El carácter de alerta (campana) ('\\u0007')
\\e	El carácter de escape ('\\u001B')
\\cx	El carácter de control correspondiente a x

Tabla 7. Coincidencia de clases de caracteres.

Clases de caracteres	Coincidencias
[abc]	a, b o c (clase simple)
[^abc]	Cualquier carácter excepto a, b o c (resta)
[a-zA-Z]	De la a a la z o de la A a la Z, ambas inclusive (rango)
[a-d[m-p]]	De la a a la d o de la m a la p (unión). Esto también puede especificarse como [a-dm-p].
[a-z&&[def]]	De la a a la z y d, e o f (intersección)
[a-z&&[^bc]]	De la a a la z, excepto b y c (resta). Esto también puede especificarse como [ad-z].
[a-z&&[^m-p]]	De la a a la z y no de la m a la p (resta). Esto también puede especificarse como [a-lq-z].

Tabla 8. Clases de caracteres predefinidas.

Clases de caracteres predefinidas	Coincidencias
.	Cualquier carácter (puede o no coincidir con los terminadores de línea)
\\d	Cualquier dígito: [0-9]
\\D	Un carácter que no es un dígito: [^0-9]

Tabla 8. Clases de caracteres predefinidas (continuación).

Clases de caracteres predefinidas	Coincidencias
\s	Un carácter de espacio en blanco : [\t\n\x0B\f\r]
\S	Un carácter de espacio que no está en blanco : [^\s]
\w	Un carácter de palabra: [a-zA-Z_0-9]
\W	Un carácter que no sea de palabra : [^\w]

Tabla 9. Coincidencias de límite.

Reconocedores de límite	Coincidencias
^	El comienzo de una línea
\$	El final de una línea
\b	Un límite alfabético
\B	Un límite no alfabético
\A	El comienzo de la entrada
\Z	El final de la entrada pero para el terminador final, si lo hay
\z	El final de la entrada

Capítulo 8. CLEM Referencia del lenguaje

Conceptos básicos de la referencia de CLEM

Esta sección describe el lenguaje para la manipulación de expresiones de control de (CLEM), una potente herramienta que se utiliza para analizar y manipular los datos utilizados en las rutas de IBM SPSS Modeler. Se puede utilizar CLEM en los nodos para realizar tareas que abarcan desde la evaluación de condiciones o la derivación de valores a la inserción de datos en informes.

Las expresiones CLEM se componen de valores, nombres de campos, operadores y funciones. Utilizando la sintaxis correcta, puede crear una amplia variedad de eficaces operaciones de datos.

Tipos de datos de CLEM

Los tipos de datos de CLEM pueden estar compuestos por lo siguiente:

- Enteros
- Reales
- Caracteres
- Cadenas
- Listas
- Fields
- Fecha/hora

Reglas de entrecorillado

Aunque IBM SPSS Modeler es flexible a la hora de determinar los campos, los valores, los parámetros y las cadenas que se utilizan en una expresión CLEM, las siguientes reglas generales ofrecen una lista de "procedimientos recomendables" que se deben seguir cuando se creen expresiones:

- Cadenas: utilice siempre comillas dobles cuando escriba cadenas, como por ejemplo "Tipo 2". También se pueden utilizar las comillas simples, pero con el riesgo de crear confusión con los campos entrecorillados.
- Campos: utilice comillas simples, sólo cuando sea necesario, para poner espacios u otros caracteres especiales entre comillas, como por ejemplo 'Número de orden'. los campos entrecorillados pero no definidos en el conjunto de datos se confundirán con cadenas.
- Parámetros: utilice siempre comillas simples cuando utilice parámetros, como por ejemplo '\$P-umbra1'.
- Caracteres: utilice siempre comillas inversas simples (`), como por ejemplo stripchar(`d`, "drogaA").

Estas reglas se tratan con mayor detalle en los siguientes temas.

Enteros

Los enteros se representan como una secuencia de dígitos decimales. Si lo desea, puede colocar un signo menos (-) delante del entero para denotar un número negativo; por ejemplo, 1234, 999, 77.

El lenguaje CLEM gestiona enteros de precisión arbitraria. El tamaño máximo de los enteros depende de cada plataforma. Si los valores son demasiado grandes para mostrarlos en un campo de entero, la acción de cambiar el tipo de campo a Real suele restaurar el valor.

Reales

Real hace referencia a un número de coma flotante. Los reales se representan con uno o más dígitos, seguidos por una coma decimal y, después, por uno o más dígitos. Los reales CLEM se guardan en doble precisión.

Si lo desea, puede colocar un signo menos (–) delante del real para denotar un número negativo; por ejemplo, 1,234, 0,999, 77,001. Utilice el formato <número> e <exponente> para expresar un número real en notación exponencial; por ejemplo, 1234,0e5, 1,7e 2. Cuando la aplicación IBM SPSS Modeler lee cadenas numéricas de archivos y las convierte automáticamente en números, se aceptan números sin dígitos antes o después de la coma decimal; por ejemplo, 999. o .11. Sin embargo, estos formatos no son válidos en las expresiones CLEM.

Note: Al hacer referencia a números reales en expresiones CLEM, se debe utilizar el punto como separador decimal, independientemente de lo que se haya definido en la configuración regional o ruta actual. Por ejemplo, especifique:

Na > 0,6

en vez de

Na > 0.6

Todas las opciones anteriores se aplican aunque se haya seleccionado una coma como símbolo decimal en el cuadro de diálogo de propiedades de la ruta y es coherente con la norma general de que la sintaxis de código debe ser independiente de cualquier convención o configuración regional específica.

Caracteres

Los caracteres (mostrados normalmente como CAR) se suelen utilizar dentro de una expresión CLEM para realizar comprobaciones en cadenas. Por ejemplo, puede utilizar la función `isuppercode` para determinar si el primer carácter de una cadena va en mayúsculas. La siguiente expresión CLEM utiliza un carácter para indicar que la comprobación se debe realizar en el primer carácter de la cadena:

```
isuppercode(subscrs(1, "MiCadena"))
```

Para expresar el código (en lugar de la ubicación) de un carácter determinado en una expresión de CLEM, utilice comillas inversas simples del tipo ``carácter``, por ejemplo, ``A``, ``Z``.

Note: No hay ningún tipo de almacenamiento CAR para los campos, por lo que si un campo se deriva o rellena con una expresión que da como resultado un CAR, este resultado se convertirá en una cadena.

Cadenas

Por lo general, debe encerrar las cadenas entre comillas dobles. Algunos ejemplos de cadenas son "c35producto2" e "IDdereferencia". Para indicar caracteres especiales en una cadena, utilice una barra inclinada invertida; por ejemplo, "\\$65443". (Para indicar un carácter de barra inclinada invertida, utilice una barra inclinada invertida doble, `\\`). Puede utilizar comillas simples para las cadenas, pero el resultado no se distinguirá entonces de un campo entrecomillado ('IDdereferencia'). Consulte el tema "Funciones de cadena" en la página 100 para obtener más información.

Listas

Una lista es una secuencia ordenada de elementos, que pueden ser de distintos tipos. Las listas se encierran entre corchetes (`[]`). Algunos ejemplos de listas son `[1 2 4 16]` y `["abc" "def"]`. Las listas no se utilizan como valores de los campos de IBM SPSS Modeler. Se utilizan para proporcionar argumentos a funciones, como por ejemplo `member` y `oneof`.

Fields

Los nombres de las expresiones CLEM que no son nombres de funciones se consideran nombres de campos. Puede escribirlos simplemente como `Potencia`, `val27`, `marca_estado`, etc., pero si el nombre comienza por un dígito o incluye caracteres no alfabéticos, como por ejemplo espacios (a excepción del guión bajo), debe poner el nombre entre comillas simples; por ejemplo, `'Aumento de potencia'`, `'2a pregunta'`, `'#101'`, `'$P-NextField'`.

Note: los campos entrecomillados pero no definidos en el conjunto de datos se confundirán con cadenas.

Fechas

Los cálculos de fecha se basan en una fecha de "línea base", la cual se especifica en el cuadro de diálogo de propiedades de la ruta. La fecha de línea base predeterminada es el 1 de enero de 1900.

El lenguaje CLEM admite los siguientes formatos de fecha.

Tabla 10. Formatos de fecha de idioma de CLEM

Formato	Ejemplos
DDMMAA	150163
MMDDAA	011563
AAMMDD	630115
AAAAMMDD	19630115
AAAADDD	Año de cuatro dígitos, seguido de un número de tres dígitos que representa el día del año; por ejemplo, 2000032 representa el día 32 de 2000 o el 1 de febrero de 2000.
DAY	Día de la semana en el entorno local actual, por ejemplo, Lunes, Martes, ..., en inglés.
MONTH	Mes en la configuración regional actual; por ejemplo, Enero, Febrero, ..., etc.
DD/MM/AA	15/01/63
DD/MM/AAAA	15/01/1963
MM/DD/AA	01/15/63
MM/DD/AAAA	01/15/1963
DD-MM-AA	15-01-63
DD-MM-AAAA	15-01-1963
MM-DD-AA	01-15-63
MM-DD-AAAA	01-15-1963
DD.MM.AA	15.01.63
DD.MM.AAAA	15.01.1963
MM.DD.AA	01.15.63
MM.DD.AAAA	01.15.1963
DD-MES-AA	15-ENE-63, 15-ene-63, 15-Ene-63
DD/MES/AA	15/ENE/63, 15/ene/63, 15/Ene/63
DD.MES.AA	15.ENE.63, 15.ene.63, 15.Ene.63
DD-MES-AAAA	15-ENE-1963, 15-ene-1963, 15-Ene-1963
DD/MES/AAAA	15/ENE/1963, 15/ene/1963, 15/Ene/1963
DD.MES.AAAA	15.ENE.1963, 15.ene.1963, 15.Ene.1963

Tabla 10. Formatos de fecha de idioma de CLEM (continuación)

Formato	Ejemplos
MES AAAA	Ene 2004
q Q YYYY	Fecha en forma de dígito (1–4), que representa el trimestre seguido de la letra <i>T</i> y un año de cuatro dígitos; por ejemplo, la fecha 25 Dic 2004 se representaría como 4 T 2004.
ss SE AAAA	Número de dos dígitos que representa la semana del año, seguido de las letras <i>SE</i> y, a continuación, un año de cuatro dígitos. La semana del año se calcula asumiendo que el primer día de la semana es lunes y que hay al menos un día en la primera semana.

Hora

El lenguaje CLEM admite los siguientes formatos de hora.

Tabla 11. Formatos de hora de idioma de CLEM

Formato	Ejemplos
HHMMSS	120112, 010101, 221212
HHMM	1223, 0745, 2207
MMSS	5558, 0100
HH:MM:SS	12:01:12, 01:01:01, 22:12:12
HH:MM	12:23, 07:45, 22:07
MM:SS	55:58, 01:00
(H)H:(M)M:(S)S	12:1:12, 1:1:1, 22:12:12
(H)H:(M)M	12:23, 7:45, 22:7
(M)M:(S)S	55:58, 1:0
HH.MM.SS	12.01.12, 01.01.01, 22.12.12
HH.MM	12.23, 07.45, 22.07
MM.SS	55.58, 01.00
(H)H.(M)M.(S)S	12.1.12, 1.1.1, 22.12.12
(H)H.(M)M	12.23, 7.45, 22.7
(M)M.(S)S	55.58, 1.0

Operadores de CLEM

Los siguientes operadores están disponibles:

Tabla 12. Operadores de idioma de CLEM.

Operación	Comentarios	Prioridad (consulte la sección siguiente)
o	Se utiliza entre dos expresiones CLEM. Devuelve un valor de verdadero si una de ellas es verdadera o si ambas lo son.	10
y	Se utiliza entre dos expresiones CLEM. Devuelve un valor de verdadero si ambas son verdaderas.	9

Tabla 12. Operadores de idioma de CLEM (continuación).

Operación	Comentarios	Prioridad (consulte la sección siguiente)
=	Se utiliza entre dos elementos comparables. Devuelve verdadero si el ELEMENTO1 NO es igual al ELEMENTO2.	7
==	Idéntico a =.	7
/=	Se utiliza entre dos elementos comparables. Devuelve verdadero si el ELEMENTO1 <i>no</i> es igual al ELEMENTO2.	7
/==	Idéntico a /=.	7
>	Se utiliza entre dos elementos comparables. Devuelve verdadero si el ELEMENTO1 es estrictamente mayor que el ELEMENTO2.	6
>=	Se utiliza entre dos elementos comparables. Devuelve verdadero si el ELEMENTO1 es mayor o igual que el ELEMENTO2.	6
<	Se utiliza entre dos elementos comparables. Devuelve verdadero si el ELEMENTO1 es estrictamente menor que el ELEMENTO2.	6
<=	Se utiliza entre dos elementos comparables. Devuelve verdadero si el ELEMENTO1 es menor o igual que el ELEMENTO2.	6
&&=_0	Se utiliza entre dos enteros. Equivalente a la expresión booleana ENT1 && ENT2 = 0.	6
&&/=_0	Se utiliza entre dos enteros. Equivalente a la expresión booleana ENT1 && ENT2 /= 0.	6
+	Suma dos números: NÚM1 + NÚM2.	5
>>	Concatena dos cadenas, por ejemplo, CADENA1 >> CADENA2	5
-	Resta un número a otro: NÚM1 - NÚM2. También se puede utilizar delante de un número:-NÚM.	5
*	Se utiliza para multiplicar dos números: NÚM1 * NÚM2	4
&&	Se utiliza entre dos enteros. El resultado es el 'y' a nivel de bit de los enteros ENT1 y ENT2.	4
&&~	Se utiliza entre dos enteros. El resultado es el 'y' a nivel de bit de ENT1 y el complemento a nivel de bit de ENT2.	4
	Se utiliza entre dos enteros. El resultado es el 'o inclusivo' a nivel de bit de ENT1 y ENT2.	4
~	Se utiliza delante de un entero. Genera el complemento a nivel de bit de ENT.	4
/&	Se utiliza entre dos enteros. El resultado es el 'o exclusivo' a nivel de bit de ENT1 y ENT2.	4
ENT1 << N	Se utiliza entre dos enteros. Genera el patrón de bits de ENT desplazado a la izquierda N posiciones.	4

Tabla 12. Operadores de idioma de CLEM (continuación).

Operación	Comentarios	Prioridad (consulte la sección siguiente)
ENT1 >> N	Se utiliza entre dos enteros. Genera el patrón de bits de ENT desplazado a la derecha N posiciones.	4
/	Se utiliza para dividir un número por otro: NÚM1 / NÚM2.	4
**	Se utiliza entre dos números: BASE ** POTENCIA. Devuelve BASE elevado a la potencia POTENCIA.	3
rem	Se utiliza entre dos enteros: ENT1 rem ENT2. Devuelve el resto, ENT1 - (ENT1 div ENT2) * ENT2.	2
div	Se utiliza entre dos enteros: ENT1 div ENT2. Realiza la división de enteros.	2

Prioridad del operador

Las prioridades determinan el análisis de expresiones complejas, especialmente aquellas que no van entre paréntesis y tienen más de un operador infijo. Por ejemplo:

$3 + 4 * 5$

se puede calcular como $3 + (4 * 5)$ en lugar de $(3 + 4) * 5$, ya que las prioridades relativas dictan que * se analiza antes que +. Cada operador del lenguaje CLEM tiene un valor de prioridad asociado y, mientras más bajo sea éste, más importante será en la lista de análisis, lo que significa que será procesado antes que otros operadores con valores de prioridad superiores.

Referencia de funciones

Las siguientes funciones de CLEM están disponibles para trabajar con datos en IBM SPSS Modeler. Puede introducir estas funciones como código en diversos cuadros de diálogo, como los de los nodos Derivar y Marcas, o puede utilizar el generador de expresiones para crear expresiones CLEM válidas sin memorizar listas de funciones ni nombres de campos.

Tabla 13. Funciones de CLEM para utilizarlas con datos de IBM SPSS Modeler

Tipo de función	Descripción
Información	Se utiliza para comprender mejor los valores de campos. Por ejemplo, la función <code>is_string</code> devuelve una true para todos los registros cuyo tipo sea una cadena.
Conversión	Se utiliza para construir nuevos campos o convertir el tipo de almacenamiento. Por ejemplo, la función <code>to_timestamp</code> convierte el campo seleccionado en una marca de tiempo.
Comparación	Se utiliza para comparar valores de campos entre ellos o con una cadena especificada. Por ejemplo, <code><=</code> se utiliza para comparar y ver si los valores de dos campos son menores o iguales.
Lógico	Se utilizan para realizar operaciones lógicas, como por ejemplo operaciones <code>if, then, else</code> .
Numéricas	Se utilizan para realizar cálculos numéricos como, por ejemplo, el logaritmo natural de valores de campos.
Trigonométricas	Se utilizan para realizar cálculos trigonométricos como, por ejemplo, el arcoseno de un ángulo especificado.

Tabla 13. Funciones de CLEM para utilizarlas con datos de IBM SPSS Modeler (continuación)

Tipo de función	Descripción
Probability	Probabilidades basadas en distintas distribuciones, como probabilidad de que un valor de una distribución <i>t</i> de Student sea menor que un valor específico.
A nivel de bit	Se utiliza para manipular enteros como patrones de bits.
Aleatorias	Se utilizan para seleccionar elementos al azar o generar números.
Cadena	Se utiliza para realizar numerosas operaciones en cadenas, como por ejemplo stripchar, que permite eliminar un carácter especificado.
SoundEx	Se utiliza para encontrar cadenas cuando no se conoce la ortografía exacta, se basa en supuestos fonéticos sobre cómo se pronuncian ciertas letras.
La fecha y la hora	Se utiliza para realizar diversas operaciones en los campos de fecha, hora y marca de tiempo.
Secuencia	Se utiliza para comprender mejor la secuencia de registros de un conjunto de datos o para realizar operaciones en función de dicha secuencia.
Globales	Se utilizan para acceder a los valores globales creados por un nodo Val. globales. Por ejemplo, @MEAN se utiliza para hacer referencia a la media de todos los valores de un campo en todo el conjunto de datos.
Vacíos y nulos	Se utiliza para acceder, marcar y, con frecuencia, rellenar elementos vacíos especificados por el usuario y valores perdidos por el sistema. Por ejemplo, @BLANK(CAMPO) se utiliza para devolver una marca verdadera para los registros donde hay elementos vacíos.
Campos especiales	Se utilizan para denotar los campos específicos que están siendo examinados. Por ejemplo, @FIELD se utiliza cuando se derivan varios campos.

Convenciones en las descripciones de las funciones

En este manual se utilizan las siguientes convenciones para referirse a los elementos de una función.

Tabla 14. Convenciones en descripciones de función.

Convención	Descripción
BOOL	Un valor booleano, o una marca, como por ejemplo verdadero o falso.
NÚM, NÚM1, NÚM2	Cualquier número.
REAL, REAL1, REAL2	Cualquier número real, como por ejemplo 1,234 ó -77,01.
ENT, ENT1, ENT2	Cualquier entero, como por ejemplo 1 ó -77.
CAR	Un código de carácter, como por ejemplo `A`.
STRING	Una cadena, como por ejemplo "IDdereferencia".
LIST	Una lista de elementos, como por ejemplo ["abc" "def"].
ELEMENTO	Un campo, como por ejemplo Cliente o extraer_concepto.
DATE	Un campo de fecha, como por ejemplo fecha_inicio, donde los valores tienen formatos tales como DD-MES-AAAA.
TIME	Un campo de hora, como por ejemplo flujo_de_potencia, donde los valores tienen formatos tales como HHMMSS.

Las funciones de este manual se enumeran con la función en una columna, el tipo de resultado (entero, cadena, etc.) en otra y una descripción (si está disponible) en una tercera columna. Por ejemplo, a continuación se muestra la descripción de la función rem.

Tabla 15. Descripción de la función rem.

Función	Resultado	Descripción
ENT1 rem ENT2	Number	Devuelve el resto de dividir ENT1 entre ENT2. Por ejemplo, ENT1 – (ENT1 div ENT2) * ENT2.

En otra sección se ofrecen detalles sobre el uso de las convenciones, como por ejemplo cómo crear una lista de elementos o especificar caracteres en una función. Consulte el tema “Tipos de datos de CLEM” en la página 85 para obtener más información.

Funciones de información

Las funciones de información se utilizan para comprender mejor los valores de un campo en particular. Se suelen utilizar para derivar campos de marcas. Por ejemplo, puede utilizar la función @BLANK para crear un campo de marca que indique los registros cuyos valores están vacíos para el campo seleccionado. De igual modo, puede comprobar el tipo de almacenamiento de un campo utilizando cualquiera de las funciones de tipo de almacenamiento, como por ejemplo is_string.

Tabla 16. CLEMfunciones de información.

Función	Resultado	Descripción
@BLANK(CAMPO)	Booleana	Devuelve verdadero para todos los registros cuyos valores están vacíos de acuerdo con el conjunto de reglas de tratamiento de los elementos vacíos en un nodo Tipo o de origen (pestaña Tipos) anterior de la ruta.
@NULL(CAMPO)	Booleana	Devuelve verdadero para todos los registros cuyos valores no están definidos. Los valores no definidos son valores nulos del sistema y se muestran en IBM SPSS Modeler como \$null\$.
is_date(ELEMENTO)	Booleana	Devuelve verdadero para todos los registros cuyo tipo es una fecha.
is_datetime(ELEMENTO)	Booleana	Devuelve verdadero para todos los registros cuyo tipo es una fecha, hora o marca de tiempo.
is_integer(ELEMENTO)	Booleana	Devuelve verdadero para todos los registros cuyo tipo es un entero.
is_number(ELEMENTO)	Booleana	Devuelve verdadero para todos los registros cuyo tipo es un número.
is_real(ELEMENTO)	Booleana	Devuelve verdadero para todos los registros cuyo tipo es un número real.
is_string(ELEMENTO)	Booleana	Devuelve verdadero para todos los registros cuyo tipo es una cadena.
is_time(ELEMENTO)	Booleana	Devuelve verdadero para todos los registros cuyo tipo es una hora.
is_timestamp(ELEMENTO)	Booleana	Devuelve verdadero para todos los registros cuyo tipo es una marca de tiempo.

Funciones de conversión

Las funciones de conversión permiten construir nuevos campos y convertir el tipo de almacenamiento de los archivos existentes. Por ejemplo, puede formar cadenas nuevas uniendo o desmontando otras. Para unir dos cadenas, utilice el operador ><. Por ejemplo, si el campo Sitio tiene el valor "BRAMLEY", "xx" >< Sitio devolverá "xxBRAMLEY". El resultado de >< siempre es una cadena, aun cuando los argumentos no sean cadenas. Así, si el campo V1 es 3 y el campo V2 es 5, V1 >< V2 devolverá "35" (una cadena, no un número).

Las funciones de conversión (y otras muchas funciones que requieren un tipo específico de entrada, como un valor de fecha u hora) dependen de los formatos actuales especificados en el cuadro de diálogo de opciones de ruta. Por ejemplo, si desea convertir un campo de cadena con los valores *Ene 2003, Feb 2003, etc.*, seleccione el formato de fecha coincidente **MES AAAA** como formato de fecha predeterminado de la ruta.

Tabla 17. CLEMfunciones de conversión

Función	Resultado	Descripción
ELEMENTO1 >< ELEMENTO2	<i>Cadena</i>	Concatena los valores de dos campos y devuelve la cadena resultante como <i>ELEMENTO1ELEMENTO2</i> .
to_integer(ELEMENTO)	<i>Entero</i>	Convierte el almacenamiento del campo especificado en un entero.
to_real(ELEMENTO)	<i>Real</i>	Convierte el almacenamiento del campo especificado en un número real.
to_number(ELEMENTO)	<i>Number</i>	Convierte el almacenamiento del campo especificado en un número.
to_string(ELEMENTO)	<i>Cadena</i>	Convierte el almacenamiento del campo especificado en una cadena.
to_time(ELEMENTO)	<i>Hora</i>	Convierte el almacenamiento del campo especificado en una hora.
to_date(ELEMENTO)	<i>Fecha</i>	Convierte el almacenamiento del campo especificado en una fecha.
to_timestamp(ELEMENTO)	<i>Marca de tiempo</i>	Convierte el almacenamiento del campo especificado en una marca de tiempo.
to_datetime(ELEMENTO)	<i>Momento_fecha</i>	Convierte el almacenamiento del campo especificado en una fecha, hora o marca de tiempo.
datetime_date(ELEMENTO)	<i>Fecha</i>	Devuelve el valor de fecha de un <i>número, cadena o marca de tiempo</i> . Tenga en cuenta que es la única función que le permite volver a convertir un número (en segundos) en una fecha. Si ELEMENTO es una cadena, crea una fecha analizando una cadena en el formato de fecha actual. El formato de fecha especificado en el cuadro de diálogo de propiedades de la ruta debe ser correcto para que esta función se ejecute correctamente. Si ELEMENTO es un número, se interpreta como un número de segundos desde la fecha de base (o época). Las fracciones de un día se truncan. Si ELEMENTO es una marca de tiempo, la parte de la fecha de la marca de tiempo se devuelve. Si ELEMENTO es una fecha, se devuelve sin modificar.

Funciones de comparación

Las funciones de comparación se utilizan para comparar valores de campos entre ellos o con una cadena especificada. Por ejemplo, puede comprobar la igualdad de cadenas utilizando =. Un ejemplo de verificación de igualdad de cadena es: `Class = "clase 1"`.

En la comparación numérica, *mayor* significa más cerca del infinito positivo y *menor*, más cerca del infinito negativo. Es decir, que todos los números negativos son menores que cualquier número positivo.

Tabla 18. CLEMfunciones de comparación

Función	Resultado	Descripción
count_equal(ELEMENTO1, LISTA)	<i>Entero</i>	Devuelve el número de valores de una lista de campos que son iguales a <i>ELEMENTO1</i> o nulos si <i>ELEMENTO1</i> es nulo.

Tabla 18. CLEMfunciones de comparación (continuación)

Función	Resultado	Descripción
count_greater_than(ELEMENTO1, LISTA)	Entero	Devuelve el número de valores de una lista de campos que son mayores que <i>ELEMENTO1</i> o nulos si <i>ELEMENTO1</i> es nulo.
count_less_than(ELEMENTO1, LISTA)	Entero	Devuelve el número de valores de una lista de campos que son menores que <i>ELEMENTO1</i> o nulos si <i>ELEMENTO1</i> es nulo.
count_not_equal(ELEMENTO1, LISTA)	Entero	Devuelve el número de valores de una lista de campos que no son iguales a <i>ELEMENTO1</i> o son nulos si <i>ELEMENTO1</i> es nulo.
count_nulls(LISTA)	Entero	Devuelve el número de valores nulos de una lista de campos.
count_non_nulls(LISTA)	Entero	Devuelve el número de valores no nulos de una lista de campos.
date_before(FECHA1, FECHA2)	Booleana	Se utiliza para comprobar el orden de los valores de fecha. Devuelve un valor verdadero si <i>FECHA1</i> es anterior a <i>FECHA2</i> .
first_index(ELEMENTO, LISTA)	Entero	Devuelve el índice del primer campo que contiene <i>ELEMENTO</i> de una <i>LISTA</i> de campos o 0 si no se encuentra el valor. Compatible con los tipos cadena, entero y real únicamente.
first_non_null(LISTA)	<i>Cualquier caracterización correcta relacionada de la que no se esté seguro del todo deberá comunicarse al departamento Global Tax o al EHQ, APG o asesores fiscales de LA.</i>	Devuelve el primer valor que no sea nulo en la lista de campos. Se admiten todos los tipos de almacenamiento.
first_non_null_index(LISTA)	Entero	Devuelve el índice del primer campo en la <i>LISTA</i> especificada con un valor no nulo o 0 si todos los valores son nulos. Se admiten todos los tipos de almacenamiento.
ELEMENTO1 = ELEMENTO2	Booleana	Devuelve verdadero para los registros donde <i>ELEMENTO1</i> es igual a <i>ELEMENTO2</i> .
ELEMENTO1 /= ELEMENTO2	Booleana	Devuelve verdadero si las dos cadenas no son idénticas o 0 si son idénticas.
ELEMENTO1 < ELEMENTO2	Booleana	Devuelve verdadero para los registros donde <i>ELEMENTO1</i> es menor que <i>ELEMENTO2</i> .
ELEMENTO1 <= ELEMENTO2	Booleana	Devuelve verdadero para los registros donde <i>ELEMENTO1</i> es menor o igual que <i>ELEMENTO2</i> .
ELEMENTO1 > ELEMENTO2	Booleana	Devuelve verdadero para los registros donde <i>ELEMENTO1</i> es mayor que <i>ELEMENTO2</i> .
ELEMENTO1 >= ELEMENTO2	Booleana	Devuelve verdadero para los registros donde <i>ELEMENTO1</i> es mayor o igual que <i>ELEMENTO2</i> .
last_index(ELEMENTO, LISTA)	Entero	Devuelve el índice del último campo que contiene <i>ELEMENTO</i> de una <i>LISTA</i> de campos o 0 si no se encuentra el valor. Compatible con los tipos cadena, entero y real únicamente.

Tabla 18. CLEMfunciones de comparación (continuación)

Función	Resultado	Descripción
last_non_null(LISTA)	<i>Cualquier caracterización correcta relacionada de la que no se esté seguro del todo deberá comunicarse al departamento Global Tax o al EHQ, APG o asesores fiscales de LA.</i>	Devuelve el último valor que no sea nulo en la lista de campos. Se admiten todos los tipos de almacenamiento.
last_non_null_index(LISTA)	<i>Entero</i>	Devuelve el índice del último campo en la LISTA especificada con un valor no nulo o 0 si todos los valores son nulos. Se admiten todos los tipos de almacenamiento.
max(ELEMENTO1, ELEMENTO2)	<i>Cualquier caracterización correcta relacionada de la que no se esté seguro del todo deberá comunicarse al departamento Global Tax o al EHQ, APG o asesores fiscales de LA.</i>	Devuelve el mayor de los dos elementos: <i>ELEMENTO1</i> o <i>ELEMENTO2</i> .
max_index(LISTA)	<i>Entero</i>	Devuelve el índice del campo con el valor máximo de una lista de campos numéricos o 0 si todos los valores son nulos. Por ejemplo, si el tercer campo de la lista contiene el máximo, se devuelve el valor de índice 3. Si varios campos contienen el valor máximo, se devuelve el valor que aparece primero (el primero por la izquierda).
max_n(LISTA)	<i>Number</i>	Devuelve el valor máximo de una lista de campos numéricos o es nulo si todos los valores de campo son nulos.
member(ELEMENTO, LISTA)	<i>Booleana</i>	Devuelve verdadero si <i>ELEMENTO</i> es un miembro de la <i>LISTA</i> especificada. En el resto de los casos, devuelve falso. También se puede especificar una lista de nombres de campo.
min(ELEMENTO1, ELEMENTO2)	<i>Cualquier caracterización correcta relacionada de la que no se esté seguro del todo deberá comunicarse al departamento Global Tax o al EHQ, APG o asesores fiscales de LA.</i>	Devuelve el menor de los dos elementos: <i>ELEMENTO1</i> o <i>ELEMENTO2</i> .
min_index(LISTA)	<i>Entero</i>	Devuelve el índice del campo con el valor mínimo de una lista de campos numéricos o 0 si todos los valores son nulos. Por ejemplo, si el tercer campo de la lista contiene el mínimo, se devuelve el valor de índice 3. Si varios campos contienen el valor mínimo, se devuelve el valor que aparece primero (el primero por la izquierda).

Tabla 18. CLEMfunciones de comparación (continuación)

Función	Resultado	Descripción
min_n(LISTA)	Number	Devuelve el valor mínimo de una lista de campos numéricos o es nulo si todos los valores de campo son nulos.
time_before(TIEMP01, TIEMP02)	Booleana	Se utiliza para comprobar el orden de los valores de hora. Devuelve un valor verdadero si HORA1 es anterior a HORA2.
value_at(ENT, LISTA)		Devuelve el valor de cada campo de la lista con un desplazamiento INT o NULL si el desplazamiento está fuera del rango de valores válidos (es decir, menor que 1 o mayor que el número de campos de la lista). Se admiten todos los tipos de almacenamiento.

Funciones lógicas

Las expresiones CLEM se pueden utilizar para realizar operaciones lógicas.

Tabla 19. CLEMfunciones lógicas

Función	Resultado	Descripción
COND1 and COND2	Booleana	Esta operación es una conjunción lógica y devuelve un valor verdadero si COND1 y COND2 son verdaderas. Si COND1 es falsa, COND2 no se evalúa; esto permite tener conjunciones donde COND1 comprueba primero si una operación en COND2 es legal. Por ejemplo, length(Etiqueta) >=6 y Etiqueta(6) = 'x'.
COND1 o COND2	Booleana	Esta operación es una disyunción (inclusiva) lógica y devuelve verdadero si COND1 o COND2 son verdaderas, o bien si ambas lo son. Si COND1 es verdadera, COND2 no se evalúa.
not(COND)	Booleana	Esta operación es una negación lógica y devuelve un valor verdadero si COND es falsa. En el resto de los casos, esta operación devuelve un valor de 0.
if COND then EXPR1 else EXPR2 endif	Cualquier caracterización correcta relacionada de la que no se esté seguro del todo deberá comunicarse al departamento Global Tax o al EHQ, APG o asesores fiscales de LA.	Esta operación es una evaluación condicional. Si COND es verdadera, esta operación devuelve el resultado de EXPR1. En el resto de los casos, devuelve el resultado de evaluar EXPR2.
if COND1 then EXPR1 elseif COND2 then EXPR2 else EXPR_N endif	Cualquier caracterización correcta relacionada de la que no se esté seguro del todo deberá comunicarse al departamento Global Tax o al EHQ, APG o asesores fiscales de LA.	Esta operación es una evaluación condicional de varias ramas. Si COND1 es verdadera, esta operación devuelve el resultado de EXPR1. En caso contrario, si COND2 es verdadera, esta operación devuelve el resultado de evaluar EXPR2. En caso contrario, devuelve el resultado de evaluar EXPR_N.

Funciones numéricas

CLEM contiene varias funciones numéricas de uso común.

Tabla 20. CLEMfunciones numéricas.

Función	Resultado	Descripción
-NÚM	Number	Se utiliza para negar NÚM. Devuelve el número correspondiente con el signo opuesto.
NÚM1 + NÚM2	Number	Devuelve la suma de NÚM1 y NÚM2.
NÚM1 - NÚM2	Number	Devuelve el valor de NÚM1 menos NÚM2.
NÚM1 * NÚM2	Number	Devuelve el valor de NÚM1 multiplicado por NÚM2.
NÚM1 / NÚM2	Number	Devuelve el valor de NÚM1 dividido por NÚM2.
ENT1 div ENT2	Number	Se utiliza para realizar una división entera. Devuelve el valor de ENT1 dividido por ENT2.
ENT1 rem ENT2	Number	Devuelve el resto de dividir ENT1 entre ENT2. Por ejemplo, ENT1 - (ENT1 div ENT2) * ENT2.
ENT1 mod ENT2	Number	Esta función está desaprobada. Utilice la función rem en su lugar.
BASE ** POTENCIA	Number	Devuelve BASE elevada a la potencia POTENCIA, donde ambas pueden ser cualquier número (con la excepción de que BASE no debe ser cero si POTENCIA es cero de cualquier tipo, menos el entero 0). Si POTENCIA es un entero, el cálculo se realiza multiplicando sucesivamente potencias de BASE. Así, si BASE es un entero, el resultado será un entero. Si POTENCIA es el entero 0, el resultado siempre será un 1 del mismo tipo que BASE. De lo contrario, si POTENCIA no es un entero, el resultado se calculará como $\exp(\text{POTENCIA} * \log(\text{BASE}))$.
abs(NÚM)	Number	Devuelve el valor absoluto de NÚM, que siempre será un número del mismo tipo.
exp(NÚM)	Real	Devuelve e elevado a la potencia de NÚM, donde e es la base de los logaritmos naturales.
fracof(NÚM)	Real	Devuelve la parte fraccional de NÚM, definida como $\text{NÚM} - \text{intof}(\text{NÚM})$.
intof(NÚM)	Entero	Trunca su argumento a un entero. Devuelve el entero del mismo signo de NÚM y con la mayor magnitud tal que $\text{abs}(\text{ENT}) \leq \text{abs}(\text{NÚM})$.
log(NÚM)	Real	Devuelve el logaritmo natural (en base e) de NÚM, que no puede ser un cero de ningún tipo.
log10(NÚM)	Real	Devuelve el logaritmo en base 10 de NÚM, que no puede ser un cero de ningún tipo. Esta función se define como $\log(\text{NUM}) / \log(10)$.
negate(NÚM)	Number	Se utiliza para negar NÚM. Devuelve el número correspondiente con el signo opuesto.
round(NÚM)	Entero	Se utiliza para redondear NÚM a un número entero tomando $\text{intof}(\text{NÚM}+0,5)$ si NÚM es positivo o $\text{intof}(\text{NÚM}-0,5)$ si NÚM es negativo.
sign(NÚM)	Number	Se utiliza para determinar el signo de NÚM. Esta operación devuelve -1, 0 ó 1 si NÚM es un entero. Si NÚM es un real, devuelve -1,0, 0,0 ó 1,0 dependiendo de si NÚM es negativo, cero o positivo.
sqrt(NÚM)	Real	Devuelve la raíz cuadrada de NÚM. NÚM debe ser positivo.
sum_n(LISTA)	Number	Devuelve la suma de valores de una lista de campos numéricos o es nulo si todos los valores de campo son nulos.

Tabla 20. CLEMfunciones numéricas (continuación).

Función	Resultado	Descripción
mean_n(LISTA)	Number	Devuelve el valor medio de una lista de campos numéricos o es nulo si todos los valores de campo son nulos.
sdev_n(LISTA)	Number	Devuelve la desviación estándar de una lista de campos numéricos o es nulo si todos los valores de campo son nulos.

Funciones trigonométricas

Todas las funciones de esta sección toman un ángulo como argumento o devuelven uno como resultado. En ambos casos, las unidades del ángulo (radianes o grados) se controlan con la configuración de la opción de ruta correspondiente.

Tabla 21. CLEMfunciones trigonométricas

Función	Resultado	Descripción
arccos(NÚM)	Real	Calcula el arcocoseno del ángulo especificado.
arccosh(NÚM)	Real	Calcula el arcocoseno hiperbólico del ángulo especificado.
arcsin(NÚM)	Real	Calcula el arcoseno del ángulo especificado.
arcsinh(NÚM)	Real	Calcula el arcoseno hiperbólico del ángulo especificado.
arctan(NÚM)	Real	Calcula la arcotangente del ángulo especificado.
arctan2(NÚM_Y, NÚM_X)	Real	Calcula la arcotangente de NÚM_Y / NÚM_X y utiliza los signos de los dos números para derivar la información sobre el cuadrante. El resultado es un número real comprendido entre $-\pi < \text{ÁNGULO} \leq \pi$ (radianes) $-180 < \text{ÁNGULO} \leq 180$ (grados).
arctanh(NÚM)	Real	Calcula la arcotangente hiperbólica del ángulo especificado.
cos(NÚM)	Real	Calcula el coseno del ángulo especificado.
cosh(NÚM)	Real	Calcula el coseno hiperbólico del ángulo especificado.
pi	Real	Esta constante es la mejor aproximación real de pi.
sin(NÚM)	Real	Calcula el seno del ángulo especificado.
sinh(NÚM)	Real	Calcula el seno hiperbólico del ángulo especificado.
tan(NÚM)	Real	Calcula la tangente del ángulo especificado.
tanh(NÚM)	Real	Calcula la tangente hiperbólica del ángulo especificado.

Funciones de probabilidad

Las funciones de probabilidad devuelven probabilidades basadas en distintas distribuciones, como probabilidad de que un valor de una distribución *t* de Student sea menor que un valor específico.

Tabla 22. CLEMfunciones de probabilidad

Función	Resultado	Descripción
cdf_chisq(NÚM, GL)	Real	Devuelve la probabilidad de que un valor de la distribución chi-cuadrado con los grados especificados de libertad sea menor que el número especificado.
cdf_f(NÚM, GL1, GL2)	Real	Devuelve la probabilidad de que un valor de la distribución <i>F</i> con los grados de libertad <i>DF1</i> y <i>DF2</i> sea menor que el número especificado.
cdf_normal(NÚM, MEDIA, DESTÍP)	Real	Devuelve la probabilidad de que un valor de la distribución normal con la desviación media y estándar especificada sea menor que el número especificado.

Tabla 22. CLEMfunciones de probabilidad (continuación)

Función	Resultado	Descripción
<code>cdf_t(NÚM, GL)</code>	<i>Real</i>	Devuelve la probabilidad de que un valor de la distribución <i>t</i> de Student con los grados especificados de libertad sea menor que el número especificado.

Operaciones de enteros a nivel de bit

Estas funciones permiten manipular los enteros como patrones de bits que representan valores en complemento a dos, donde la posición del bit *N* tiene la ponderación 2^{**N} . Los bits se numeran de 0 en adelante. Estas operaciones actúan como si el bit de signo de un entero se extendiera indefinidamente hacia la izquierda. Así, por encima de su bit más significativo, un entero positivo tiene 0 bits y un entero negativo tiene 1 bit.

Tabla 23. Operaciones de enteros bit a bit de CLEM.

Función	Resultado	Descripción
<code>~~ ENT1</code>	<i>Entero</i>	Genera el complemento a nivel de bit del entero <i>ENT1</i> . Es decir, hay un 1 en el resultado de cada posición de bits para la que <i>ENT1</i> tiene 0. Siempre es verdadero que <code>~~ ENT = -(ENT + 1)</code> .
<code>ENT1 ENT2</code>	<i>Entero</i>	El resultado de esta operación es el "o inclusivo" a nivel de bit de <i>ENT1</i> y <i>ENT2</i> . Es decir, hay un 1 en el resultado de cada posición de bits para la que hay un 1 en <i>ENT1</i> o <i>ENT2</i> , o en ambos.
<code>ENT1 /& ENT2</code>	<i>Entero</i>	El resultado de esta operación es el "o exclusivo" a nivel de bit de <i>ENT1</i> y <i>ENT2</i> . Es decir, hay un 1 en el resultado de cada posición de bits para la que hay un 1 en <i>ENT1</i> o <i>ENT2</i> , pero no en ambos.
<code>ENT1 && ENT2</code>	<i>Entero</i>	Genera el "y" a nivel de bit de los enteros <i>ENT1</i> y <i>ENT2</i> . Es decir, hay un 1 en el resultado de cada posición de bits para la que hay un 1 en <i>ENT1</i> y <i>ENT2</i> .
<code>ENT1 &&~~ ENT2</code>	<i>Entero</i>	Genera el "y" a nivel de bit de <i>ENT1</i> y el complemento a nivel de bit de <i>ENT2</i> . Es decir, hay un 1 en el resultado de cada posición de bits para la que hay un 1 en <i>ENT1</i> y un 0 en <i>ENT2</i> . Es lo mismo que <code>ENT1 && (~~ENT2)</code> y resulta útil para borrar bits de <i>ENT1</i> definidos en <i>ENT2</i> .
<code>ENT << N</code>	<i>Entero</i>	Genera el patrón de bits de <i>ENT1</i> desplazado a la izquierda <i>N</i> posiciones. Un valor negativo para <i>N</i> genera un desplazamiento a la derecha.
<code>ENT >> N</code>	<i>Entero</i>	Genera el patrón de bits de <i>ENT1</i> desplazado a la derecha <i>N</i> posiciones. Un valor negativo para <i>N</i> genera un desplazamiento a la izquierda.
<code>ENT1 &&=_0 ENT2</code>	<i>Booleana</i>	Equivalente a la expresión booleana <code>ENT1 && ENT2 != 0</code> , pero es más eficaz.
<code>ENT1 &&/=_0 ENT2</code>	<i>Booleana</i>	Equivalente a la expresión booleana <code>ENT1 && ENT2 == 0</code> , pero es más eficaz.
<code>integer_bitcount(ENT)</code>	<i>Entero</i>	Cuenta el número de bits 1 ó 0 de la representación en complemento a dos de <i>ENT</i> . Si <i>ENT</i> no es negativo, <i>N</i> es el número de bits 1. Si <i>ENT</i> es negativo, es el número de bits 0. Debido a la extensión del signo, hay un número infinito de bits 0 en un entero no negativo, o de bits 1 en un entero negativo. Siempre se da el caso de que <code>integer_bitcount(ENT) = integer_bitcount(-(ENT+1))</code> .

Tabla 23. Operaciones de enteros bit a bit de CLEM (continuación).

Función	Resultado	Descripción
integer_leastbit(ENT)	Entero	Devuelve la posición de bits N correspondiente al bit menos significativo definido en el entero ENT . N es la potencia más elevada de 2 por la que ENT tiene una división exacta.
integer_length(ENT)	Entero	Devuelve la longitud en bits de ENT como un entero en complemento a dos. Es decir, N es el entero más pequeño, tal que $ENT < (1 \ll N)$ si $ENT \geq 0$ $ENT \geq ((-1 \ll N)$ si $ENT < 0$. Si INT no es negativo, la representación de ENT como entero sin signo requiere un campo de al menos NN bits. También se requiere un mínimo de $N+1$ bits para representar ENT como un entero con signo, independientemente de su signo.
testbit(ENT, N)	Booleana	Comprueba el bit situado en la posición N del entero ENT y devuelve el estado del bit N como un valor booleano, que es verdadero para 1 y falso para 0.

Funciones aleatorias

Las siguientes funciones se utilizan para seleccionar elementos o generar números al azar.

Tabla 24. CLEMfunciones aleatorias.

Función	Resultado	Descripción
oneof(LISTA)	Cualquiera	Devuelve un elemento seleccionado al azar de $LISTA$. Los elementos de la lista se deben introducir como $[ELEMENT01, ELEMENT02, \dots, ELEMENTO_N]$. Recuerde que también se puede especificar una lista de nombres de campo.
random(NÚM)	Number	Devuelve un número aleatorio del mismo tipo (ENT o $REAL$) distribuido de modo uniforme, desde 1 a $NÚM$. Si utiliza un entero, sólo se devuelven enteros. Si utiliza un número real (decimal), se devuelven números reales (precisión decimal determinada por las opciones de la ruta). El número aleatorio más grande que devuelve la función puede ser igual a $NÚM$.
random0(NÚM)	Number	Este tiene las mismas propiedades que $random(NÚM)$, pero empezando desde 0. El número aleatorio más grande devuelto por la función nunca será igual a $NÚM$.

Funciones de cadena

En CLEM, puede realizar las siguientes operaciones con cadenas:

- Comparar cadenas
- Crear cadenas
- Acceder a caracteres

En CLEM, una cadena es cualquier secuencia de caracteres que se encuentre entre un par de comillas dobles ("comillas de cadena"). Los caracteres (CAR) pueden ser cualquier carácter alfanumérico simple. Se declaran en expresiones CLEM utilizando comillas inversas simples del tipo ``<carácter>``, como por ejemplo ``z``, ``A`` o ``2``. Los caracteres que se encuentran fuera de los límites o los índices negativos de una cadena darán como resultado un comportamiento indefinido.

Nota: Las comparaciones entre cadenas que utilizan y no utilizan retrotracción SQL pueden generar diferentes resultados donde existen espacios finales.

Tabla 25. CLEMfunciones de cadena.

Función	Resultado	Descripción
allbutfirst(N, CADENA)	Cadena	Devuelve una cadena, que es CADENA sin los N primeros caracteres.
allbutlast(N, CADENA)	Cadena	Devuelve una cadena, que es CADENA sin los últimos caracteres.
alphabefore(CADENA1, CADENA2)	Booleana	Se utiliza para comprobar el orden alfabético de las cadenas. Devuelve verdadero si CADENA1 precede a CADENA2.
endstring(LONGITUD, CADENA)	Cadena	Extrae los últimos N caracteres de la cadena especificada. Si la longitud de la cadena es menor o igual que la longitud especificada, no hay cambios.
hasendstring(CADENA, SUBCADENA)	Entero	Esta función es la misma que isendstring(SUBCADENA, CADENA).
hasmidstring(CADENA, SUBCADENA)	Entero	Esta función es la misma que ismidstring(SUBCADENA, CADENA) (subcadena incrustada).
hasstartstring(CADENA, SUBCADENA)	Entero	Esta función es la misma que isstartstring(SUBCADENA, CADENA).
hassubstring(CADENA, N, SUBCADENA)	Entero	Esta función es la misma que issubstring(SUBCADENA, N, CADENA) donde, de forma predeterminada, N tiene el valor 1.
count_substring(CADENA, SUBCADENA)	Entero	Devuelve el número de veces que aparece la subcadena especificada en la cadena. Por ejemplo: count_substring("foooo.txt", "oo") devuelve 3.
hassubstring(CADENA, SUBCADENA)	Entero	Esta función es la misma que issubstring(SUBCADENA, 1, CADENA) donde N tiene el valor 1 de forma predeterminada.
isalphacode(CAR)	Booleana	Devuelve verdadero si CAR es un carácter en la cadena especificada (normalmente un nombre de campo) cuyo código de carácter es una letra. De lo contrario, esta función devuelve 0. Por ejemplo, isalphacode(producir_núm(1)).
isendstring(SUBCADENA, CADENA)	Entero	Si la cadena CADENA finaliza con la subcadena SUBCADENA, esta función devuelve el subíndice entero de SUBCADENA en CADENA. En el resto de los casos, esta función devuelve 0.
islowercode(CAR)	Booleana	Devuelve verdadero si CAR es un carácter de una letra minúscula para la cadena especificada (normalmente un nombre de campo). En el resto de los casos, esta función devuelve 0. Por ejemplo, islowercode(``) y islowercode(nombre_país(2)) son ambas expresiones válidas.
ismidstring(SUBCADENA, CADENA)	Entero	Si SUBCADENA es una subcadena de CADENA pero no comienza en el primer carácter de CADENA ni termina en el último, esta función devuelve el subíndice donde dicha subcadena comienza. En el resto de los casos, esta función devuelve 0.

Tabla 25. CLEMfunciones de cadena (continuación).

Función	Resultado	Descripción
isnumbercode(CAR)	Booleana	Devuelve verdadero si CAR, para la cadena especificada (normalmente un nombre de campo), es un carácter cuyo código de carácter es un número. De lo contrario, esta función devuelve un valor de 0. Por ejemplo, isnumbercode(product_id(2)).
isstartstring(SUBCADENA, CADENA)	Entero	Si la cadena CADENA comienza con la subcadena SUBCADENA, esta función devuelve el subíndice 1. En el resto de los casos, esta función devuelve un valor de 0.
issubstring(SUBCADENA, N, CADENA)	Entero	Busca la cadena CADENA, a partir del carácter número N, para una subcadena igual a la cadena SUBCADENA. Si la encuentra, esta función devuelve el subíndice entero donde comienza el subcadena coincidente. En el resto de los casos, esta función devuelve 0. Si no se proporciona N, esta función toma el valor 1 de forma predeterminada.
issubstring(SUBCADENA, CADENA)	Entero	Busca la cadena CADENA, a partir del carácter número N, para una subcadena igual a la cadena SUBCADENA. Si la encuentra, esta función devuelve el subíndice entero donde comienza el subcadena coincidente. En el resto de los casos, esta función devuelve 0. Si no se proporciona N, esta función toma el valor 1 de forma predeterminada.
issubstring_count(SUBCADENA, N, CADENA)	Entero	Devuelve el índice de la instancia número N de la SUBCADENA en la CADENA especificada. Si hay menos de N instancias de SUBCADENA, se devuelve 0.
issubstring_lim(SUBCADENA, N, LÍMINICIAL, LÍMFINAL, CADENA)	Entero	Esta función es la misma que issubstring, pero la coincidencia se limita de manera que se inicia en el subíndice LÍMINICIAL (o antes) y finaliza en el subíndice LÍMFINAL (o antes). Las restricciones STARTLIM o ENDLIM se pueden desactivar proporcionando un valor de falso para cada argumento, por ejemplo, issubstring_lim(SUBSTRING, N, falso, falso, STRING) es lo mismo que issubstring.
isuppercode(CAR)	Booleana	Devuelve verdadero si CAR es un carácter de una letra mayúscula. En el resto de los casos, esta función devuelve 0. Por ejemplo, isuppercode(``) y isuppercode(nombre_país(2)) son ambas expresiones válidas.
last(CAR)	Cadena	Devuelve el último carácter CAR de CADENA (que debe tener una longitud de al menos un carácter).
length(CADENA)	Entero	Devuelve la longitud de la cadena CADENA; es decir el número de caracteres que hay en ella.

Tabla 25. CLEMfunciones de cadena (continuación).

Función	Resultado	Descripción
<p>locchar(CAR, N, CADENA)</p>	<p>Entero</p>	<p>Se utiliza para identificar la ubicación de los caracteres en los campos simbólicos. La función busca en la cadena CADENA el carácter CAR, comenzando la búsqueda en el carácter número N de la CADENA. Esta función devuelve un valor que indica la ubicación (comenzando por N) en la que se encuentra el carácter. Si no se encuentra el carácter, esta función devuelve 0. Si la función tiene un desplazamiento no válido (N) (por ejemplo, un desplazamiento que supera la longitud de la cadena), esta función devuelve \$null\$.</p> <p>Por ejemplo, locchar(`n`, 2, página_web) busca el carácter `n` en el campo con el nombre página_web, comenzando a partir del segundo carácter del valor del campo.</p> <p>Note: Asegúrese de utilizar comillas inversas simples para entrecomillar el carácter especificado.</p>
<p>locchar_back(CAR, N, CADENA)</p>	<p>Entero</p>	<p>Similar a locchar, con la diferencia de que la búsqueda se realiza hacia atrás, comenzando a partir del carácter número N. Por ejemplo, locchar_back(`n`, 9, página_web) realiza búsquedas en el campo página_web comenzando por el noveno carácter y yendo hacia atrás, hacia el inicio de la cadena. Si la función tiene un desplazamiento no válido (por ejemplo, un desplazamiento superior a la longitud de la cadena), esta función devuelve \$null\$. Lo ideal sería usar locchar_back junto con la función length(<campo>) para utilizar dinámicamente la longitud del valor actual del campo. Por ejemplo, locchar_back(`n`, (length(página_web)), página_web).</p>
<p>lowertoupper(CAR) lowertoupper(CADENA)</p>	<p>CAR o Cadena</p>	<p>La entrada puede ser una cadena o un carácter, y se utiliza en esta función para devolver un nuevo elemento del mismo tipo, con los caracteres en minúscula convertidos a sus equivalentes en mayúscula. Por ejemplo, lowertoupper(`a`), lowertoupper("Mi cadena") y lowertoupper(nombre_campo(2)) son todas expresiones válidas.</p>
<p>coincidencias</p>	<p>Booleana</p>	<p>Devuelve verdadero si una cadena coincide con un patrón determinado. El patrón debe ser una cadena literal, no puede ser un nombre de campo que contenga un patrón. Debe utilizarse un signo de cierre de interrogación (?) se puede incluir en el patrón para que coincida exactamente con un carácter, un asterisco (*) coincide con cero o más caracteres. Para que coincida con un asterisco o una interrogación literal (en lugar de utilizarlos como comodines), se puede utilizar una barra inclinada invertida (\) como carácter de escape.</p>

Tabla 25. CLEMfunciones de cadena (continuación).

Función	Resultado	Descripción
replace(SUBCADENA, NUEVASUBCADENA, CADENA)	Cadena	En la CADENA especificada, sustituye todas las instancias de SUBCADENA con NUEVASUBCADENA.
replicate (RECUENTO, CADENA)	Cadena	Devuelve una cadena que consta de la cadena original copiada el número específico de veces.
stripchar(CAR, CADENA)	Cadena	Permite eliminar determinados caracteres de una cadena o campo. Por ejemplo, puede utilizar esta función para eliminar de los datos símbolos adicionales (como notaciones de moneda) con el fin de obtener un nombre o número simple. Por ejemplo, la sintaxis stripchar(``, 'Coste') devuelve un nuevo campo con el signo de dólar eliminado en todos los valores. <i>Note:</i> Asegúrese de utilizar comillas inversas simples para entrecomillar el carácter especificado.
skipchar(CAR, N, CADENA)	Entero	Busca en la cadena CADENA cualquier carácter que no sea CAR, comenzando por el carácter número N. Esta función devuelve una subcadena de enteros que indica el punto en que se encuentra cada uno, o 0 si todos los caracteres a partir del Nson CAR. Si la función tiene un desplazamiento no válido (por ejemplo, un desplazamiento superior a la longitud de la cadena), esta función devuelve \$null\$. locchar se suele utilizar junto con las funciones skipchar para determinar el valor de N (el punto en que comenzar la búsqueda en la cadena). Por ejemplo, skipchar(`s`, (locchar(`s`, 1, "MiCadena")), "MiCadena").
skipchar_back(CAR, N, CADENA)	Entero	Similar a skipchar, con la diferencia de que la búsqueda se realiza hacia atrás , comenzando a partir del carácter número N.
startstring(LONGITUD, CADENA)	Cadena	Extrae los primeros N caracteres de la cadena especificada. Si la longitud de la cadena es menor o igual que la longitud especificada, no hay cambios.
strmember(CAR, CADENA)	Entero	Equivalente a locchar(CAR, 1, CADENA). Devuelve una subcadena de enteros que indica el punto en que CHAR aparece por primera vez o bien 0. Si la función tiene un desplazamiento no válido (por ejemplo, un desplazamiento superior a la longitud de la cadena), esta función devuelve \$null\$.
subscrs(N, CADENA)	CAR	Devuelve el carácter CAR número N de la cadena de entrada CADENA. Esta función también se puede escribir de forma abreviada como CADENA(N). Por ejemplo, lowertoupper("nombre"(1)) es una expresión válida.
substring(N, LON, CADENA)	Cadena	Devuelve una cadena SUBCADENA, que consiste en los caracteres LON de la cadena CADENA, a partir del carácter situado en el subíndice N.

Tabla 25. CLEMfunciones de cadena (continuación).

Función	Resultado	Descripción
substring_between(N1, N2, CADENA)	Cadena	Devuelve la subcadena de CADENA que comienza en el subíndice N1 y termina en el subíndice N2.
trim(CADENA)	Cadena	Extrae los espacios en blanco iniciales y finales de la cadena especificada.
trim_start(CADENA)	Cadena	Extrae los espacios en blanco iniciales de la cadena especificada.
trimend(CADENA)	Cadena	Extrae los espacios en blanco finales de la cadena especificada.
unicode_char(NÚM)	CAR	Devuelve el carácter con el valor Unicode NÚM.
unicode_value(CAR)	NUM	Devuelve el valor Unicode de CAR
uppertolower(CAR) uppertolower(CADENA)	CAR o Cadena	La entrada puede ser una cadena o un carácter y se utiliza en esta función para devolver un nuevo elemento del mismo tipo, con los caracteres en mayúscula convertidos a sus equivalentes en minúscula. <i>Note:</i> No olvide especificar las cadenas con comillas dobles y los caracteres con comillas inversas simples. Los nombres de campos simples deben especificarse sin comillas.

Funciones SoundEx

SoundEx es un método utilizado para buscar cadenas cuando se conoce el sonido pero no la ortografía exacta. Desarrollado en 1918, busca palabras con sonidos similares según ciertos supuestos fonéticos sobre cómo se pronuncian ciertas letras. Se puede utilizar para buscar nombres en una base de datos, por ejemplo, cuando la ortografía o la pronunciación de ciertos nombres puede variar. El algoritmo SoundEx básico se documenta en varios orígenes y, a pesar de las limitaciones conocidas (por ejemplo, combinaciones de letras iniciales como ph y f no coincidirán aunque el sonido sea el mismo), es compatible en cierto modo con la mayoría de las bases de datos.

Tabla 26. Funciones soundex de CLEM

Función	Resultado	Descripción
soundex(CADENA)	Entero	Devuelve el código SoundEx de cuatro caracteres para la CADENA especificada.
soundex_difference(CADENA1, CADENA2)	Entero	Devuelve un entero entre 0 y 4 que indica el número de caracteres que son iguales en la codificación SoundEx para las dos cadenas, donde 0 indica que no hay similitud y 4 indica una fuerte similitud o cadenas idénticas.

Funciones de fecha y hora

CLEM incluye un grupo de funciones para la gestión de campos con almacenamiento de fecha y hora de variables de cadena que representan fechas y horas. Los formatos de fecha y hora utilizados son específicos de cada ruta y se especifican en el cuadro de diálogo de propiedades de la ruta. Las funciones de fecha y hora analizan las cadenas de fecha y hora en función del formato seleccionado actualmente.

Cuando se especifica un año en una fecha que utiliza sólo dos dígitos (es decir, el siglo no se especifica), IBM SPSS Modeler utiliza el siglo predeterminado que se ha especificado en el cuadro de diálogo de propiedades de la ruta.

Tabla 27. Funciones de fecha y hora de CLEM.

Función	Resultado	Descripción
@TODAY	Cadena	Si selecciona Admitir fecha/mín. negativos en el cuadro de diálogo de propiedades de la ruta, esta función devuelve la fecha actual, como una cadena, con el formato de fecha actual. Si utiliza un formato de fecha de dos dígitos y no selecciona Admitir fecha/mín. negativos , esta función devuelve \$null\$ en el servidor actual.
to_time(ELEMENTO)	Hora	Convierte el almacenamiento del campo especificado en una hora.
to_date(ELEMENTO)	Fecha	Convierte el almacenamiento del campo especificado en una fecha.
to_timestamp(ELEMENTO)	Marca de tiempo	Convierte el almacenamiento del campo especificado en una marca de tiempo.
to_datetime(ELEMENTO)	Momento_fecha	Convierte el almacenamiento del campo especificado en una fecha, hora o marca de tiempo.
datetime_date(ELEMENTO)	Fecha	Devuelve el valor de fecha de un <i>número, cadena o marca de tiempo</i> . Tenga en cuenta que es la única función que le permite volver a convertir un número (en segundos) en una fecha. Si ELEMENTO es una cadena, crea una fecha analizando una cadena en el formato de fecha actual. El formato de fecha especificado en el cuadro de diálogo de propiedades de la ruta debe ser correcto para que esta función se ejecute correctamente. Si ELEMENTO es un número, se interpreta como un número de segundos desde la fecha de base (o época). Las fracciones de un día se truncan. Si ELEMENTO es una marca de tiempo, la parte de la fecha de la marca de tiempo se devuelve. Si ELEMENTO es una fecha, se devuelve sin modificar.
date_before(FECHA1, FECHA2)	Booleana	Devuelve un valor de verdadero si FECHA1 representa una fecha o marca de tiempo anterior a la representada por FECHA2. En el resto de los casos, esta función devuelve 0.
date_days_difference(FECHA1, FECHA2)	Entero	Devuelve el tiempo en días desde la fecha o marca de tiempo representada por FECHA1 hasta la fecha representada por FECHA2, como un entero. Si FECHA2 es anterior a FECHA1, esta función devuelve un número negativo.
date_in_days(FECHA)	Entero	Devuelve el tiempo en días desde la fecha de línea base hasta la fecha o marca de tiempo representada por FECHA, como un entero. Si FECHA es anterior a la fecha de línea base, esta función devuelve un número negativo. Debe incluir una fecha válida para que el cálculo se realice correctamente. Por ejemplo, no debe especificar 29 de febrero de 2001 como fecha. Dado que 2001 no es un año bisiesto, esta fecha no existe.
date_in_months(FECHA)	Real	Devuelve el tiempo en meses desde la fecha de línea base hasta la fecha o marca de tiempo representada por FECHA, como un número real. Esta cifra es aproximada; se basa en un mes de 30.4375 días. Si FECHA es anterior a la fecha de línea base, esta función devuelve un número negativo. Debe incluir una fecha válida para que el cálculo se realice correctamente. Por ejemplo, no debe especificar 29 de febrero de 2001 como fecha. Dado que 2001 no es un año bisiesto, esta fecha no existe.

Tabla 27. Funciones de fecha y hora de CLEM (continuación).

Función	Resultado	Descripción
date_in_weeks(FECHA)	Real	Devuelve el tiempo en semanas desde la fecha de línea base hasta la fecha o marca de tiempo representada por <i>FECHA</i> , como un número real. Éste se basa en una semana de 7,0 días. Si <i>FECHA</i> es anterior a la fecha de línea base, esta función devuelve un número negativo. Debe incluir una fecha válida para que el cálculo se realice correctamente. Por ejemplo, no debe especificar 29 de febrero de 2001 como fecha. Dado que 2001 no es un año bisiesto, esta fecha no existe.
date_in_years(FECHA)	Real	Devuelve el tiempo en años desde la fecha de línea base hasta la fecha o marca de tiempo representada por <i>FECHA</i> , como un número real. Esta cifra es aproximada; se basa en un año de 365.25 días. Si <i>FECHA</i> es anterior a la fecha de línea base, esta función devuelve un número negativo. Debe incluir una fecha válida para que el cálculo se realice correctamente. Por ejemplo, no debe especificar 29 de febrero de 2001 como fecha. Dado que 2001 no es un año bisiesto, esta fecha no existe.
date_months_difference (FECHA1, FECHA2)	Real	Devuelve el tiempo en meses desde la fecha o marca de tiempo representada por <i>FECHA1</i> hasta la fecha representada por <i>FECHA2</i> , como un número real. Esta cifra es aproximada; se basa en un mes de 30.4375 días. Si <i>FECHA2</i> es anterior a <i>FECHA1</i> , esta función devuelve un número negativo.
datetime_date(AÑO, MES, DÍA)	Fecha	Crea un valor de fecha para los valores <i>AÑO</i> , <i>MES</i> y <i>DÍA</i> especificados. Los argumentos deben ser enteros.
datetime_day(FECHA)	Entero	Devuelve el día del mes de una <i>FECHA</i> o marca de tiempo especificada. El resultado es un entero comprendido entre 1 y 31.
datetime_day_name(DÍA)	Cadena	Devuelve el nombre completo del <i>DÍA</i> especificado. El argumento debe ser un entero comprendido entre 1 (domingo) y 7 (sábado).
datetime_hour(TIEMPO)	Entero	Devuelve la hora de una <i>HORA</i> o marca de tiempo. El resultado es un entero comprendido entre 0 y 23.
datetime_in_seconds(TIEMPO)	Real	Devuelve la segunda parte almacenada en <i>TIEMPO</i> .
datetime_in_seconds(FECHA), datetime_in_seconds(FECHAHORA)	Real	Devuelve el número acumulado, convertido en segundos, desde la diferencia entre la <i>FECHA</i> o <i>MOMENTO DE FECHA</i> actual y la fecha de línea base (01-01-1900).
datetime_minute(TIEMPO)	Entero	Devuelve el minuto de una <i>HORA</i> o marca de tiempo. El resultado es un entero comprendido entre 0 y 59.
datetime_month(FECHA)	Entero	Devuelve el mes de una <i>FECHA</i> o marca de tiempo. El resultado es un entero comprendido entre 1 y 12.
datetime_month_name (MES)	Cadena	Devuelve el nombre completo del <i>MES</i> especificado. El argumento debe ser un entero comprendido entre 1 y 12.
datetime_now	Marca de tiempo	Devuelve la hora actual como una marca de tiempo.
datetime_second(TIEMPO)	Entero	Devuelve el segundo de una <i>HORA</i> o marca de tiempo. El resultado es un entero comprendido entre 0 y 59.
datetime_day_short_name (DÍA)	Cadena	Devuelve el nombre abreviado del <i>DÍA</i> especificado. El argumento debe ser un entero comprendido entre 1 (domingo) y 7 (sábado).

Tabla 27. Funciones de fecha y hora de CLEM (continuación).

Función	Resultado	Descripción
datetime_month_short_name (MES)	Cadena	Devuelve el nombre abreviado del MES especificado. El argumento debe ser un entero comprendido entre 1 y 12.
datetime_time(HORA, MINUTO, SEGUNDO)	Hora	Devuelve el valor de tiempo para los valores HORA, MINUTO y SEGUNDO especificados. Los argumentos deben ser enteros.
datetime_time(ELEMENTO)	Hora	Devuelve el valor de tiempo del ELEMENTO especificado.
datetime_timestamp(AÑO, MES, DÍA, HORA, MINUTO, SEGUNDO)	Marca de tiempo	Devuelve el valor de marca de tiempo para los valores AÑO, MES, DÍA, HORA, MINUTO y SEGUNDO especificados.
datetime_timestamp(FECHA, TIEMPO)	Marca de tiempo	Devuelve el valor de marca de tiempo para los valores FECHA y HORA especificados.
datetime_timestamp (NÚM)	Marca de tiempo	Devuelve el valor de marca de tiempo del número especificado de segundos.
datetime_weekday(FECHA)	Entero	Devuelve el día de la semana a partir de la FECHA o marca de tiempo especificada.
datetime_year(FECHA)	Entero	Devuelve el año a partir de una FECHA o marca de tiempo. El resultado es un entero como, por ejemplo, 2002.
date_weeks_difference (FECHA1, FECHA2)	Real	Devuelve el tiempo en semanas desde la fecha o marca de tiempo representada por FECHA1 hasta la fecha representada por FECHA2, como un número real. Éste se basa en una semana de 7,0 días. Si FECHA2 es anterior a FECHA1, esta función devuelve un número negativo.
date_years_difference (FECHA1, FECHA2)	Real	Devuelve el tiempo en años desde la fecha o marca de tiempo representada por FECHA1 hasta la fecha representada por FECHA2, como un número real. Esta cifra es aproximada; se basa en un año de 365.25 días. Si FECHA2 es anterior a FECHA1, esta función devuelve un número negativo.
time_before(TIEMPO1, TIEMPO2)	Booleana	Devuelve un valor verdadero si HORA1 representa una hora o marca de tiempo anterior a la representada por HORA2. En el resto de los casos, esta función devuelve 0.
time_hours_difference (TIEMPO1, TIEMPO2)	Real	Devuelve el número de horas de diferencia que hay entre las horas o marcas de tiempo representadas por HORA1 y HORA2, como un número real. Si selecciona Admitir fecha/mín. negativos en el cuadro de diálogo de propiedades de la ruta, se toma un valor superior de HORA1 para hacer referencia al día anterior. Si no selecciona la opción de admisión de negativos, un valor superior de HORA1 hará que el valor devuelto sea negativo.
time_in_hours(TIEMPO)	Real	Devuelve el número de horas representado por HORA, como un número real. Por ejemplo, con el formato de hora HHMM, la expresión time_in_hours('0130') se evalúa como 1,5. HORA puede representar una hora o una marca de tiempo.
time_in_mins(TIEMPO)	Real	Devuelve el número de minutos representado por HORA, como un número real. HORA puede representar una hora o una marca de tiempo.
time_in_secs(TIEMPO)	Entero	Devuelve el número de segundos representado por HORA, como un entero. HORA puede representar una hora o una marca de tiempo.

Tabla 27. Funciones de fecha y hora de CLEM (continuación).

Función	Resultado	Descripción
<code>time_mins_difference(TIEMP01, TIEMP02)</code>	<i>Real</i>	Devuelve el número de minutos de diferencia que hay entre las horas o marcas de tiempo representadas por <i>HORA1</i> y <i>HORA2</i> , como un número real. Si selecciona Admitir fecha/mín. negativos en el cuadro de diálogo de propiedades de la ruta, se toma un valor superior de <i>HORA1</i> para hacer referencia al día anterior (o la hora anterior, si sólo se especifican minutos y segundos en el formato actual). Si no selecciona la opción de admisión de negativos, un valor superior de <i>HORA1</i> hará que el valor devuelto sea negativo.
<code>time_secs_difference(TIEMP01, TIEMP02)</code>	<i>Entero</i>	Devuelve el número de segundos de diferencia que hay entre las horas o marcas de tiempo representadas por <i>HORA1</i> y <i>HORA2</i> , como un entero. Si selecciona Admitir fecha/mín. negativos en el cuadro de diálogo de propiedades de la ruta, se toma un valor superior de <i>HORA1</i> para hacer referencia al día anterior (o la hora anterior, si sólo se especifican minutos y segundos en el formato actual). Si no selecciona la opción de admisión de negativos, un valor superior de <i>HORA1</i> hará que el valor devuelto sea negativo.

Conversión de valores de fecha y hora

Tenga en cuenta que las funciones de conversión (y otras muchas funciones que requieren un tipo específico de entrada, como un valor de fecha u hora) dependen de los formatos actuales especificados en el cuadro de diálogo de opciones de ruta. Por ejemplo, si tiene un campo denominado *FECHA* que se guarda como una cadena con los valores *Ene 2003*, *Feb 2003*, etcétera, puede convertirla en campos de almacenamiento de la forma siguiente:

```
to_date(FECHA)
```

Para que esta conversión funcione, seleccione el formato de fecha coincidente **MES AAAA** como formato de fecha predeterminado de la ruta.

Para ver un ejemplo que convierte valores de cadena a fecha utilizando un nodo Rellenar, consulte la ruta *broadband_create_models.str*, en la carpeta *\Demos* en la subcarpeta *streams*.

Fechas almacenadas como números. Tenga en cuenta que *FECHA* del ejemplo anterior es el nombre de un campo, mientras que `to_date` es una función de CLEM. Si tiene datos guardados como números, puede convertirlos utilizando la función `datetime_date`, donde el número se interpreta como un número de segundos desde la fecha de base (o época).

```
datetime_date(FECHA)
```

Al convertir una fecha a un número de segundos (y viceversa), puede ejecutar operaciones como calcular la fecha actual más o menos un número fijo de días, por ejemplo:

```
datetime_date((date_in_days(FECHA)-7)*60*60*24)
```

Funciones de secuencia

Para algunas operaciones, la secuencia de eventos tiene su importancia. La aplicación permite trabajar con las siguientes secuencias de registros:

- Secuencias y series temporales
- Funciones de secuencia
- Indización de registros

- Media, suma y comparación de valores
- Control de cambio - diferenciación
- @SINCE
- Valores de desplazamiento
- Recursos de secuencias adicionales

Para muchas aplicaciones, cada registro que pasa por una ruta se puede considerar como un caso individual, independiente de todos los demás. En estas situaciones, el orden de los registros suele no ser importante.

Sin embargo, para algunos tipos de problemas, la secuencia de registros es de gran importancia. Suelen ser situaciones de series temporales, en las que la secuencia de registros representa una secuencia de eventos o instancias ordenada. Cada registro representa una instantánea de un momento concreto; sin embargo, gran parte de la información más importante puede no estar contenida en valores instantáneos, sino en el modo en que tales valores cambian y se comportan durante un período de tiempo.

Por supuesto, el parámetro relevante puede no ser temporal. Por ejemplo, los registros podrían representar análisis realizados a distintas distancias a lo largo de una línea, pero se aplicarían los mismos principios.

Las funciones de secuencia y especiales se reconocen de inmediato por las siguientes características:

- Todas tienen el prefijo @.
- Sus nombres se proporcionan en mayúsculas.

Las funciones de secuencia pueden hacer referencia al registro que está siendo procesado actualmente por un nodo, los registros que ya han pasado por un nodo e, incluso, a los registros que aún tienen que pasar por un nodo. Las funciones de secuencia se pueden mezclar libremente con otros componentes de expresiones CLEM, aunque algunas tienen restricciones en cuanto a qué pueden utilizar como sus argumentos.

Ejemplos

Puede que le resulte útil saber cuánto tiempo ha pasado desde que se produjo un determinado evento o una condición fue verdadera. Utilice la función @SINCE; por ejemplo:

```
@SINCE(Ingresos > Pagos)
```

Esta función devuelve el desplazamiento del último registro donde esta condición fue verdadera; es decir, el número de registros anteriores a éste en el que la condición fue verdadera. Si la condición nunca ha sido verdadera, @SINCE devuelve @INDEX + 1.

Puede que en algunas ocasiones desee hacer referencia a un valor del registro actual en la expresión utilizada por @SINCE. Puede hacerlo mediante la función @THIS, que especifica que un nombre de campo siempre se aplique al registro actual. Para buscar el desplazamiento del último registro que tuvo un valor de campo concentración más de dos veces que el registro actual, podría utilizar:

```
@SINCE(Concentración > 2 * @THIS(Concentración))
```

En algunos casos, la condición especificada para @SINCE es verdadera en el registro actual por definición; por ejemplo,

```
@SINCE(ID == @THIS(ID))
```

Por este motivo, @SINCE no evalúa su condición para el registro actual. Utilice una función similar, @SINCE0, si desea evaluar la condición para el registro actual además de los anteriores; si la condición es verdadera en el registro actual, @SINCE0 devuelve 0.

Tabla 28. CLEMfunciones de secuencia.

Función	Resultado	Descripción
MEAN(CAMPO)	<i>Real</i>	Devuelve el valor promediado de los valores para el CAMPO o CAMPOS especificados.
@MEAN(CAMPO, EXPR)	<i>Real</i>	Devuelve el valor promediado de los valores para CAMPO en los últimos registros de EXPR recibidos por el nodo actual, incluyendo el registro actual. CAMPO debe ser el nombre de un campo numérico. EXPR puede ser cualquier expresión que se evalúe como un entero mayor que 0. Si EXPR se omite, o si supera el número de registros recibidos hasta el momento, se devuelve el promedio de todos los registros recibidos hasta el momento.
@MEAN(CAMPO, EXPR, ENT)	<i>Real</i>	Devuelve el valor promediado de los valores para CAMPO en los últimos registros de EXPR recibidos por el nodo actual, incluyendo el registro actual. CAMPO debe ser el nombre de un campo numérico. EXPR puede ser cualquier expresión que se evalúe como un entero mayor que 0. Si EXPR se omite, o si supera el número de registros recibidos hasta el momento, se devuelve el promedio de todos los registros recibidos hasta el momento. ENT especifica el número máximo de valores que buscar hacia atrás, lo que es mucho más eficaz que utilizar tan sólo dos argumentos.
@DIFF1(CAMPO)	<i>Real</i>	Devuelve el primer diferencial de CAMPO1. El formato de un único argumento devuelve simplemente la diferencia entre el valor actual y el valor anterior del campo. Devuelve 0 si los registros anteriores relevantes no existen.
@DIFF1(CAMPO1, CAMPO2)	<i>Real</i>	El formato de dos argumentos proporciona el primer diferencial de CAMPO1 respecto a CAMPO2. Devuelve 0 si los registros anteriores relevantes no existen.
@DIFF2(CAMPO)	<i>Real</i>	Devuelve el segundo diferencial de CAMPO1. El formato de un único argumento devuelve simplemente la diferencia entre el valor actual y el valor anterior del campo. Devuelve 0 si los registros anteriores relevantes no existen.
@DIFF2(CAMPO1, CAMPO2)	<i>Real</i>	El formato de dos argumentos proporciona el primer diferencial de CAMPO1 respecto a CAMPO2. Devuelve 0 si los registros anteriores relevantes no existen.
@INDEX	<i>Entero</i>	Devuelve el índice del registro actual. Se asignan los índices a los registros a medida que llegan al nodo actual. Al primer registro se le asigna el índice 1, y el índice se va incrementando en 1 para cada registro posterior.
@LAST_NON_BLANK(CAMPO)	<i>Cualquiera</i>	Devuelve el último valor de CAMPO que no está vacío, tal y como se define en un nodo Tipo o de origen anterior de la ruta. Si no hay ningún valor no vacío para CAMPO en los registros leídos hasta el momento, se devuelve \$null\$. Recuerde que los valores vacíos, también llamados valores perdidos por el usuario, se pueden definir por separado para cada campo.
@MAX(CAMPO)	<i>Number</i>	Devuelve el valor máximo del CAMPO especificado.
@MAX(CAMPO, EXPR)	<i>Number</i>	Devuelve el valor máximo de CAMPO en los últimos registros de EXPR recibidos hasta el momento, incluyendo el registro actual. CAMPO debe ser el nombre de un campo numérico. EXPR puede ser cualquier expresión que se evalúe como un entero mayor que 0.

Tabla 28. CLEMfunciones de secuencia (continuación).

Función	Resultado	Descripción
@MAX(CAMPO, EXPR, ENT)	Number	Devuelve el valor máximo de CAMPO en los últimos registros de EXPR recibidos hasta el momento, incluyendo el registro actual. CAMPO debe ser el nombre de un campo numérico. EXPR puede ser cualquier expresión que se evalúe como un entero mayor que 0. Si EXPR se omite, o si supera el número de registros recibidos hasta el momento, se devuelve el valor máximo de todos los registros recibidos hasta el momento. ENT especifica el número máximo de valores que buscar hacia atrás, lo que es mucho más eficaz que utilizar tan sólo dos argumentos.
@MIN(CAMPO)	Number	Devuelve el valor mínimo del CAMPO especificado.
@MIN(CAMPO, EXPR)	Number	Devuelve el valor mínimo del CAMPO en los últimos registros de EXPR recibidos hasta el momento, incluyendo el registro actual. CAMPO debe ser el nombre de un campo numérico. EXPR puede ser cualquier expresión que se evalúe como un entero mayor que 0.
@MIN(CAMPO, EXPR, ENT)	Number	Devuelve el valor mínimo del CAMPO en los últimos registros de EXPR recibidos hasta el momento, incluyendo el registro actual. CAMPO debe ser el nombre de un campo numérico. EXPR puede ser cualquier expresión que se evalúe como un entero mayor que 0. Si EXPR se omite, o si supera el número de registros recibidos hasta el momento, se devuelve el valor mínimo de todos los registros recibidos hasta el momento. ENT especifica el número máximo de valores que buscar hacia atrás, lo que es mucho más eficaz que utilizar tan sólo dos argumentos.
@OFFSET(CAMPO, EXPR)	Cualquiera	Devuelve el valor de CAMPO en el desplazamiento de registro respecto al registro actual con el valor EXPR. Un positivo hace referencia a un registro que ya ha pasado, mientras que uno negativo especifica una "lectura anticipada" de un registro que aún tiene que llegar. Por ejemplo, @OFFSET(Status, 1) devuelve el valor del campo Status en el registro anterior, mientras que @OFFSET(Status, -4) hace una "lectura anticipada" de cuatro registros de la secuencia (es decir, en los registros que aún no han pasado por este nodo) para obtener el valor. Tenga en cuenta que un desplazamiento negativo (lectura anticipada) se debe especificar como una constante. En desplazamientos positivos únicamente, EXPR también debe ser una expresión CLEM arbitraria, que se evalúa para que el registro actual proporcione el desplazamiento. En este caso, la versión de tres argumentos de esta función debería mejorar el rendimiento (ver función siguiente). Si la expresión devuelve cualquier cosa que no sea un entero no negativo, se producirá un error, es decir, no es legal haber calculado desplazamientos de lectura anticipada. <i>Note:</i> Una función @OFFSET autorreferida no puede utilizar lectura anticipada literal. Por ejemplo, en un nodo Rellenar, no puede sustituir el valor de campo1 utilizando una expresión como @OFFSET(campo1, -2).

Tabla 28. CLEMfunciones de secuencia (continuación).

Función	Resultado	Descripción
@OFFSET(CAMPO, EXPR, ENT)	Cualquiera	Realiza la misma operación que la función @OFFSET con la adición de un tercer argumento, <i>ENT</i> , que especifica el número máximo de valores que buscar hacia atrás. En casos en los que el desplazamiento se calcula a partir de una expresión, este tercer argumento debería mejorar el rendimiento. Por ejemplo, en una expresión como @OFFSET(Foo, Mes, 12), el sistema sabe que debe mantener sólo los últimos doce valores de Foo; de lo contrario, tiene que almacenar todos los valores, por si acaso. Para los desplazamientos fijos, incluyendo los desplazamientos de "lectura anticipada" negativos, que deben ser fijos, el tercer argumento no es importante, y debe utilizarse la versión de dos argumentos de esta función. Consulte también la nota acerca de las funciones autorreferidas en la versión de dos argumentos descritas anteriormente.
@SDEV(CAMPO)	Real	Devuelve la desviación estándar de los valores para el CAMPO o CAMPOS especificados.
@SDEV(CAMPO, EXPR)	Real	Devuelve la desviación estándar de los valores para CAMPO en los últimos registros de EXPR recibidos por el nodo actual, incluyendo el registro actual. CAMPO debe ser el nombre de un campo numérico. EXPR puede ser cualquier expresión que se evalúe como un entero mayor que 0. Si EXPR se omite, o si supera el número de registros recibidos hasta el momento, se devuelve la desviación estándar de todos los registros recibidos hasta el momento.
@SDEV(CAMPO, EXPR, ENT)	Real	Devuelve la desviación estándar de los valores para CAMPO en los últimos registros de EXPR recibidos por el nodo actual, incluyendo el registro actual. CAMPO debe ser el nombre de un campo numérico. EXPR puede ser cualquier expresión que se evalúe como un entero mayor que 0. Si EXPR se omite, o si supera el número de registros recibidos hasta el momento, se devuelve la desviación estándar de todos los registros recibidos hasta el momento. ENT especifica el número máximo de valores que buscar hacia atrás, lo que es mucho más eficaz que utilizar tan sólo dos argumentos.
@SINCE(EXPR)	Cualquiera	Devuelve el número de registros que se han aceptado ya que EXPR, una expresión CLEM arbitraria, era verdadera.
@SINCE(EXPR, ENT)	Cualquiera	La adición del segundo argumento, <i>ENT</i> , especifica el número máximo de registros a buscar hacia atrás. Si EXPR nunca ha sido verdadera, ENT es @INDEX+1.
@SINCE0(EXPR)	Cualquiera	Considera el registro actual, mientras que @SINCE no lo hace; @SINCE0 devuelve 0 si EXPR es verdadera para el registro actual.
@SINCE0(EXPR, ENT)	Cualquiera	La adición del segundo argumento, <i>ENT</i> , especifica el número máximo de registros a buscar hacia atrás.
@SUM(CAMPO)	Number	Devuelve la suma de los valores del CAMPO o CAMPOS especificados.

Tabla 28. CLEMfunciones de secuencia (continuación).

Función	Resultado	Descripción
@SUM(CAMPO, EXPR)	Number	Devuelve la suma de los valores de CAMPO en los últimos EXPR registros recibidos por el nodo actual, incluyendo el registro actual. CAMPO debe ser el nombre de un campo numérico. EXPR puede ser cualquier expresión que se evalúe como un entero mayor que 0. Si se omite EXPR, o si supera el número de registros recibidos hasta el momento, se devuelve la suma de todos los registros recibidos hasta el momento.
@SUM(CAMPO, EXPR, ENT)	Number	Devuelve la suma de los valores de CAMPO en los últimos EXPR registros recibidos por el nodo actual, incluyendo el registro actual. CAMPO debe ser el nombre de un campo numérico. EXPR puede ser cualquier expresión que se evalúe como un entero mayor que 0. Si se omite EXPR, o si supera el número de registros recibidos hasta el momento, se devuelve la suma de todos los registros recibidos hasta el momento. ENT especifica el número máximo de valores que buscar hacia atrás, lo que es mucho más eficaz que utilizar tan sólo dos argumentos.
@THIS(CAMPO)	Cualquiera	Devuelve el valor del campo con el nombre CAMPO en el registro actual. Sólo se utiliza en las expresiones @SINCE.

Funciones globales

Las funciones @MEAN,@SUM, @MIN, @MAX y @SDEV funcionan, como máximo, en todos los registros leídos hasta el registro actual, éste incluido. Sin embargo, en algunos casos resulta útil poder saber cómo los valores del registro actual se comparan con los valores considerados en todo el conjunto de datos. Utilizando un nodo Val. globales para generar valores en todo el conjunto de datos, puede acceder a los mismos en una expresión CLEM mediante las funciones globales.

Por ejemplo:

@GLOBAL_MAX(Edad)

devuelve al valor más alto de Edad en el conjunto de datos, mientras que la expresión
 $(\text{Valor} - @GLOBAL_MEAN(\text{Valor})) / @GLOBAL_SDEV(\text{Valor})$

expresa la diferencia entre el Valor de este registro y la media global como un número de desviaciones estándar. Puede utilizar valores globales sólo una vez que éstos han sido calculados por un nodo Val. globales. Todos los valores globales actuales se pueden cancelar pulsando en el botón **Borrar valores globales** de la pestaña Valores globales del cuadro de diálogo de propiedades de la ruta.

Tabla 29. CLEMfunciones globales.

Función	Resultado	Descripción
@GLOBAL_MAX(CAMPO)	Number	Devuelve el valor máximo de CAMPO en todo el conjunto de datos, según lo haya generado previamente un nodo Val. globales. CAMPO debe ser el nombre de un campo numérico. Si no se ha establecido el valor global correspondiente, se produce un error.
@GLOBAL_MIN(CAMPO)	Number	Devuelve el valor mínimo de CAMPO en todo el conjunto de datos, según lo haya generado previamente un nodo Val. globales. CAMPO debe ser el nombre de un campo numérico. Si no se ha establecido el valor global correspondiente, se produce un error.

Tabla 29. CLEMfunciones globales (continuación).

Función	Resultado	Descripción
@GLOBAL_SDEV(CAMPO)	Number	Devuelve el la desviación estándar de los valores de CAMPO en todo el conjunto de datos, según lo haya generado previamente un nodo Val. globales. CAMPO debe ser el nombre de un campo numérico. Si no se ha establecido el valor global correspondiente, se produce un error.
@GLOBAL_MEAN(CAMPO)	Number	Devuelve el valor promediado de los valores de CAMPO en todo el conjunto de datos, según lo haya generado previamente un nodo Val. globales. CAMPO debe ser el nombre de un campo numérico. Si no se ha establecido el valor global correspondiente, se produce un error.
@GLOBAL_SUM(CAMPO)	Number	Devuelve la suma de los valores de CAMPO en todo el conjunto de datos, según lo haya generado previamente un nodo Val. globales. CAMPO debe ser el nombre de un campo numérico. Si no se ha establecido el valor global correspondiente, se produce un error.

Funciones para gestionar los valores vacíos y nulos

Con CLEM, puede especificar que ciertos valores de un campo se consideren "vacíos" o perdidos. Las siguientes funciones trabajan con valores o elementos vacíos.

Tabla 30. Funciones de valores vacíos y nulos de CLEM.

Función	Resultado	Descripción
@BLANK(CAMPO)	Booleana	Devuelve verdadero para todos los registros cuyos valores están vacíos de acuerdo con el conjunto de reglas de tratamiento de los elementos vacíos en un nodo Tipo o de origen (pestaña Tipos) anterior de la ruta.
@LAST_NON_BLANK(CAMPO)	Cualquiera	Devuelve el último valor de CAMPO que no está vacío, tal y como se define en un nodo Tipo o de origen anterior de la ruta. Si no hay ningún valor no vacío para CAMPO en los registros leídos hasta el momento, se devuelve \$null\$. Recuerde que los valores vacíos, también llamados valores perdidos por el usuario, se pueden definir por separado para cada campo.
@NULL(CAMPO)	Booleana	Devuelve verdadero si el valor de CAMPO es el valor \$null\$ que falta del sistema. Devuelve falso para todos los demás valores, incluyendo los vacíos definidos por el usuario. Si desea comprobar ambos casos, utilice @BLANK(CAMPO) y @NULL(CAMPO).
undef	Cualquiera	Se utiliza generalmente en CLEM para introducir un valor \$null\$; por ejemplo, para rellenar valores vacíos con nulos en el nodo Rellenar.

Los campos vacíos se pueden "rellenar" con el nodo Rellenar. En los nodos Rellenar y Derivar (sólo modo múltiple), la función CLEM especial @FIELD hace referencia al campo o campos actuales que están siendo examinados.

Campos especiales

Las funciones especiales se utilizan para denotar los campos específicos que están siendo examinados o para generar una lista de campos como entrada. Por ejemplo, cuando se derivan varios campos al mismo

tiempo, debe utilizar @FIELD para denotar "realizar esta acción de derivación en los campos seleccionados". Con la expresión log(@FIELD) se deriva un nuevo campo de log para cada campo seleccionado.

Tabla 31. Campos especiales de CLEM.

Función	Resultado	Descripción
@FIELD	<i>Cualquiera</i>	Realiza una acción en todos los campos especificados en el contexto de la expresión.
@TARGET	<i>Cualquiera</i>	Cuando se utiliza una expresión CLEM en una función de análisis definida por el usuario, @TARGET representa el campo objetivo o "valor correcto" de la pareja objetivo/predicho que se está analizando. Esta función se suele utilizar en un nodo Análisis.
@PREDICTED	<i>Cualquiera</i>	Cuando se utiliza una expresión CLEM en una función de análisis definida por el usuario, @PREDICTED representa el valor predicho de la pareja objetivo/predicho que se está analizando. Esta función se suele utilizar en un nodo Análisis.
@PARTITION_FIELD	<i>Cualquiera</i>	Sustituye el nombre del campo de partición actual.
@TRAINING_PARTITION	<i>Cualquiera</i>	Devuelve el valor de la partición de entrenamiento actual. Por ejemplo, para seleccionar registros mediante un nodo Seleccionar, utilice la expresión CLEM:@PARTITION_FIELD = @TRAINING_PARTITION; de este modo se garantiza que el nodo Seleccionar funcione siempre independientemente de los valores que se utilicen para representar cada partición de los datos.
@TESTING_PARTITION	<i>Cualquiera</i>	Devuelve el valor de la partición de comprobación actual.
@VALIDATION_PARTITION	<i>Cualquiera</i>	Devuelve el valor de la partición de validación actual.
@FIELDS_BETWEEN(inicio, fin)	<i>Cualquiera</i>	Devuelve la lista de nombres de campo entre los campos de inicio y fin especificados (inclusive) según el orden natural (es decir, de inserción) de los campos en los datos.

Tabla 31. Campos especiales de CLEM (continuación).

Función	Resultado	Descripción
@FIELDS_MATCHING(patrn)	<i>Cualquiera</i>	Devuelve una lista de nombres de campo que coinciden un patrón especificado. Debe utilizarse un signo de cierre de interrogación (?) se puede incluir en el patrón para que coincida exactamente con un carácter, un asterisco (*) coincide con cero o más caracteres. Para que coincida con un asterisco o una interrogación literal (en lugar de utilizarlos como comodines), se puede utilizar una barra inclinada invertida (\) como carácter de escape.
@MULTI_RESPONSE_SET	<i>Cualquiera</i>	Devuelve la lista de campos del conjunto de respuestas múltiples nombrado.

Capítulo 9. Uso de IBM SPSS Modeler con un repositorio

Acerca de IBM SPSS Collaboration and Deployment Services Repository

IBM SPSS Modeler puede utilizarse junto con un repositorio IBM SPSS Collaboration and Deployment Services, lo que permite administrar el ciclo vital de modelos de minería de datos y los objetos predictivos relacionados, así como utilizar estas aplicaciones, herramientas y soluciones empresariales. Entre los objetos IBM SPSS Modeler que se pueden compartir de este modo figuran rutas, nodos, resultados de rutas, escenarios, proyectos y modelos. Los objetos se almacenan en un repositorio central, desde donde se pueden compartir con otras aplicaciones y supervisar usando versiones extendidas, metadatos y capacidades de búsqueda.

Note: Es necesario disponer de una licencia independiente para acceder a un repositorio IBM SPSS Collaboration and Deployment Services. Si desea obtener más información, consulte <http://www.ibm.com/software/analytics/spss/products/deployment/cds/>

Antes de poder usar IBM SPSS Modeler con el repositorio, debe instalar un adaptador en el host del repositorio. Sin este adaptador, es posible que aparezcan los siguientes mensajes cuando intente acceder a los objetos del repositorio de algunos nodos o modelos de IBM SPSS Modeler:

Es posible que necesite actualizar el repositorio para que sea compatible con los nuevos tipos de nodo, modelo y resultados.

Si desea instrucciones sobre cómo instalar el adaptador, consulte la guía *IBM SPSS Modeler Instalación del despliegue de*, disponible en el DVD de despliegue de IBM SPSS Modeler. Se proporcionan detalles sobre cómo acceder a objetos de repositorio de IBM SPSS Modeler desde IBM SPSS Collaboration and Deployment Services Deployment Manager en el *Manual de despliegue de IBM SPSS Modeler*.

Las secciones siguientes proporcionan información sobre cómo acceder al repositorio desde IBM SPSS Modeler.

Compatibilidad con versiones extendidas y búsquedas

El repositorio ofrece capacidades amplias de compatibilidad con versiones de objeto y búsquedas. Por ejemplo, supongamos que crea una ruta y la almacena en el repositorio, donde se puede compartir con investigadores de otras divisiones. Si posteriormente se actualiza la ruta en IBM SPSS Modeler, se puede añadir la versión actualizada al repositorio sin que se sobrescriba la versión anterior. Todas las versiones permanecerán disponibles y en ellas se podrán realizar búsquedas por nombre, etiqueta, campos usados y otros atributos. Por ejemplo, podría buscar todas las versiones del modelo que usen ingresos netos como entrada o aquellos modelos creados por un autor determinado. (Para hacer esto en un sistema de archivo tradicional, tendría que guardar cada versión con un nombre de archivo diferente y las relaciones entre las versiones serían desconocidas para el software.)

Inicio de sesión único

La característica de inicio de sesión único permite que los usuarios se conecten al repositorio sin tener que introducir cada vez los detalles de nombre de usuario y contraseña. Los detalles de inicio de sesión en la red local existente del usuario proporcionan la autenticación necesaria para IBM SPSS Collaboration and Deployment Services. Esta característica depende de lo siguiente:

- IBM SPSS Collaboration and Deployment Services se debe configurar para usar un proveedor de inicio de sesión único.
- El usuario debe iniciar la sesión en un host compatible con el proveedor.

Consulte el tema “Conexión con el repositorio” en la página 121 para obtener más información.

Almacenamiento y despliegue de objetos de repositorio

Las rutas creadas en IBM SPSS Modeler se pueden **almacenar** en el repositorio tal cual, como archivos con la extensión *.str*. De este modo, varios usuarios de la empresa pueden acceder a una única ruta. Consulte el tema “Almacenamiento de objetos en el repositorio” en la página 122 para obtener más información.

También es posible **desplegar** una ruta en el repositorio. Una ruta desplegada se almacena como un archivo con metadatos adicionales. Una ruta desplegada puede aprovechar al máximo las características de ámbito empresarial de IBM SPSS Collaboration and Deployment Services, como la puntuación automatizada y la actualización de modelos. Por ejemplo, se puede actualizar automáticamente un modelo a intervalos periódicos programados, según se vaya disponiendo de nuevos datos. Asimismo, se puede desplegar un conjunto de rutas para el análisis Champion Challenger, en el que se comparen rutas para determinar cuál contiene el modelo predictivo más efectivo.

Puede desplegar una ruta de una de estas formas: en forma de ruta (con la extensión *.str*) o en forma de escenario (con la extensión *.scn*). El despliegue en forma de ruta permite que la aplicación cliente de baja intensidad IBM SPSS Modeler Advantage utilice la ruta. Consulte el tema “Apertura de una ruta en IBM SPSS Modeler Advantage” en la página 139 para obtener más información. El despliegue en forma de escenario permite que Predictive Applications versión 5, el predecesor de IBM SPSS Modeler Advantage, utilice la ruta.

Si desea obtener más información, consulte “Opciones de despliegue de rutas” en la página 133.

Requisitos de rutas desplegadas como escenarios

- Para garantizar un acceso coherente a datos empresariales, debe accederse a las rutas desplegadas como escenarios a través del componente Enterprise View de IBM SPSS Collaboration and Deployment Services. Eso significa que en IBM SPSS Modeler, debe haber al menos un nodo de origen Enterprise View dentro de cada rama de puntuación o modelado designada en la ruta.
- Para utilizar el nodo Enterprise View, es necesario instalar, configurar y permitir el acceso a IBM SPSS Collaboration and Deployment Services desde el sitio, con Enterprise View, vistas de aplicación y definiciones de proveedor de datos (DPD) ya definidas. Si desea obtener más información, póngase en contacto con el administrador local o consulte la página Web corporativa en <http://www.ibm.com/software/analytics/spss/products/deployment/cds/>.
- Una DPD se define respecto a un determinado origen de datos ODBC. Para usar una DPD de IBM SPSS Modeler, debe tener un origen de datos ODBC definido en el host del servidor de IBM SPSS Modeler que tenga el mismo nombre y que conecte con el mismo almacén de datos que aquel al que se hace referencia en la DPD.
- Además, es necesario instalar Controlador de IBM SPSS Collaboration and Deployment Services Enterprise View Driver en cada ordenador que se utilice para modificar o ejecutar la ruta. Para Windows, basta con instalar el controlador en el ordenador donde se ha instalado SPSS Modeler o SPSS Modeler Server y no será necesario realizar ninguna configuración adicional del controlador. En UNIX, es necesario añadir una referencia al script *pev.sh* al script de inicio. Póngase en contacto con el administrador local si desea obtener más información acerca de la instalación del controlador Controlador de IBM SPSS Collaboration and Deployment Services Enterprise View Driver.

Otras opciones de despliegue

Mientras que IBM SPSS Collaboration and Deployment Services ofrece las características más amplias para administrar el contenido empresarial, existen también otros mecanismos para desplegar o exportar rutas, incluyendo:

- Exporte la ruta y el modelo para su uso posterior con IBM SPSS Modeler Solution Publisher Runtime.

- Exporte uno o varios modelos en PMML, un formato basado en XML para la codificación de información de modelos. Consulte el tema “Cómo importar y exportar modelos como PMML” en la página 140 para obtener más información.

Conexión con el repositorio

1. Para conectarse al repositorio, en menú principal de IBM SPSS Modeler, pulse en:

Herramientas > Repositorio > Opciones

2. En el campo **URL del repositorio**, especifique o seleccione la vía de acceso del repositorio, o el URL, de la instalación del repositorio a la que desea acceder. Sólo es posible conectarse a un repositorio al mismo tiempo.

Las opciones de configuración son específicas de cada sitio o instalación. Si desea obtener información sobre el inicio de sesión, póngase en contacto con el administrador del sistema local.

Note: Es necesario disponer de una licencia independiente para acceder a un repositorio IBM SPSS Collaboration and Deployment Services. Si desea obtener más información, consulte <http://www.ibm.com/software/analytics/spss/products/deployment/cds/>

Establecer credenciales. Deje esta casilla sin seleccionar para activar la característica de **inicio de sesión único**, que tratará de iniciar la sesión del usuario con los detalles de nombre de usuario y contraseña del equipo local. Si no es posible el inicio de sesión único para este o si selecciona esta casilla para desactivar el inicio de sesión único (por ejemplo, para iniciar la sesión en una cuenta de administrador), aparecerá otra pantalla para que introduzca las credenciales.

Introducción de credenciales para el repositorio

En función de los valores, pueden ser necesarios los campos siguientes en el cuadro de diálogo Repositorio: Credenciales:

ID de usuario y contraseña. Especifique un nombre de usuario y una contraseña válidos para iniciar sesión. Si es necesario, póngase en contacto con el administrador local para obtener más información.

Proveedor. Seleccione un proveedor de seguridad para la autenticación. El repositorio se puede configurar para usar proveedores de seguridad distintos; si es necesario, póngase en contacto con el administrador local para obtener más información.

Recordar repositorio e ID del usuario. Guarda la configuración actual como configuración predeterminada para no tener que volver a introducir los datos cada vez que se desee conectar.

Exploración del contenido del repositorio

El repositorio permite explorar contenido almacenado de manera parecida al Explorador de Windows; también puede explorar *versiones* de cada objeto almacenado.

1. Para abrir la ventana IBM SPSS Collaboration and Deployment Services Repository, en los menús de SPSS Modeler pulse:

Herramientas > Repositorio > Explorar...

1. Especifique los valores de conexión con el repositorio si es necesario. Consulte el tema “Conexión con el repositorio” para obtener más información. Si desea obtener información sobre el puerto concreto, la contraseña y otros datos de conexión, póngase en contacto con el administrador del sistema local.

La ventana del explorador muestra inicialmente una vista en árbol de la jerarquía de carpetas. Pulse en el nombre de una carpeta para ver su contenido.

Los objetos que reúnen los criterios de la búsqueda o la selección actual se muestran en el panel derecho, y la información detallada acerca de la versión seleccionada se muestra en el panel inferior derecho. Los atributos mostrados se aplican a versión más reciente.

Almacenamiento de objetos en el repositorio

Puede almacenar rutas, nodos, modelos, paletas de modelos, proyectos y objetos de resultados en el repositorio, desde el que otros usuarios y aplicaciones pueden acceder a ellos.

Note: Es necesario disponer de una licencia independiente para acceder a un repositorio IBM SPSS Collaboration and Deployment Services. Si desea obtener más información, consulte <http://www.ibm.com/software/analytics/spss/products/deployment/cds/>

También puede publicar un resultado de rutas en el repositorio en un formato que permita a otros usuarios verlo en Internet mediante Portal de despliegue de IBM SPSS Collaboration and Deployment Services.

Configuración de propiedades de objeto

Cuando almacena un objeto, se muestra el cuadro de diálogo Repositorio: Almacenar, que permite configurar los valores de distintas propiedades del objeto. Tiene la posibilidad de:

- Elegir el nombre y la carpeta del repositorio en la que se almacenará el objeto
- Añadir información sobre el objeto, como la etiqueta de la versión y otras propiedades de búsqueda
- Asignar uno o más temas de clasificación al objeto
- Definir las opciones de seguridad del objeto

Las siguientes secciones describen las propiedades que se pueden configurar.

Elección de la ubicación de almacenamiento de objetos

En el cuadro de diálogo Repositorio: Almacenar, especifique lo siguiente.

Guardar en Muestra la carpeta actual, en la que se almacenará el objeto. Pulse dos veces en un nombre de carpeta de la lista para configurar la carpeta como carpeta actual. Utilice el botón Carpeta superior para navegar a la carpeta padre. Utilice el botón Nueva carpeta para crear una carpeta en el nivel actual.

Nombre del archivo. El nombre con el que se guardará el objeto.

Almacenar. Almacena el objeto en la ubicación actual.

Adición de información sobre objetos almacenados

Todos los campos de la pestaña Información del cuadro de diálogo Repositorio: Almacenar son opcionales.

Autor. Nombre de usuario de quien crea el objeto en el repositorio. De forma predeterminada, muestra el nombre de usuario empleado para la conexión de repositorio, pero puede cambiarlo aquí.

Etiqueta de la versión. Seleccione una etiqueta de la lista para indicar la versión del objeto, o pulse en **Añadir** para crear una nueva etiqueta. Evite utilizar el carácter "[" en la etiqueta. Asegúrese de que no hay ninguna casilla seleccionada si no desea asignar una etiqueta a esta versión de objeto. Consulte el tema "Visualización y edición de propiedades de objetos" en la página 131 para obtener más información.

Descripción. Una descripción del objeto. Los usuarios pueden buscar objetos por descripción (consulte la nota).

Palabras clave. Una o más palabras clave relacionadas con el objeto, que se pueden utilizar para realizar búsquedas (consulte la nota).

Caducidad. Fecha tras la cual el objeto deja de estar visible para los usuarios en general, aunque lo pueden seguir viendo el propietario del objeto y el administrador del repositorio. Para configurar una fecha de caducidad, seleccione la opción **Fecha** y escriba la fecha, o bien elija una utilizando el botón de calendario.

Almacenar. Almacena el objeto en la ubicación actual.

Note: La información de los campos **Descripción** y **Palabras clave** se trata como distinta de cualquier cosa introducida en SPSS Modeler en la pestaña Anotaciones del objeto. Una búsqueda en el repositorio por descripción o palabra clave no devuelve información de la pestaña Anotaciones. Consulte el tema “Búsqueda de objetos en el repositorio” en la página 127 para obtener más información.

Asignación de temas a un objeto almacenado

Los temas son un sistema de clasificación jerárquica para el contenido almacenado en el repositorio. Puede seleccionar de entre los temas disponibles cuando almacene objetos, y los usuarios pueden buscar objetos por tema. Pueden configurar la lista de temas disponibles los usuarios del repositorio con los privilegios pertinentes (para obtener más información, consulte la *Guía del usuario de Deployment Manager*).

Para asignar un tema al objeto, en la pestaña Temas del cuadro de diálogo Repositorio: Almacenar:

1. Pulse en el botón **Añadir**.
2. Pulse en un nombre de tema de la lista de temas disponibles.
3. Pulse en **Aceptar**.

Para eliminar una asignación de tema:

4. Seleccione el tema en la lista de temas asignados.
5. Pulse en **Eliminar**.

Configuración de las opciones de seguridad para objetos almacenados

Puede configurar o cambiar varias opciones de seguridad para un objeto almacenado en la pestaña Seguridad del cuadro de diálogo Repositorio: Almacenar. Para uno o más **principales** (es decir, usuarios o grupos de usuarios), puede:

- Asignar derechos de acceso al objeto
- Modificar derechos de acceso al objeto
- Eliminar derechos de acceso al objeto

Principal. El nombre de usuario del repositorio para el usuario o el grupo con derechos de acceso sobre el objeto.

Permisos. Los derechos de acceso que este usuario o grupo tienen para el objeto.

Añadir. Permite añadir uno o más usuarios o grupos a la lista de las personas con derechos de acceso sobre el objeto. Consulte el tema “Adición de un usuario a la lista de permisos” para obtener más información.

Modificar. Permite modificar los derechos de acceso del usuario o grupo seleccionado para el objeto. El acceso de lectura se concede de forma predeterminada. Esta opción permite garantizar derechos de acceso adicionales, en concreto Propietario, Escribir, Eliminar y Modificar permisos.

Eliminar. Elimina el usuario o el grupo seleccionado de la lista de permisos del objeto.

Adición de un usuario a la lista de permisos: Están disponibles los campos siguientes cuando selecciona **Añadir** en la pestaña Seguridad del cuadro de diálogo Repositorio: Almacenar.

Seleccionar proveedor. Seleccione un proveedor de seguridad para la autenticación. El repositorio se puede configurar para usar proveedores de seguridad distintos; si es necesario, póngase en contacto con el administrador local para obtener más información.

Buscar. Escriba el nombre de usuario del repositorio para el usuario o el grupo que desee añadir, y pulse en **Buscar** para mostrar dicho nombre en la lista de usuarios. Para añadir más de un nombre de usuario a la vez, deje en blanco este campo y pulse simplemente en **Buscar** para mostrar una lista de todos los nombres de usuario del repositorio.

Lista de usuarios. Seleccione uno o más nombres de usuario de la lista y pulse en Aceptar para añadirlos a la lista de permisos.

Modificación de los derechos de acceso de un objeto: Esstán disponibles los campos siguientes cuando selecciona **Modificar** en la pestaña Seguridad del cuadro de diálogo Repositorio: Almacenar.

Propietario. Seleccione esta opción para dar a este usuario o grupo derechos de acceso al objeto como propietario. El propietario tiene pleno control del objeto, incluidos derechos de acceso para eliminar y modificar.

Leer. De forma predeterminada, un usuario o grupo que no sea propietario del objeto tendrá solamente derechos de acceso de lectura al objeto. Seleccione las casillas de verificación correspondientes para añadir derechos de acceso de escritura, eliminación y modificación de permisos para ese usuario o grupo.

Almacenamiento de rutas

Puede almacenar una ruta como archivo *.str* en el repositorio, desde donde podrán acceder a ella otros usuarios.

Note: Si desea más información sobre el despliegue de una ruta, para aprovechar las características adicionales del repositorio, consulte “Despliegue de rutas” en la página 132.

Para almacenar la ruta actual:

1. En el menú principal, pulse en:
Archivo > Almacenar > Almacenar ruta...
2. Especifique los valores de conexión con el repositorio si es necesario. Consulte el tema “Conexión con el repositorio” en la página 121 para obtener más información. Si desea obtener información sobre el puerto concreto, la contraseña y otros datos de conexión, póngase en contacto con el administrador del sistema local.
3. En el cuadro de diálogo Repositorio: Almacenar, seleccione la carpeta donde desea almacenar el objeto, especifique cualquier otra información que desee registrar y pulse en el botón **Almacenar**. Consulte el tema “Configuración de propiedades de objeto” en la página 122 para obtener más información.

Almacenamiento de proyectos

Puede almacenar un proyecto completo de IBM SPSS Modeler como archivo *.cpj* en el repositorio, para que otros usuarios puedan acceder a él.

Dado que un archivo de proyecto contiene otros objetos de IBM SPSS Modeler, necesitará indicar a IBM SPSS Modeler que almacene los objetos del proyecto en el repositorio. Esta operación se puede realizar utilizando una configuración del cuadro de diálogo Propiedades de proyecto. Consulte el tema “Configuración de las propiedades de un proyecto” en la página 146 para obtener más información.

Una vez configurado un proyecto para almacenar objetos en el repositorio, IBM SPSS Modeler solicitará automáticamente que se almacene el objeto siempre se añada un nuevo objeto al proyecto.

Cuando haya terminado la sesión de IBM SPSS Modeler, deberá almacenar una nueva versión del archivo de proyecto para que recuerde las adiciones. El archivo de proyecto contiene automáticamente (y recupera) las últimas versiones de sus objetos. Si no añade ningún objeto a un proyecto durante una sesión de IBM SPSS Modeler, no tendrá que volver a almacenar el archivo de proyecto. Sin embargo, deberá almacenar nuevas versiones para los objetos del proyecto (ruta, resultados, etc.) que haya cambiado

Para almacenar un proyecto

1. Seleccione el proyecto en la pestaña CRISP-DM o Clases del panel de gestores, en IBM SPSS Modeler, y en el menú principal pulse:
Archivo > Proyecto > Almacenar proyecto...
2. Especifique los valores de conexión con el repositorio si es necesario. Consulte el tema “Conexión con el repositorio” en la página 121 para obtener más información. Si desea obtener información sobre el puerto concreto, la contraseña y otros datos de conexión, póngase en contacto con el administrador del sistema local.
3. En el cuadro de diálogo Repositorio: Almacenar, seleccione la carpeta donde desea almacenar el objeto, especifique cualquier otra información que desee registrar y pulse en el botón **Almacenar**. Consulte el tema “Configuración de propiedades de objeto” en la página 122 para obtener más información.

Almacenamiento de nodos

Puede almacenar una única definición de nodo de la ruta actual en forma de archivo *.nod* en el repositorio, desde donde otros usuarios pueden acceder a ella.

Para almacenar un nodo:

1. Pulse con el botón derecho del ratón en el nodo del lienzo de rutas y pulse **Almacenar nodo**.
2. Especifique los valores de conexión con el repositorio si es necesario. Consulte el tema “Conexión con el repositorio” en la página 121 para obtener más información. Si desea obtener información sobre el puerto concreto, la contraseña y otros datos de conexión, póngase en contacto con el administrador del sistema local.
3. En el cuadro de diálogo Repositorio: Almacenar, seleccione la carpeta donde desea almacenar el objeto, especifique cualquier otra información que desee registrar y pulse en el botón **Almacenar**. Consulte el tema “Configuración de propiedades de objeto” en la página 122 para obtener más información.

Almacenamiento de objetos de resultado

Puede almacenar un objeto de resultado de la ruta actual en forma de archivo *.cou* en el repositorio, desde donde otros usuarios pueden acceder a ella.

Para almacenar un objeto de resultado:

1. Pulse en el objeto en la pestaña Resultados del panel de gestores, en SPSS Modeler, y en el menú principal pulse:
Archivo > Resultados > Almacenar resultado...
2. También puede pulsar con el botón derecho en la pestaña Resultados y pulsar **Almacenar**.
3. Especifique los valores de conexión con el repositorio si es necesario. Consulte el tema “Conexión con el repositorio” en la página 121 para obtener más información. Si desea obtener información sobre el puerto concreto, la contraseña y otros datos de conexión, póngase en contacto con el administrador del sistema local.
4. En el cuadro de diálogo Repositorio: Almacenar, seleccione la carpeta donde desea almacenar el objeto, especifique cualquier otra información que desee registrar y pulse en el botón **Almacenar**. Consulte el tema “Configuración de propiedades de objeto” en la página 122 para obtener más información.

Almacenamiento de modelos y paletas de modelos

Puede almacenar una ruta concreta como archivo *.gm* en el repositorio, desde donde podrán acceder a ella otros usuarios. También puede almacenar el contenido completo de la paleta de modelos como archivo *.gen* en el repositorio.

Almacenamiento de modelos:

1. Pulse en el objeto de la paleta Modelos en SPSS Modeler, y en el menú principal pulse:
Archivo > Modelos > Almacenar modelo...
2. También puede pulsar con el botón derecho en un objeto de la paleta Modelos y pulsar **Almacenar modelo**.
3. Continúe a partir de "Finalización del procedimiento de almacenamiento", más abajo.

Almacenamiento de paletas de modelos

1. Pulse con el botón derecho en el fondo de la paleta Modelos.
2. En el menú emergente, pulse **Almacenar paleta**.
3. Continúe a partir de "Finalización del procedimiento de almacenamiento", más abajo.

Finalización del procedimiento de almacenamiento:

1. Especifique los valores de conexión con el repositorio si es necesario. Consulte el tema "Conexión con el repositorio" en la página 121 para obtener más información. Si desea obtener información sobre el puerto concreto, la contraseña y otros datos de conexión, póngase en contacto con el administrador del sistema local.
2. En el cuadro de diálogo Repositorio: Almacenar, seleccione la carpeta donde desea almacenar el objeto, especifique cualquier otra información que desee registrar y pulse en el botón **Almacenar**. Consulte el tema "Configuración de propiedades de objeto" en la página 122 para obtener más información.

Recuperación de objetos desde el repositorio

Puede recuperar rutas, modelos, paletas de modelos, nodos, proyectos y objetos de resultados que se hayan almacenado en el repositorio.

Note: Además de utilizar las opciones de menú descritas, también puede recuperar rutas, objetos de resultados, modelos y paletas de modelos pulsando con el botón derecho en la pestaña correspondiente del panel de gestor, en la parte superior derecha de la ventana de SPSS Modeler.

1. Para recuperar una ruta, en el menú principal de IBM SPSS Modeler, pulse:
Archivo > Recuperar ruta...
2. Para recuperar un modelo, paleta de modelos, proyecto u objeto de resultados, en el menú principal de IBM SPSS Modeler pulse:
Archivo > Modelos > Recuperar modelo...
o
Archivo > Modelos > Recuperar paleta de modelos...
o
Archivo > Proyecto > Recuperar proyecto...
o
Archivo > Resultados > Recuperar resultado...
3. También puede pulsar con el botón derecho en el panel del proyecto o gestores y pulsar **Recuperar** en el menú emergente.
4. Para recuperar un nodo, en el menú principal de IBM SPSS Modeler, pulse:
Insertar > Nodo (o Supernodo) de repositorio...

- a. Especifique los valores de conexión con el repositorio si es necesario. Consulte el tema “Conexión con el repositorio” en la página 121 para obtener más información. Si desea obtener información sobre el puerto concreto, la contraseña y otros datos de conexión, póngase en contacto con el administrador del sistema local.
5. En el cuadro de diálogo Repositorio: Recuperar, busque el objeto mediante Examinar, selecciónelo y pulse el botón **Recuperar**. Consulte el tema para obtener más información.

Elección de un objeto que recuperar

Los campos siguientes están disponibles en el cuadro de diálogo Repositorio: Recuperar/Buscar:

Buscar en: Muestra la jerarquía de carpetas de la carpeta actual. Para desplazarse a otra carpeta, selecciónela en la lista para navegar directamente hasta ella, o bien utilice la lista de objetos que aparece bajo este campo para llegar a ella.

Botón Carpeta superior Sube un nivel con respecto a la carpeta actual en la jerarquía.

Botón Carpeta nueva. Crea una nueva carpeta en el nivel actual de la jerarquía.

Nombre del archivo. El nombre de archivo del repositorio para el objeto seleccionado. Para recuperar dicho objeto, pulse en **Recuperar**.

Archivos del tipo. El tipo de objeto que ha seleccionado para recuperar. Solamente se mostrarán, en la lista de objetos, objetos de este tipo, junto con las carpetas. Para mostrar objetos de otro tipo para su recuperación, seleccione el tipo en cuestión en la lista.

Abrir como bloqueado. De forma predeterminada, cuando se recupera un objeto, éste se bloquea en el repositorio para que otras personas no puedan actualizarlo. Si no desea que el objeto se bloquee al recuperarlo, desactive esta casilla.

Descripción, palabras clave. Si, al almacenar el objeto, se definieron detalles adicionales al respecto, éstos aparecerán aquí. Consulte el tema “Adición de información sobre objetos almacenados” en la página 122 para obtener más información.

Versión. Para recuperar una versión de un objeto distinta de la más reciente, pulse este botón. Se mostrará información de todas las versiones para que pueda seleccionar la que desee.

Selección de una versión de objeto

Para seleccionar una versión concreta de un objeto de repositorio, en el cuadro de diálogo Repositorio: Seleccionar versión:

1. (Opcional) Ordene la lista por versión, etiqueta, tamaño, fecha de creación o usuario de creación, pulsando dos veces en la cabecera de la columna apropiada.
2. Seleccione la versión del objeto con la que desee trabajar.
3. Pulse en Continuar.

Búsqueda de objetos en el repositorio

Puede buscar objetos por nombre, carpeta, tipo, etiqueta, fecha u otro criterio.

Búsqueda por nombre

Para buscar objetos por nombre:

1. En el menú principal de IBM SPSS Modeler, pulse en:
Herramientas > Repositorio > Explorar...

- a. Especifique los valores de conexión con el repositorio si es necesario. Consulte el tema “Conexión con el repositorio” en la página 121 para obtener más información. Si desea obtener información sobre el puerto concreto, la contraseña y otros datos de conexión, póngase en contacto con el administrador del sistema local.
2. Pulse en la pestaña **Buscar**.
3. En el campo **Buscar objetos con el nombre**, especifique el nombre del objeto que desea buscar.

Cuando busca objetos por nombre, se puede utilizar un asterisco (*) como carácter comodín para que represente cualquier cadena de caracteres, y un signo de interrogación (?) representa un carácter individual. Por ejemplo, **cluster** coincide con todos los objetos que incluyen la cadena *cluster* en cualquier parte del nombre. La cadena de búsqueda *m0?_** coincide con *M01_cluster.str* y *M02_cluster.str* pero no con *M01a_cluster.str*. Las búsquedas no distinguen entre mayúsculas y minúsculas (*cluster* coincide con *Cluster* y coincide con *CLUSTER*).

Note: si el número de objetos es grande, las búsquedas pueden tardar unos momentos.

Búsqueda por otros criterios

Puede realizar una búsqueda según el título, la etiqueta, las fechas, el autor, las palabras clave, el contenido indexado o la descripción. Solamente se buscarán los objetos que reúnan *todos* los criterios de búsqueda especificados. Por ejemplo, podría localizar todas las rutas que contengan uno o más modelos de agrupación en clústeres que también tengan aplicada una etiqueta específica, y se hayan modificado con posterioridad a una fecha específica.

Tipos de objeto. Puede restringir la búsqueda a modelos, rutas, resultados, nodos, Supernodos, proyectos, paletas de modelos, escenarios y otros tipos de objetos.

- **Modelos.** Puede realizar búsquedas de modelos por categoría (clasificación, aproximación, agrupación en clústeres, etc.), o por un algoritmo específico de modelo, como Kohonen.

También puede realizar búsquedas por campos utilizados, por ejemplo, todos los modelos que usen un campo denominado *ingreso* como objetivo o resultado.

- **Rutas.** Para las rutas, puede restringir la búsqueda por campos utilizados o tipo de modelo (categoría o algoritmo) que contenga la ruta.

Temas. Puede realizar búsquedas por modelos asociados a temas específicos desde un conjunto de listas de usuarios del repositorio con los privilegios correspondientes (si desea más información, consulte la *Guía del usuario de Deployment Manager*). Para obtener la lista, active esta casilla, pulse el botón Añadir temas que aparece, seleccione uno o más temas de la lista y pulse en Aceptar.

Etiqueta. La búsqueda se restringe a las etiquetas de versión de objeto seleccionadas.

Fechas. Se puede especificar una fecha de creación o modificación y buscar objetos anteriores o posteriores o pertenecientes a un intervalo determinado.

Autor. Restringe la búsqueda a objetos creados por un usuario concreto.

Palabras clave. Búsqueda por palabras clave determinadas. En IBM SPSS Modeler las palabras clave se especifican en la pestaña Anotaciones de una ruta, un modelo o un objeto de resultados.

Descripción. Búsqueda basada en términos concretos del campo de descripción. En IBM SPSS Modeler la descripción se especifica en la pestaña Anotaciones de una ruta, un modelo o un objeto de resultados. Se pueden especificar varias frases de búsqueda separadas por punto y coma (por ejemplo, ingresos; tipo recorte; valor de reclamación). (Tenga en cuenta que dentro de una frase de búsqueda, los espacios cuentan. Por ejemplo, tipo recorte, con un espacio, y tipo recorte, con dos espacios, no es lo mismo.)

Modificación de objetos del repositorio

Puede modificar los objetos existentes en el repositorio directamente desde SPSS Modeler. Tiene la posibilidad de:

- Crear, cambiar de nombre o eliminar carpetas
- Bloquear o desbloquear objetos
- Eliminar objetos

Creación, cambio de nombre y eliminación de carpetas

1. Para realizar operaciones en carpetas del repositorio, en el menú principal de SPSS Modeler, pulse en: **Herramientas > Repositorio > Explorar...**
 - a. Especifique los valores de conexión con el repositorio si es necesario. Consulte el tema “Conexión con el repositorio” en la página 121 para obtener más información. Si desea obtener información sobre el puerto concreto, la contraseña y otros datos de conexión, póngase en contacto con el administrador del sistema local.
2. Asegúrese de que la pestaña **Carpetas** está activa.
3. Para añadir una nueva carpeta, pulse con el botón derecho en la carpeta padre y pulse en **Nueva carpeta**.
4. Para cambiar el nombre de una carpeta, pulse con el botón derecho del ratón y pulse **Cambiar nombre de carpeta**.
5. Para eliminar una carpeta, pulse con el botón derecho del ratón y seleccione **Eliminar carpeta**.

Bloqueo y desbloqueo de objetos y repositorio

Puede bloquear un objeto para evitar que otros usuarios actualicen cualquiera de las versiones existentes o creen nuevas versiones. Un objeto bloqueado se identifica por un icono de candado sobre el mismo.



Figura 15. Objeto bloqueado

Para bloquear un objeto:

1. En la ventana del explorador del repositorio, pulse con el botón derecho del ratón en el objeto deseado.
2. Pulse en **Bloquear**.

Para desbloquear un objeto:

1. En la ventana del explorador del repositorio, pulse con el botón derecho del ratón en el objeto deseado.
2. Pulse **Desbloquear**.

Eliminación de objetos del repositorio

Antes de eliminar un objeto del repositorio, debe decidir si desea eliminar todas las versiones del objeto o solamente una versión determinada.

Para eliminar todas las versiones de un objeto

1. En la ventana del explorador del repositorio, pulse con el botón derecho del ratón en el objeto deseado.
2. Pulse en **Eliminar objetos**.

Para eliminar la versión más reciente de un objeto

1. En la ventana del explorador del repositorio, pulse con el botón derecho del ratón en el objeto deseado.
2. Pulse en **Eliminar**.

Para eliminar una versión anterior de un objeto

1. En la ventana del explorador del repositorio, pulse con el botón derecho del ratón en el objeto deseado.
2. Pulse en **Eliminar versiones**.
3. Seleccione las versiones que va a eliminar y pulse en **Aceptar**.

Gestión de las propiedades de objetos del repositorio

Puede controlar distintas propiedades de objetos desde SPSS Modeler. Tiene la posibilidad de:

- Ver las propiedades de una carpeta
- Ver y editar las propiedades de un objeto
- Crear, aplicar y eliminar etiquetas de versión de un objeto

Visualización de propiedades de carpeta

Para ver las propiedades de cualquier carpeta en la ventana del repositorio, pulse con el botón derecho del ratón en la carpeta deseada. Pulse en **Propiedades de la carpeta**.

pestaña General

Muestra el nombre de la carpeta y las fechas de creación y modificación.

Pestaña Permisos

La pestaña Permisos permite establecer permisos de lectura y escritura para la carpeta. Se enumeran todos los usuarios y grupos con acceso a la carpeta padre. Los permisos siguen una jerarquía. Por ejemplo, si no tiene permisos de lectura, no podrá tener permisos de escritura. Si no tiene permisos de escritura, no podrá tener permisos de eliminación.

Usuarios y grupos. Muestra los usuarios y grupos del repositorio que tienen al menos acceso de lectura a la carpeta. Seleccione las casillas de verificación de escritura y lectura para añadir derechos de acceso a la carpeta para un usuario o un grupo concretos. Pulse en el icono **Añadir usuarios/grupos** de la parte derecha de la pestaña Permisos para asignar acceso a grupos y usuarios adicionales. El administrador controla la lista de usuarios y grupos disponibles.

Representar en cascada los permisos. Elija una opción para controlar cómo se aplicarán en las carpetas hijas los cambios realizados en la carpeta actual, si procede.

- **Representar en cascada todos los permisos.** Organiza en cascada la configuración de los permisos desde la carpeta actual hasta todas las carpetas hijas y descendientes. Se trata de una forma rápida de establecer permisos para varias carpetas a la vez. Establezca los permisos necesarios en la carpeta padre y, a continuación, propague en cascada como desee.
- **Representar en cascada solamente los cambios.** Se representan en cascada solamente los cambios realizados desde la última vez que se aplicaron cambios. Por ejemplo, si se ha añadido un nuevo grupo y desea otorgarle a éste acceso a todas las carpetas bajo la rama Ventas, puede darle acceso a la carpeta raíz Ventas y representar en cascada los cambios realizados en todas las subcarpetas. El resto de permisos para las subcarpetas existentes permanecen sin cambios.
- **No representar en cascada.** Todos los cambios efectuados afectarán solamente a la carpeta actual y no se propagarán en cascada a las carpetas hijas.

Visualización y edición de propiedades de objetos

En el cuadro de diálogo Propiedades del objeto, puede ver y editar propiedades. Aunque algunas propiedades no se puedan modificar, siempre se puede actualizar un objeto añadiendo una nueva versión.

1. En la ventana del repositorio, pulse con el botón derecho del ratón en el objeto deseado.
2. Pulse en **Propiedades del objeto**.

Pestaña General

Nombre. El nombre del objeto tal como se ve en el repositorio.

Creado el. Fecha de creación del objeto (no la versión).

Última modificación. Fecha de la última modificación efectuada en el objeto.

Autor. Nombre de inicio de sesión del usuario.

Descripción. De forma predeterminada, contiene la descripción especificada en la pestaña Anotaciones del objeto en SPSS Modeler.

Temas enlazados. El repositorio permite organizar por temas los modelos y objetos relacionados si se estima oportuno. Pueden configurar la lista de temas disponibles los usuarios del repositorio con los privilegios pertinentes (para obtener más información, consulte la *Guía del usuario de Deployment Manager*).

Palabras clave. Permite especificar palabras clave en la pestaña Anotaciones para una ruta, un modelo o un objeto de resultados. Las series de palabras clave deben estar separadas por espacios y no superar los 255 caracteres. (Si las palabras clave contienen espacios, utilice las comillas para separarlas.)

Pestaña Versiones

Los objetos almacenados en el repositorio pueden tener varias versiones. En la pestaña Versiones se muestra información acerca de cada versión.

Las propiedades que se detallan a continuación se pueden especificar o modificar en determinadas versiones de un objeto almacenado:

Versión. Identificador exclusivo de la versión, generado basándose en el momento en que se almacenó la versión.

Etiqueta. Etiqueta actual de la versión, si existe. Al contrario que ocurre con el identificador de la versión, las etiquetas se pueden trasladar de una versión de un objeto a otra.

El tamaño del archivo, la fecha de creación y el autor también se muestran en cada versión.

Editar etiquetas. Pulse en el icono **Editar etiquetas** situado en la parte superior derecha de la pestaña Versiones para definir, aplicar o eliminar las etiquetas de objetos almacenados. Consulte el tema “Gestión de etiquetas de versión de objetos” en la página 132 para obtener más información.

Pestaña Permisos

La pestaña Permisos permite establecer permisos de lectura y escritura para el objeto. Se enumeran todos los usuarios y grupos con acceso al objeto actual. Los permisos siguen una jerarquía. Por ejemplo, si no tiene permisos de lectura, no podrá tener permisos de escritura. Si no tiene permisos de escritura, no podrá tener permisos de eliminación.

Usuarios y grupos. Muestra los usuarios y grupos del repositorio que tienen al menos acceso de lectura al objeto. Seleccione las casillas de verificación de escritura y lectura para añadir derechos de acceso a este objeto para un usuario o un grupo concretos. Pulse en el icono **Añadir usuarios/grupos** de la parte derecha de la pestaña Permisos para asignar acceso a grupos y usuarios adicionales. El administrador controla la lista de usuarios y grupos disponibles.

Gestión de etiquetas de versión de objetos

El cuadro de diálogo Editar etiquetas de versión le permite:

- Aplicar etiquetas al objeto seleccionado
- Eliminar las etiquetas del objeto seleccionado
- Definir una nueva etiqueta y aplicarla al objeto

Aplicar etiquetas al objeto seleccionado

1. Seleccione una o más etiquetas de la lista **Etiquetas disponibles**.
2. Pulse el botón de flecha derecha para mover las etiquetas seleccionadas a la lista **Etiquetas aplicadas**.
3. Pulse en **Aceptar**.

Eliminar etiquetas del objeto

1. Seleccione una o más etiquetas de la lista **Etiquetas aplicadas**.
2. Pulse el botón de flecha izquierda para quitar las etiquetas seleccionadas de la lista **Etiquetas disponibles**.
3. Pulse en **Aceptar**.

Definir una nueva etiqueta y aplicarla al objeto

1. Introduzca el nombre de la etiqueta en el campo **Nueva etiqueta**.
2. Pulse el botón de flecha derecha para mover las nuevas etiquetas a la lista **Etiquetas aplicadas**.
3. Pulse en **Aceptar**.

Despliegue de rutas

Para activar la ruta que se utilizará con la aplicación IBM SPSS Modeler Advantage de baja intensidad, se debe desplegar como una ruta (archivo *.str*) en el repositorio.

Independientemente de si una ruta se despliegan como una ruta (archivo *.str*) o como un escenario (archivo *.scn*), el objeto puede aprovechar al máximo las características de ámbito empresarial de IBM SPSS Collaboration and Deployment Services. Consulte el tema "Almacenamiento y despliegue de objetos de repositorio" en la página 120 para obtener más información.

Para desplegar la ruta actual (método del menú Archivo)

1. En el menú principal, pulse en:
Archivo > Almacenar > Desplegar
2. Seleccione el tipo de despliegue y complete el resto del cuadro de diálogo si fuera necesario.
3. Pulse en **Desplegar como ruta** para desplegar la ruta para su uso con IBM SPSS Modeler Advantage o IBM SPSS Collaboration and Deployment Services. Pulse en **Desplegar como escenario** para desplegar la ruta para su uso con IBM SPSS Collaboration and Deployment Services o Predictive Applications versión 5.
4. Pulse en **Almacenar**. Si desea obtener más información, pulse en **Ayuda**.
5. Continúa a partir de "Finalización del proceso de despliegue".

Para desplegar la ruta actual (método del menú Herramientas)

1. En el menú principal, pulse en:

Herramientas > Propiedades de ruta > Despliegue

2. Elija el tipo de despliegue, complete el resto de la pestaña Despliegue según sea necesario y pulse en **Almacenar**. Consulte el tema “Opciones de despliegue de rutas” para obtener más información.

Finalización del proceso de despliegue

1. Especifique los valores de conexión con el repositorio si es necesario. Consulte el tema “Conexión con el repositorio” en la página 121 para obtener más información. Si desea obtener información sobre el puerto concreto, la contraseña y otros datos de conexión, póngase en contacto con el administrador del sistema local.
2. En el cuadro de diálogo Repositorio: Almacenar, seleccione la carpeta donde desea almacenar el objeto, especifique cualquier otra información que desee registrar y pulse en el botón **Almacenar**. Consulte el tema “Configuración de propiedades de objeto” en la página 122 para obtener más información.

Opciones de despliegue de rutas

La pestaña Despliegue del cuadro de diálogo Opciones de ruta permite especificar opciones para desplegar la ruta. Puede realizar el despliegue como ruta o escenario.

Cuando despliega una ruta, puede abrir y modificar la ruta en la aplicación IBM SPSS Modeler Advantage de baja intensidad. La ruta se guarda en el repositorio como un archivo con la extensión *.str*.

Si la despliega como escenario guarda la ruta en el repositorio como un archivo con la extensión *.scn*. De esta forma activará una ruta que utilizará Predictive Applications versión 5.

Independientemente de si la despliega como ruta o como escenario, puede aprovechar al máximo las funciones adicionales disponibles con IBM SPSS Collaboration and Deployment Services, como el acceso para varios usuarios, la puntuación automatizada, la actualización de modelos y el análisis Champion Challenger.

Desde la pestaña Despliegue también puede obtener una presentación preliminar de la descripción de ruta que IBM SPSS Modeler crea para la ruta. Consulte el tema “Descripciones de ruta” en la página 53 para obtener más información.

Note: Para garantizar un acceso coherente a datos empresariales, una ruta desplegada en forma de escenario debe acceder a los datos de origen a través de IBM SPSS Collaboration and Deployment Services Enterprise View, de modo que debe haber al menos un nodo de origen Enterprise View dentro de cada rama de puntuación o modelado designada, según proceda.

Tipo de despliegue. Selecciona cómo desea desplegar la ruta. Antes de su despliegue, todas las rutas requieren un nodo de puntuación específico; las opciones y requisitos adicionales dependen del tipo de despliegue.

- **<ninguno>**. La ruta no se desplegará en el repositorio. Todas las opciones están desactivadas menos la presentación preliminar de la descripción de rutas.
- **Solamente puntuación**. La ruta se despliega al repositorio cuando pulsa en el botón **Almacenar**. Los datos pueden puntuarse utilizando el nodo que designe en el campo **Nodo Puntuación**.
- **Actualización de modelos**. Hace lo mismo que Solamente puntuación pero además, el modelo puede actualizarse en el repositorio utilizando los objetos que designe en los campos **Nodo de modelado** y **Nugget de modelo**. *Nota:* no se admite el modelo automático de forma predeterminada en IBM SPSS Collaboration and Deployment Services, de modo que debe elegir este tipo de despliegue si desea utilizar esta característica al ejecutar una ruta desde el repositorio. Consulte el tema “Actualización del modelo” en la página 136 para obtener más información.

Nodo Puntuación. Seleccione un nodo de gráfico, resultado o exportación para identificar la rama de la ruta que se utilizará para puntuar los datos. Mientras que la ruta puede contener un número indefinido de ramas, modelos y nodos de terminación válidos, sólo se puede designar una rama de puntuación con fines de despliegue. Éste es el requisito principal para desplegar una ruta.

Parámetros de puntuación. Permite especificar los parámetros que se pueden modificar cuando se ejecuta la rama de puntuación. Consulte el tema “Parámetros de puntuación y modelado” en la página 135 para obtener más información.

Nodo Modelado. Para la actualización de modelos, especifica el nodo de modelado que se utilizará para volver a crear o actualizar el modelo en el repositorio. Debe ser un nodo de modelado del mismo tipo que el especificado para **Nugget de modelo**.

Parámetros del generador de modelos. Permite especificar los parámetros que se pueden modificar al ejecutar el nodo de generación de modelado. Consulte el tema “Parámetros de puntuación y modelado” en la página 135 para obtener más información.

Nugget de modelo. Para la actualización de modelos, especifica el nugget de modelo que se actualizará o regenerará cada vez que se actualice la ruta en el repositorio (normalmente, como parte de un trabajo programado). El modelo se debe encontrar en la rama de puntuación. Aunque puede haber varios modelos en la rama de puntuación, solamente se puede designar a uno. Tenga en cuenta que cuando se crea inicialmente la ruta, puede ser realmente un modelo marcador que se actualiza o regenera a medida que hay nuevos datos disponibles.

Desplegar como ruta. Pulse en esta opción si desea utilizar la ruta con IBM SPSS Modeler Advantage o IBM SPSS Collaboration and Deployment Services (y consulte la nota siguiente).

Desplegar como escenario. Pulse en esta opción si desea utilizar la ruta con IBM SPSS Collaboration and Deployment Services o Predictive Applications versión 5 (y consulte la nota siguiente).

Comprobar. Pulse en este botón para comprobar si se trata de una ruta válida para desplegar. Todas las rutas deben tener un nodo de puntuación específico antes de poder desplegarse. Si está realizando un despliegue en forma de escenario, la ruta también debe contener un nodo de origen Enterprise View válido. Si no se cumplen estas condiciones, aparecerán mensajes de error.

Almacenar. Despliega la ruta si es válida. En caso contrario, aparecerá un mensaje de error. Pulse en el botón **Reparar**, corrija el error y vuelva a intentarlo.

Presentación preliminar de descripción de ruta. Le permite visualizar los contenidos de la descripción de la ruta que IBM SPSS Modeler crea para la ruta. Consulte el tema “Descripciones de ruta” en la página 53 para obtener más información.

Note: (Desplegar como ruta o escenario) Es posible utilizar varios nodos Enterprise View en una rama de modelado. Si es así, en la mayoría de los casos es recomendable utilizar una única conexión de datos para todos los nodos de vista empresariales dentro de una rama determinada, y es necesario el análisis con Champion Challenger.

- Si no es necesaria la compatibilidad con Champion Challenger, se pueden utilizar diferentes conexiones Enterprise View dentro de la misma rama, siempre que éstas solamente varíen según la definición del proveedor de datos (DPD).
- Estas limitaciones solamente se aplican dentro de una rama determinada. Entre las ramas de puntuación y generación del modelo, se pueden utilizar diferentes conexiones Enterprise View sin restricciones.

Parámetros de puntuación y modelado

Al desplegar una ruta a IBM SPSS Collaboration and Deployment Services, puede elegir los parámetros que se podrán ver o editar cada vez que se actualice o puntúe el modelo. Por ejemplo, puede especificar los valores máximo y mínimo o algún otro valor que puede estar sujeto a cambio cada vez que se ejecute un trabajo.

1. Para hacer que un parámetro sea visible y se pueda ver o editar después de desplegar la ruta, selecciónelo en la lista del cuadro de diálogo Parámetros de puntuación.

La lista de parámetros disponibles se define en la pestaña Parámetros del cuadro de diálogo de propiedades de ruta. Consulte el tema “Cómo configurar parámetros de sesión y ruta.” en la página 48 para obtener más información.

La rama de puntuación

Si despliega una ruta, una de sus ramas debe designarse como **rama de puntuación** (es decir, la que contiene el nodo de puntuación). Al designar una rama como rama de puntuación, ésta se resalta en el lienzo de rutas, al igual que el enlace del modelo al nugget en la rama de puntuación. Esta representación visual es particularmente útil en rutas complejas con varias ramas, en las que tal vez no se pueda identificar inmediatamente la ruta de puntuación.

Note: Solamente se puede designar una rama de la ruta como rama de puntuación.

Si la ruta tenía ya una rama de puntuación definida, la nueva rama la sustituirá como tal. Puede definir el color que identifique la rama de puntuación mediante la opción Color personalizado. Consulte el tema “Opciones de configuración de representación” en la página 154 para obtener más información.

Puede mostrar u ocultar la identificación de la rama de puntuación por medio del botón Mostrar/ocultar de la barra de herramientas de marcación de ruta.



Figura 16. Botón Mostrar/ocultar de la barra de herramientas de marcación de ruta:

Identificación de la rama de puntuación para el despliegue

Puede designar la rama de puntuación desde el menú emergente de un nodo terminal o desde el menú Herramientas. Si utiliza el menú emergente, el nodo de puntuación se define automáticamente en la pestaña Despliegue de las propiedades de ruta.

Para designar una rama como rama de puntuación (menú emergente)

1. Conecte el nugget de modelo a un nodo terminal (un nodo de procesamiento o de resultados en dirección descendente desde el nugget).
2. Pulse con el botón derecho en el nodo terminal.
3. En el menú, pulse **Utilizar como rama de puntuación**.

Para designar una rama como rama de puntuación (menú Herramientas)

1. Conecte el nugget de modelo a un nodo terminal (un nodo de procesamiento o de resultados en dirección descendente desde el nugget).
2. En el menú principal, pulse en:
Herramientas > Propiedades de ruta > Despliegue
3. En la lista **Tipo de despliegue**, pulse **Solamente puntuación** o **Actualización de modelos**, según desee. Consulte el tema “Opciones de despliegue de rutas” en la página 133 para obtener más información.

4. Pulse en el campo **Nodo de puntuación** y seleccione un nodo terminal en la lista.
5. Pulse en **Aceptar**.

Actualización del modelo

La actualización de modelos es un proceso de regeneración de un modelo existente en una ruta utilizando datos más recientes. La ruta en sí no cambia en el repositorio. Por ejemplo, el tipo de algoritmo y la configuración específica de la ruta siguen siendo los mismos, pero el modelo vuelve a entrenarse con nuevos datos, y se actualiza si la nueva versión del modelo funciona mejor que la anterior.

En una ruta solamente se puede definir la actualización de un nugget de modelo, que se conoce como **modelo de actualización**. Si elige la opción **Actualización de modelos** en la pestaña Despliegue de las propiedades de ruta (consulte "Opciones de despliegue de rutas" en la página 133), el nugget de modelo que designe en ese momento pasará a ser el modelo de actualización. También puede designar un modelo como modelo de actualización en el menú emergente de un nugget de modelo. Para ello, el nugget debe estar ya en la rama de puntuación.

Desactivar el estado de "modelo de actualización" de un nugget equivale a configurar el tipo de despliegue de la ruta como Solamente puntuación, y la pestaña Despliegue del cuadro de diálogo de propiedades de ruta se actualizará en consecuencia. Puede activar o desactivar este estado con ayuda de la opción **Utilizar como modelo de actualización** del menú emergente del nugget en la rama de puntuación actual.

Si se elimina el enlace de modelo de un nugget en la rama de puntuación, también se eliminará el estado de "modelo de actualización" del nugget. Puede deshacer la eliminación del enlace de modelo con el menú Editar o con la barra de herramientas; de esta forma restablecerá el estado "modelo de actualización" del nugget.

Cómo se selecciona el modelo de actualización

Al igual que la rama de puntuación, también se resalta en la ruta el enlace al modelo de actualización. El nugget de modelo elegido como modelo de actualización, y por tanto el enlace resaltado, dependerán del número de nuggets presentes en la ruta.

Modelo único en ruta

Si en la rama de puntuación hay un único nugget de modelo con enlace, se identificará como tal y pasará a ser el modelo de actualización para la ruta.

Varios modelos en la ruta

Si hay más de un nugget enlazado en la ruta, el modelo de actualización se seleccionará de la siguiente manera:

Si se ha definido un nugget de modelo en la pestaña Despliegue del cuadro de diálogo de propiedades de ruta y también en la ruta, éste pasará a ser el modelo de actualización.

Si no se ha definido ningún nugget en la pestaña Despliegue, o si se ha definido uno, pero no en la rama de puntuación, se seleccionará como modelo de actualización el nugget más cercano al nodo terminal.

Si más tarde elimina la selección de todos los enlaces de modelo como enlaces de actualización, solamente se resaltará la rama de puntuación y no los enlaces. El tipo de despliegue se define en Solamente puntuación.

Note: Puede decidir definir uno de los enlaces para reemplazar el estado, pero no el otro. En ese caso, el nugget de modelo seleccionado como modelo de actualización será el que tenga un enlace de actualización y se encuentre más cerca del nodo terminal cuando se designe la rama de puntuación.

Ausencia de modelos en la ruta

Si no hay modelos en la ruta, o solamente modelos sin enlaces de modelo, el tipo de despliegue se definirá en Solamente puntuación.

Supervisión de una rama de puntuación para detectar errores

Cuando se designa la rama de puntuación, se supervisa a fin de detectar errores, como la ausencia de un nodo Enterprise View en la ruta al desplegarla como escenario.

Si se detecta un error, se resalta la rama de puntuación en el color correspondiente a errores de dicha rama, y se muestra un mensaje de error. Puede definir el color de errores mediante la opción Color personalizado. Consulte el tema “Opciones de configuración de representación” en la página 154 para obtener más información.

Si se detecta un error, proceda de la siguiente manera:

1. Corrija el error según el contenido del mensaje.
2. Realice una de las siguientes acciones:
 - Pulse con el botón derecho en el nodo terminal y pulse en **Supervisar escenario** en el menú emergente.
 - En el menú principal, pulse en:
Herramientas > Propiedades de ruta > Despliegue
y pulse en **Comprobar**.
3. Si es necesario, repita el proceso hasta que dejen de detectarse errores.

Capítulo 10. Exportación a aplicaciones externas

Acerca de Exportación a aplicaciones externas

IBM SPSS Modeler ofrece diversos mecanismos para exportar todo el proceso de minería de datos a aplicaciones externas para que el trabajo que desarrolla para la preparación de datos y generación de modelos pueda aprovecharse también fuera de IBM SPSS Modeler.

La sección anterior mostraba cómo puede desplegar rutas en un repositorio de IBM SPSS Collaboration and Deployment Services para aprovechar su acceso multiusuario, su programación de trabajos y otras características. De forma análoga, las rutas de IBM SPSS Modeler también pueden utilizarse en conjunción con:

- IBM SPSS Modeler Advantage
- Aplicaciones que pueden importar y exportar archivos en formato PMML

Si desea obtener más información sobre uso de rutas con IBM SPSS Modeler Advantage, consulte “Apertura de una ruta en IBM SPSS Modeler Advantage”.

Para obtener información sobre la exportación e importación de modelos como archivos PMML, lo que hace posible compartir modelos con otras aplicaciones que admitan este formato, consulte “Cómo importar y exportar modelos como PMML” en la página 140.

Apertura de una ruta en IBM SPSS Modeler Advantage

Las rutas de IBM SPSS Modeler se pueden utilizar en conjunción con la aplicación IBM SPSS Modeler Advantage de baja intensidad. Aunque es posible crear aplicaciones personalizadas completas con IBM SPSS Modeler Advantage, también puede utilizar una ruta ya creada en IBM SPSS Modeler como base de un flujo de trabajo de aplicaciones.

Para abrir una ruta en IBM SPSS Modeler Advantage:

1. Despliegue la ruta en el repositorio de IBM SPSS Collaboration and Deployment Services, asegurándose de pulsar la opción **Desplegar como ruta**. Consulte el tema “Despliegue de rutas” en la página 132 para obtener más información.
2. Pulse en el botón de la barra de herramientas Abrir en IBM SPSS Modeler Advantage o en el menú principal, pulse:

Archivo > Abrir en IBM SPSS Modeler Advantage

1. Especifique los valores de conexión con el repositorio si es necesario. Consulte el tema “Conexión con el repositorio” en la página 121 para obtener más información. Si desea obtener información sobre el puerto concreto, la contraseña y otros datos de conexión, póngase en contacto con el administrador del sistema local.

Note: El servidor también debe tener el software de IBM SPSS Modeler Advantage instalado.

1. En el cuadro de diálogo Repositorio: Almacenar, seleccione la carpeta donde desea almacenar el objeto, especifique cualquier otra información que desee registrar y pulse en el botón **Almacenar**. Consulte el tema “Configuración de propiedades de objeto” en la página 122 para obtener más información.

De este modo también se inicia IBM SPSS Modeler Advantage con la ruta ya abierta. La ruta está cerrada en IBM SPSS Modeler.

Cómo importar y exportar modelos como PMML

PMML, o lenguaje de códigos para modelos predictivos, es un formato XML para describir modelos estadísticos y de minería de datos, incluyendo entradas a modelos, transformaciones utilizadas para preparar los datos para minería de datos, y los parámetros que definen los propios modelos. IBM SPSS Modeler importa y exporta PMML, con lo que se permite compartir modelos con otras aplicaciones que admitan este formato, como IBM SPSS Statistics.

Si desea obtener más información sobre PMML, consulte el sitio Web del grupo de minería de datos (<http://www.dmg.org>).

Para exportar un modelo

La mayoría de tipos de modelos generados por IBM SPSS Modeler admite la exportación PMML. Consulte el tema “Tipos de modelos que admiten PMML” para obtener más información.

1. Pulse con el botón derecho del ratón en un nugget en la paleta de modelos. (también puede pulsar dos veces un nugget de modelo en el lienzo y seleccionar el menú Archivo.)
2. En el menú, pulse **Exportar PMML**.
3. En el cuadro de diálogo Exportar (o Guardar), especifique un directorio objetivo y un nombre exclusivo para el modelo.

Note: Puede cambiar las opciones de exportación PMML en el cuadro de diálogo Opciones de usuario. En el menú principal, pulse en:

Herramientas > Opciones > Opciones de usuario

y pulse la pestaña PMML.

Consulte el tema “Opciones de configuración de exportación de PMML” en la página 155 para obtener más información.

Para importar un modelo guardado como PMML

Los modelos exportados como PMML desde IBM SPSS Modeler o cualquier otra aplicación se pueden importar a la paleta de modelos. Consulte el tema “Tipos de modelos que admiten PMML” para obtener más información.

1. En la paleta de modelos, pulse con el botón derecho en la paleta y seleccione **Importar PMML** del menú.
2. Seleccione el archivo que desea importar y especifique las opciones de las etiquetas de valores y variables como desee.
3. Pulse en **Abrir**.

Utilice las etiquetas de variables si están presentes en el modelo. El lenguaje PMML puede especificar tanto nombres de variables como etiquetas de variables (como ID de referencia para *IDRef*) para las variables del diccionario de datos. Seleccione esta opción para utilizar etiquetas de variables si están presentes en el PMML exportado originalmente.

Si ha seleccionado las opciones anteriores de etiqueta pero en el PMML no hay ninguna etiqueta de variable o de valor, entonces los nombres de variables y valores literales se utilizarán como normales.

Tipos de modelos que admiten PMML

Exportación de PMML

IBM SPSS Modeler modelos ALSCAL asimétricos. Los siguientes modelos creados en IBM SPSS Modeler pueden exportarse como PMML 4.0:

- Árbol C&R
- QUEST
- CHAID
- Regresión lineal
- Red neuronal
- C5.0
- Regresión Logística
- Genlin
- SVM
- A priori
- Carma
- K-medias
- Kohonen
- Dos fases
- GLMM (soporte únicamente para modelos GLMM de solo efecto fijo)
- Lista de decisiones
- Cox
- Secuencia (no se admite la puntuación para modelos PMML de secuencia)
- Estadísticas Modelo

Modelos nativos de bases de datos. Para modelos generados mediante algoritmos nativos de bases de datos, la exportación PMML está disponible solamente para modelos de IBM InfoSphere Warehouse. Los modelos creados mediante Analysis Services desde Microsoft o Oracle Data Miner no se pueden exportar. Tenga en cuenta también que los modelos IBM exportados como PMML no se pueden volver a importar a IBM SPSS Modeler.

Importación de PMML

IBM SPSS Modeler puede importar y puntuar modelos PMML generados por versiones actuales de todos los productos de IBM SPSS Statistics, incluidos los modelos exportados desde IBM SPSS Modeler, así como cualquier modelo o transformación PMML generado mediante IBM SPSS Statistics 17.0 o posterior. Básicamente, esto significa cualquier PMML que pueda puntuar el motor de puntuación, con las siguientes excepciones:

- Los modelos Apriori, CARMA, de detección de anomalías y de secuencia no pueden importarse.
- Es posible que no pueda navegar por los modelos de PMML después de importar a IBM SPSS Modeler aunque se puedan utilizar para la puntuación. (Tenga en cuenta que esto incluye los modelos que se exportaron de IBM SPSS Modeler para comenzar. Para evitar esta limitación, exporte el modelo como un archivo del modelo generado [**.gm*] en lugar de como PMML.)
- Los modelos de IBM InfoSphere Warehouse exportados como PMML no se pueden importar.
- La validación limitada se produce al importar, pero la validación completa se realiza al intentar puntuar el modelo. Por lo tanto es posible que la importación sea correcta pero que la puntuación falle o genere resultados incorrectos.

Capítulo 11. Proyectos e informes

Introducción a los proyectos

Un **proyecto** es un grupo de archivos relacionados con una tarea de minería de datos. Los proyectos incluyen rutas de datos, diagramas, modelos generados, informes y cualquier otro elemento creado en IBM SPSS Modeler. A primera vista parece que los proyectos de IBM SPSS Modeler son simplemente una forma de organizar resultados, pero en realidad pueden hacer mucho más. Con los proyectos puede:

- Anotar cada objeto en el archivo de proyecto.
- Usar la metodología CRISP-DM para orientar sus esfuerzos en la minería de datos. Los proyectos también contienen un sistema de ayuda de CRISP-DM en el que se ofrece información detallada y ejemplos reales de minería de datos con CRISP-DM.
- Añadir objetos externos a IBM SPSS Modeler al proyecto, como una presentación de diapositivas de PowerPoint que se usa para presentar los objetivos de la minería de datos o notas técnicas sobre los algoritmos que se pretenden utilizar.
- Generar informes de actualización exhaustivos o sencillos basados en las anotaciones. Estos informes se pueden generar en HTML para publicarlos fácilmente en la intranet de su organización.

Note: Si el panel de proyectos no está visible en la ventana de IBM SPSS Modeler, pulse **Proyecto** en el menú Ver.

Los objetos que se añaden a un proyecto se pueden ver de dos formas: **vista Clases view** y **vista CRISP-DM**. Cualquier elemento que se añada a un proyecto se incorporará a ambas vistas; puede pasar de una a otra para crear la organización que le resulte más adecuada.

Vista CRISP-DM

Al ser compatible con el estándar CRISP-DM (Cross-Industry Standard Process for Data Mining), los proyectos de IBM SPSS Modeler ofrecen una forma independiente y probada en el sector para la organización de los esfuerzos invertidos en la minería de datos. CRISP-DM emplea seis fases para describir el proceso desde el comienzo (recopilación de necesidades empresariales) hasta el final (despliegue de los resultados). Incluso pese a que algunas fases no impliquen generalmente trabajar en IBM SPSS Modeler, el panel de proyectos incluye las seis fases que permiten disponer de una ubicación central para el almacenamiento y el seguimiento de todos los materiales asociados con el proyecto. Por ejemplo, la fase Comprensión del negocio implica generalmente la recopilación de necesidades y reuniones con colegas para determinar objetivos en lugar de trabajar con datos en IBM SPSS Modeler. El panel de proyectos permite almacenar las notas tomadas en estas reuniones en la carpeta *Comprensión del negocio* para tenerlas como futura referencia e incluirlas en los informes.

La vista CRISP-DM del panel de proyectos también incorpora su propio sistema de ayuda, que le guiará a través del ciclo vital de la minería de datos. Desde IBM SPSS Modeler, se puede acceder a esta ayuda pulsando **Ayuda de CRISP-DM** en el menú Ayuda.

Note: Si el panel de proyectos no está visible en la ventana, pulse **Proyecto** en el menú Ver.

Establecimiento de la fase del proyecto predeterminada

Los objetos añadidos a un proyecto se incorporan en una fase predeterminada de CRISP-DM. Esto quiere decir que necesita organizar los objetos manualmente de acuerdo con la fase de minería de datos en la que los haya usado. Por tanto, es conveniente establecer la carpeta predeterminada en la fase en la que esté trabajando actualmente.

Para seleccionar la fase que desea usar como predeterminada:

1. En la vista CRISP-DM, pulse con el botón de derecho del ratón en la carpeta cuya fase desea establecer como valor predeterminado.
2. En el menú, pulse **Establecer como valor predeterminado**.

La carpeta predeterminada se mostrará con letras en negrita.

Vista Clases

La vista Clases del panel de proyectos permite organizar el trabajo en IBM SPSS Modeler de forma categórica por los tipos de los objetos que se hayan creado. Los objetos guardados se pueden añadir a cualquiera de las siguientes categorías:

- Rutas
- Nodos
- Modelos
- Tablas, gráficos e informes
- Otros (archivos externos a IBM SPSS Modeler, como presentaciones de diapositivas o notas técnicas sobre su trabajo de minería de datos)

Al añadir objetos a la vista Clases, también se añaden a la carpeta de fase predeterminada de la vista CRISP-DM.

Note: Si el panel de proyectos no está visible en la ventana, pulse **Proyecto** en el menú Ver.

Generación de un proyecto

Un proyecto es, esencialmente, un archivo que contiene referencias a todos los archivos que se asocien con él. Esto quiere decir que los elementos de proyectos se almacenan de forma individual y como referencia en el archivo de proyecto (.cpj). Debido a esta estructura referencial, tenga en cuenta lo siguiente:

- Los elementos de proyectos se deben guardar primero de forma individual antes de añadirse a un proyecto. Si un elemento no se guarda, se le solicitará que lo haga antes de añadirlo al proyecto actual.
- Los objetos que se actualizan individualmente, como las rutas, también se actualizan en el archivo de proyecto.
- Al mover o eliminar objetos manualmente (como rutas, nodos y objetos de resultados) del sistema de archivos, los enlaces del archivo de proyecto serán no válidos.

Creación de un nuevo proyecto

Es muy fácil crear nuevos proyectos en la ventana de IBM SPSS Modeler. Puede comenzar a crear uno, si no hay ninguno abierto, o cerrar un proyecto existente y comenzar desde cero.

En el menú principal, pulse en:

Archivo > Proyecto > Nuevo proyecto...

Adición a un proyecto

Una vez que ha creado o abierto un proyecto, puede añadir objetos como rutas de datos, nodos e informes mediante diferentes métodos.

Adición de objetos desde los gestores

El uso de gestores en la esquina superior derecha de la ventana de IBM SPSS Modeler permite la adición de rutas o resultados.

1. Seleccione un objeto, como una tabla o una ruta, de una de las pestañas de gestores.

2. Pulse con el botón derecho y elija **Añadir al proyecto**.

Si el objeto se ha guardado previamente, se añadirá automáticamente a la carpeta de objetos apropiada (en la vista Clases) o a la carpeta de fase predeterminada (en la vista CRISP-DM).

3. Si lo prefiere, también puede arrastrar y soltar objetos de los gestores al panel de proyectos.

Note: Es posible que se le solicite que guarde antes el objeto. Al hacerlo, asegúrese de seleccionar **Añadir archivo al proyecto** en el cuadro de diálogo Guardar. Así se añadirá el objeto automáticamente al proyecto después de guardarlo.

Adición de nodos desde el lienzo

Se pueden añadir nodos individuales desde el lienzo de rutas mediante el cuadro de diálogo Guardar.

1. Seleccione un nodo en el lienzo.
2. Pulse con el botón derecho y seleccione **Guardar nodo**. En el menú principal, también se puede pulsar en:
Editar > Nodo > Guardar nodo...
3. En el cuadro de diálogo Guardar, seleccione **Añadir archivo al proyecto**.
4. Asigne un nombre al nodo y pulse en **Guardar**.

Así se guarda el archivo y se añade al proyecto. Los nodos se añaden a la carpeta *Nodos* en la vista Clases y a la carpeta de fase predeterminada en la vista CRISP-DM.

Adición de archivos externos

Si lo desea puede añadir una gran variedad de objetos externos a IBM SPSS Modeler en un proyecto. Esto puede resultar útil en la gestión de todo el proceso de minería de datos realizado en IBM SPSS Modeler. Por ejemplo, puede almacenar enlaces a datos, notas, presentaciones y gráficos en un proyecto. En la vista CRISP-DM se pueden añadir archivos externos a la carpeta que se desee. En la vista Clases se pueden añadir archivos externos sólo a la carpeta *Otros*.

Para añadir archivos externos a un proyecto:

1. Arrastre los archivos del escritorio al proyecto.
o
2. Pulse con el botón derecho en la carpeta de destino en las vistas CRISP-DM o Clases.
3. En el menú, pulse **Añadir a carpeta**.
4. Seleccione un archivo en el cuadro de diálogo y pulse en **Abrir**.

Se añadirá una referencia al objeto seleccionado dentro de los proyectos de IBM SPSS Modeler.

Transferencia de proyectos a IBM SPSS Collaboration and Deployment Services Repository

Puede transferir un proyecto entero, incluyendo todos los archivos de componente, a IBM SPSS Collaboration and Deployment Services Repository en un paso. No se moverá ningún objeto que ya esté en la ubicación objetivo. Esta característica también funciona a la inversa; puede transferir proyectos enteros desde IBM SPSS Collaboration and Deployment Services Repository al sistema de archivos local.

Note: Es necesario disponer de una licencia independiente para acceder a un repositorio IBM SPSS Collaboration and Deployment Services. Si desea obtener más información, consulte <http://www.ibm.com/software/analytics/spss/products/deployment/cds/>

Transferencia de un proyecto

Asegúrese de que el proyecto que desea transferir está abierto en el panel de proyectos.

Para transferir un proyecto:

1. Pulse con el botón derecho del ratón en la carpeta raíz del proyecto y pulse **Transferir proyecto**.
2. Si se le solicita, inicie sesión en IBM SPSS Collaboration and Deployment Services Repository.
3. Especifique la nueva ubicación para el proyecto y pulse en **Aceptar**.

Configuración de las propiedades de un proyecto

Puede personalizar el contenido y la documentación de un proyecto a través del cuadro de diálogo de propiedades de proyecto. Para acceder a las propiedades del proyecto:

1. Pulse con el botón derecho en un objeto o una carpeta en el panel de proyectos y elija **Propiedades de proyecto**.
2. Pulse en la pestaña **Proyecto** para especificar información básica del proyecto.

Creado. Muestra la fecha de creación del proyecto (no editable).

Resumen. Puede escribir un resumen del proyecto de minería de datos que se mostrará en el informe del proyecto.

Contenido. Enumera el tipo y el número de componentes a los que hace referencia el archivo de proyecto (no editable).

Guardar objeto no guardado como. Especifica si los objetos que no se han guardado deben guardarse en el sistema de archivos local o, en su lugar, en el repositorio. Consulte el tema “Acerca de IBM SPSS Collaboration and Deployment Services Repository” en la página 119 para obtener más información.

Actualizar referencias del objeto al cargar el proyecto. Seleccione esta opción para actualizar las referencias del proyecto a sus componentes. *Note:* Los archivos añadidos a un proyecto no se guardan en el propio archivo de proyecto. En su lugar, se almacena una referencia a los archivos en el proyecto. Esto quiere decir que cualquier desplazamiento o eliminación de un archivo conllevará la eliminación de dicho objeto del proyecto.

Anotaciones de un proyecto

El panel de proyectos incluye una gran variedad de formas de anotar los esfuerzos invertidos en la minería de datos. Las anotaciones en el nivel de proyecto se usan a menudo para hacer un seguimiento de objetivos y decisiones generales, mientras que las anotaciones de los nodos o las carpetas proporcionan un detalle adicional. La pestaña Anotaciones proporciona espacio suficiente para documentar detalles en el nivel de proyecto, como la exclusión de datos con contenido perdido irrecuperable o hipótesis halagüeñas formadas durante la exploración de los datos.

Para anotar un proyecto:

1. Seleccione la carpeta del proyecto adecuada en las vistas CRISP-DM o Clases.
2. Pulse con el botón derecho en la carpeta y a continuación en **Propiedades de proyecto**.
3. Pulse en la pestaña **Anotaciones**.
4. Introduzca las palabras clave y el texto para describir el proyecto.

Anotaciones y propiedades de las carpetas

Se pueden añadir notas en las carpetas de proyecto individuales de las vistas CRISP-DM o Clases. En la vista CRISP-DM podría tratarse de una forma extremadamente efectiva de documentar los objetivos de su organización en cada fase de la minería de datos. Por ejemplo, el uso de la herramienta de anotación en la carpeta *Comprensión del negocio* le permite incluir información como "El objetivo empresarial de este estudio es reducir el abandono de clientes importantes." Posteriormente, este texto se puede incluir automáticamente en el proyecto seleccionando la opción **Incluir en informe**.

Para anotar una carpeta:

1. Seleccione una carpeta en el panel de proyectos.
2. Pulse con el botón derecho en la carpeta y a continuación en **Propiedades de carpeta**.

En la vista CRISP-DM, las carpetas se anotan con un resumen del propósito de cada fase, así como a modo de guía para la realización de las tareas de minería de datos relevantes. Cualquiera de las anotaciones es susceptible de ser eliminada o modificada.

Nombre. Esta área muestra el nombre del campo seleccionado.

Texto de información sobre herramientas. Cree información personalizada sobre herramientas que se mostrará al colocar el puntero del ratón sobre una carpeta del proyecto. Esto es útil en la vista CRISP-DM, por ejemplo, para detallar los conceptos básicos de los objetivos de cada fase o para marcar el estado de una fase como "En curso" o "Finalizada".

Campo de anotación. Use este campo para incluir anotaciones más extensas que se pueden incorporar en el informe del proyecto. En la vista CRISP-DM se incluye una descripción de cada fase de la minería de datos en la anotación pero puede personalizarla, si lo desea, de acuerdo con las características de su proyecto.

Incluir en informe. Para incluir las anotaciones en el informe, seleccione **Incluir en informe**.

Propiedades de objeto

Se pueden ver las propiedades de los objetos y seleccionar si se desea incluir objetos concretos en el informe del proyecto. Para acceder a las propiedades de un objeto:

1. Pulse con el botón derecho en un objeto en el panel del proyecto.
2. En el menú, pulse **Propiedades de objeto**.

Nombre. Esta área indica el nombre del objeto guardado.

Ruta de acceso. Esta área indica la ubicación del objeto guardado.

Incluir en informe. Seleccione esta opción para incluir los detalles del objeto en un informe generado.

Cierre de un proyecto

Si selecciona salir de IBM SPSS Modeler o abrir un proyecto nuevo, el archivo de proyecto existente (.cpj) se cierra.

Algunos archivos asociados al proyecto (como rutas, nodos o gráficos) pueden seguir abiertos. Si desea dejar estos archivos abiertos, responda No al mensaje ... **¿Desea guardar y cerrar estos archivos?**

Si modifica y guarda cualquier archivo asociado después de cerrar un proyecto, las versiones actualizadas de éstos se incluirán en el proyecto la próxima vez que se abra. Para evitarlo, elimine el archivo del proyecto o guárdelo con un nombre diferente.

Generación de un informe

Una de las características más útiles de los proyectos es la capacidad de generar informes sobre los elementos y las anotaciones del proyecto. Se trata de un componente crucial en la minería de datos eficaz, como se destaca en la metodología CRISP-DM. Puede generar un informe directamente en un archivo de una serie de tipos de archivos o en una ventana de resultados en pantalla para una visualización inmediata. Desde ahí, puede imprimir, guardar o ver el informe en un explorador de Web. Puede distribuir los informes guardados a otras personas en su organización.

A veces, los informes de archivos de proyecto se generan varias veces durante el proceso de minería de datos para su distribución entre los participantes en el proyecto. El informe reúne información sobre los objetos referenciados desde el archivo de proyecto, así como cualquier anotación que se haya creado. Puede crear informes basados en las vistas Clases o CRISP-DM.

Para generar un informe:

1. Seleccione la carpeta del proyecto adecuada en las vistas CRISP-DM o Clases.
2. Pulse con el botón derecho en la carpeta y a continuación en **Informe de proyecto**.
3. Especifique las opciones del informe y pulse en **Generar informe**.

Las opciones del cuadro de diálogo del informe ofrecen varias posibilidades para generar el tipo de informe que necesita:

Nombre del resultado. Especifique el nombre de la ventana de resultados si elige enviar el resultado del informe a la pantalla. Puede especificar un nombre personalizado o dejar que IBM SPSS Modeler asigne automáticamente nombre a la ventana.

Salida a pantalla. Seleccione esta opción para generar y mostrar el informe en una ventana de resultados. Tenga en cuenta que tiene la opción de exportar el informe a varios tipos de archivos desde la ventana de resultados.

Salida a archivo. Seleccione esta opción para generar y guardar el informe como un archivo del tipo especificado en la lista Tipo de archivo.

Nombre de archivo. Especifique un nombre de archivo para el informe generado. Los archivos se guardan de forma predeterminada en el directorio `\bin` de IBM SPSS Modeler. Utilice el botón de puntos suspensivos (...) para especificar una ubicación diferente.

Tipo de archivo. Los tipos de archivo disponibles son:

- **Documento HTML.** El informe se guarda como un archivo HTML único. Si su informe contiene gráficos, se guardan como archivos PNG y son referenciados por el archivo HTML. Al publicar el informe en Internet, asegúrese de cargar el archivo HTML y cualquier imagen a la que hace referencia.
- **Documento de texto.** El informe se guarda como un archivo de texto único. Si su informe contiene gráficos, sólo se incluyen en él las referencias a la ruta y el nombre de archivo.
- **Documento de Microsoft Word.** El informe se guarda como un documento único, con cualquier gráfico incrustado directamente en el documento.
- **Documento de Microsoft Excel.** El informe se guarda como una hoja de cálculo única, con cualquier gráfico incrustado directamente en la hoja de cálculo.
- **Documento de Microsoft Power Point.** Cada fase se muestra en una diapositiva nueva. Cualquier gráfico se incrusta directamente en las diapositivas de Power Point.
- **Objeto de resultados.** Al abrirlo en IBM SPSS Modeler, este archivo (.cou) es igual a la opción **Salida a pantalla** en el grupo **Formato de informe**.

Note: Para exportar a un archivo de Microsoft Office, debe tener la aplicación correspondiente instalada.

Título. Especifique un título para el informe.

Estructura del informe. Seleccione o **CRISP-DM** o **Clases**. La vista CRISP-DM proporciona un informe de estado con una sinopsis a grandes rasgos y detalles acerca de cada fase de la minería de datos. La vista Clases es una vista basada en objetos que es más apropiada para realizar el seguimiento interno de datos y rutas.

Autor. Se muestra el nombre de usuario predeterminado, pero puede cambiarlo.

El informe incluye. Seleccione un método para incluir objetos en el informe. Seleccione **todas las carpetas y objetos** para incluir todos los elementos añadidos al archivo de proyecto. También puede incluir elementos basándose en si en las propiedades de cada objeto se ha seleccionado **Incluir en informe**. Si lo prefiere, para revisar los elementos no incluidos en el informe, puede optar por incluir sólo los elementos marcados para ser excluidos (en los que la opción **Incluir en informe** no está seleccionada).

Seleccionar. Esta opción permite proporcionar actualizaciones del proyecto seleccionando sólo los **elementos recientes** en el informe. También puede hacer un seguimiento de los problemas más antiguos, tal vez aún no resueltos, configurando parámetros para los **elementos antiguos**. Seleccione **todos los elementos** para descartar el tiempo como parámetro del informe.

Ordenar por. Puede seleccionar una combinación de las siguientes características de objetos para ordenarlas dentro de una carpeta:

- **Tipo.** Objetos de grupo por tipo.
- **Nombre.** Organiza los objetos alfabéticamente.
- **Fecha de adición.** Ordena los objetos tomando la fecha en que se añadieron al proyecto.

Almacenamiento y exportación de informes generados

Un informe generado en la pantalla se muestra en una ventana de resultados nueva. Cualquier gráfico incluido en el informe aparece como en las imágenes en línea.

Terminología de informe

El número total de nodos en cada ruta se incluye en el informe. Los números se muestran en las siguientes cabeceras, que utilizan la terminología de IBM SPSS Modeler, no terminología de CRISP-DM.

- **Lectores de datos.** Nodos de origen.
- **Escritor de datos.** Nodos de exportación.
- **Generadores de modelo.** Nodo de generación o modelado.
- **Aplicadores de modelo.** Modelos generados, también conocidos como nugget.
- **Generadores de resultados.** Nodos Gráfico o Resultado.
- **Otro.** Cualquier otro nodo relacionado con el proyecto. Por ejemplo, aquellos disponibles en la pestaña Operaciones con campos o la pestaña Operaciones con registros en la Paleta de nodos.

Para guardar un informe:

1. En el menú Archivo, pulse en **Guardar**.
2. Especifique un nombre de archivo.
El informe se guarda como un objeto de resultado.

Para exportar un informe:

3. En el menú Archivo, pulse en **Exportar** y el tipo de archivo al que quiere exportarlo.
4. Especifique un nombre de archivo.

El informe se guarda en el formato que elija.

Puede exportarlo a los siguientes tipos de archivos:

- HTML
- Texto
- Microsoft Word
- Microsoft Excel
- Microsoft PowerPoint

Note: Para exportar a un archivo de Microsoft Office, debe tener la aplicación correspondiente instalada.

Use los botones de la parte superior de la ventana para:

- Imprimir el informe.
- Ver el informe como HTML en un explorador Web externo.

Capítulo 12. Personalización de IBM SPSS Modeler

Personalización de opciones de IBM SPSS Modeler

Existen diferentes operaciones que puede ejecutar para personalizar IBM SPSS Modeler en función de sus necesidades. Principalmente, la personalización consiste en definir opciones específicas del usuario, como asignación de memoria, directorios predeterminados y uso de sonidos y color. También puede personalizar la paleta de nodos en la parte inferior de la ventana de IBM SPSS Modeler.

Configuración de las opciones de IBM SPSS Modeler

Hay varias formas de personalizar y definir las opciones de IBM SPSS Modeler:

- Configure las opciones del sistema, tales como el uso de memoria y la configuración regional, pulsando en **Opciones de sistema** en el menú **Herramientas > Opciones**.
- Configure las opciones de usuario, como fuentes y colores de visualización, pulsando en **Opciones de usuario** en el menú **Herramientas > Opciones**.
- Especifique la ubicación de aplicaciones que funcionan con IBM SPSS Modeler pulsando en **Aplicaciones de ayuda** en el menú **Herramientas > Opciones**.
- Especifique los directorios predeterminados utilizados en IBM SPSS Modeler pulsando en **Definir directorio** o **Definir directorio de servidor** del menú Archivo.

También puede definir las opciones que se aplican a la totalidad o parte de las rutas. Consulte el tema "Opciones de configuración de las rutas" en la página 41 para obtener más información.

Opciones de sistema

Se puede especificar el idioma preferido o la configuración regional de IBM SPSS Modeler pulsando en **Opciones de sistema** en el menú **Herramientas > Opciones**. Aquí también se puede establecer el uso de memoria máxima de IBM SPSS Modeler. Tenga en cuenta que las modificaciones realizadas en este cuadro de diálogo no tendrán efecto hasta que no reinicie IBM SPSS Modeler.

Memoria máxima. Selecciona la aplicación de un límite en megabytes para el uso de la memoria de IBM SPSS Modeler. En algunas plataformas, IBM SPSS Modeler limita el tamaño del proceso para reducir el número de ordenadores con recursos limitados o las sobrecargas. Si se trabaja con grandes cantidades de datos, es posible que se muestre un mensaje de error de "memoria agotada". Se pueden reducir las sobrecargas de memoria especificando un umbral nuevo.

Uso de la configuración regional del sistema. Esta opción está seleccionada de forma predeterminada y está configurada en inglés (Estados Unidos). Anule la selección para especificar otro idioma de la lista de idiomas y configuraciones regionales disponibles.

Administración de la memoria

Además de la configuración de **Memoria máxima** especificada en el cuadro de diálogo Opciones de sistema, existen muchas otras formas de optimizar el uso de la memoria:

- Ajuste la opción **Máximo número de miembros para campos nominales** del cuadro de diálogo de propiedades de la ruta. Esta opción define un número máximo de miembros para los campos nominales después del cual el nivel de medición del campo se convierte en *Sin tipo*. Consulte el tema "Configuración de opciones generales de las rutas" en la página 42 para obtener más información.
- Fuerce a IBM SPSS Modeler a liberar espacio en memoria pulsando en la esquina inferior derecha de la ventana de IBM SPSS Modeler donde se muestra la memoria utilizada y la cantidad asignada (*xx* MB/*xx* MB). Si se pulsa en esta zona se oscurece el color un momento y, después, las cifras de

asignación de memoria disminuyen. Cuando esta zona vuelva a adquirir su color habitual, IBM SPSS Modeler ya habrá liberado todo el espacio posible en la memoria.

Configuración de directorios predeterminados

Se puede especificar el directorio predeterminado utilizado en los exploradores de archivos y el resultado seleccionando **Definir directorio** o **Definir directorio de servidor** en el menú Archivo.

- **Definir directorio.** Se puede utilizar esta opción para establecer el directorio de trabajo. El directorio de trabajo predeterminado se basa en la ruta de instalación de la versión de IBM SPSS Modeler utilizada o en la ruta de líneas de comandos utilizada para iniciar IBM SPSS Modeler. En el modo local, el directorio de trabajo es la ruta utilizada para todas las operaciones del equipo cliente y los archivos de resultados (en caso de que estén referenciados con rutas relacionadas).
- **Definir directorio de servidor.** La opción Definir directorio de servidor del menú Archivo se activa siempre que existe una conexión con el servidor remoto. Utilice esta opción para especificar el directorio predeterminado de todos los archivos del servidor y de datos especificados para entrada y salida. El directorio de servidor predeterminado es *\$CLEO/data*, donde *\$CLEO* es el directorio en el que se instala la versión del servidor de IBM SPSS Modeler. También se puede sobrescribir esta opción predeterminada con la línea de comandos, a través de la marca `-directorio_servidor` con el argumento de la línea de comandos `modelerclient`.

Opciones de configuración de usuario

Se pueden configurar las opciones de IBM SPSS Modeler seleccionando **Opciones de usuario** en el menú **Herramientas > Opciones**. Estas opciones se aplican a todas las rutas utilizadas en IBM SPSS Modeler.

Los siguientes tipos de opciones se pueden configurar pulsando en la pestaña correspondiente:

- Opciones de notificación, como la sobrescritura de modelos y los mensajes de error.
- Opciones de representación, como los colores de gráficos y de fondo.
- Opciones de visualización de color de sintaxis.
- Las opciones de exportación de PMML se utilizan para exportar modelos como lenguaje de códigos para modelos predictivos (PMML).
- Información de autor o usuario, como el nombre, las iniciales y la dirección de correo electrónico. Esta información se puede visualizar en la pestaña Anotaciones para nodos y para otros objetos que cree.

Para configurar opciones específicas de rutas, como los separadores decimales, los formatos de hora y datos, optimización, diseño de rutas y scripts de ruta, utilice el cuadro de diálogo de propiedades de la ruta, disponible en los menús Archivo y Herramientas.

Opciones de configuración de notificación

Se pueden configurar distintas opciones relacionadas con las instancias y el tipo de advertencias y con las ventanas de confirmación mediante la pestaña Notificaciones del cuadro de diálogo Opciones de usuario de IBM SPSS Modeler. También se puede especificar el comportamiento de las pestañas Resultado y Modelos en el panel de gestores cuando se generan nuevos modelos y resultados.

Mostrar cuadro de diálogo de comentarios de ejecución de la ruta. Seleccione esta opción para que se abra un cuadro de diálogo con un indicador de progreso cuando una ruta se haya estado ejecutando durante tres segundos. El cuadro de diálogo también incluye los detalles de los objetos de resultados que crea la ruta.

- **Cerrar el cuadro de diálogo al terminar.** De forma predeterminada, el cuadro de diálogo se cierra cuando la ruta finaliza su ejecución. Cancele la selección de esta casilla de verificación si desea que el cuadro de diálogo siga siendo visible cuando finalice la ejecución de la ruta.

Avisar cuando un nodo sobrescriba un archivo. Seleccione la aparición de advertencias con un mensaje de error cuando las operaciones de nodos sobrescriben un archivo existente.

Avisar cuando un nodo sobrescriba una tabla de la base de datos. Seleccione la aparición de advertencias con un mensaje de error cuando las operaciones de nodos sobrescriben una base de datos existente.

Sonidos de notificación

Utilice la lista para especificar si se utilizan sonidos para notificar cuándo se produce un evento o un error. Hay varios tipos de sonido disponibles. Utilice el botón Reproducir (altavoz) para reproducir un sonido seleccionado. Utilice el botón de puntos suspensivos (...) para buscar y seleccionar un sonido.

Note: Los archivos *.wav* utilizados para crear sonidos en IBM SPSS Modeler se almacenan en el directorio */media/sounds* de la instalación.

- **Silenciar todos los sonidos.** Seleccione la desactivación de las notificaciones con sonido para todos los eventos.

Notificaciones visuales

Las opciones de este grupo se utilizan para especificar el comportamiento de las pestañas Resultados y Modelos en el panel de gestores de la parte superior derecha de la pantalla cuando se generan nuevos elementos. Seleccione **Nuevo modelo** o **Nuevo resultado** en la lista para especificar el comportamiento de la pestaña correspondiente.

Las siguientes opciones están disponibles para **Nuevo modelo**:

Añadir modelo a ruta. Si está seleccionada (valor predeterminado), añade un nuevo modelo a la ruta, así como a la pestaña Modelos, en el momento de la generación del modelo. En la ruta, el modelo se muestra con un enlace al nodo de modelado desde el que se creó el modelo. Si no marca esta casilla, el modelo se añade únicamente a la pestaña Modelos.

Sustituir modelo anterior. Si está seleccionada (valor predeterminado), sobrescribe un modelo existente desde esta ruta en la pestaña Modelos y en el lienzo de rutas. Si esta casilla no está marcada, el modelo se añade a los modelos existentes en la pestaña y en el lienzo. Tenga en cuenta que se ajuste se sobrescribe con el ajuste de sustitución de modelo en un enlace de modelo.

Las siguientes opciones están disponibles para **Nuevo salida**:

Avisar cuando los resultados superen [n]. Seleccione la aparición de una advertencia si el número de elementos de la pestaña Resultados supera una cantidad especificada previamente. La cantidad predeterminada es 20; sin embargo, puede cambiar este ajuste si fuera necesario.

Las siguientes opciones están disponibles en todos los casos:

Seleccionar pestaña. Seleccionar si se cambia a las pestañas Resultados y Modelos cuando se genera el objeto correspondiente durante la ejecución de la ruta.

- Seleccionar **Siempre** para cambiar a la pestaña correspondiente en el panel de gestores.
- Seleccionar **Si la ha generado la ruta actual** para cambiar a la pestaña correspondiente sólo para los objetos generados por la ruta visibles en el lienzo.
- Seleccione **Nunca** para restringir que el software cambie a la pestaña correspondiente para notificar los resultados o los modelos generados.

Hacer parpadear pestaña. Seleccionar que la pestaña Resultados o Modelos parpadee en el panel de gestores cuando se generan nuevos resultados o modelos.

- Seleccionar **Si no está seleccionada** para hacer parpadear la pestaña correspondiente (si no se ha seleccionado aún) cuando se generen objetos nuevos en el panel de gestores.

- Seleccionar **Nunca** para restringir que el software produzca parpadeos en la pestaña correspondiente con el fin de notificar los objetos generados.

Desplazar paleta para poder mostrarlo (sólo Nuevo modelo). Seleccionar el desplazamiento automático de la pestaña Modelos en el panel de gestores para que se muestre el modelo más reciente.

- Seleccionar **Siempre** para permitir el desplazamiento.
- Seleccionar **Si la ha generado la ruta actual** para desplazar sólo los objetos generados por la ruta visibles en el lienzo.
- Seleccionar **Nunca** para restringir que el software desplace automáticamente la pestaña Modelos.

Abrir ventana (sólo Nuevo resultado). Seleccione la apertura automática de una ventana de resultados al generarlos.

- Seleccionar **Siempre** para que siempre se abra la ventana de nuevos resultados.
- Seleccionar **Si la ha generado la ruta actual** para abrir una ventana nueva de los resultados que ha generado la ruta visible en el lienzo.
- Seleccionar **Nunca** para restringir que el software abra automáticamente nuevas ventanas de resultados generados.

Pulse en **Valores predeterminados** para invertir la configuración del sistema predeterminada de esta pestaña.

Opciones de configuración de representación

Se pueden configurar las opciones de representación de las fuentes y los colores en IBM SPSS Modeler mediante la pestaña Representación del cuadro de diálogo Opciones de usuario.

Mostrar cuadro de diálogo al inicio. Seleccione esta opción para que el cuadro de diálogo de bienvenida aparezca al inicio. El cuadro de diálogo de bienvenida incluye opciones para iniciar el tutorial de ejemplos de la aplicación, abrir una ruta de demostración o una ruta o proyecto existente o crear una nueva ruta.

Mostrar marcaciones de ruta y supernodo. Si está seleccionada, hace que la marcación (si la hay) en rutas o supernodos se muestre de forma predeterminada. La marcación incluye los comentarios de rutas, los enlaces de modelos y el resaltado de las ramas de puntuación.

Fuentes y colores estándar (tendrán efecto al reiniciar). Las opciones de este cuadro de control se utilizan para especificar el diseño de la pantalla de IBM SPSS Modeler, el esquema de colores y el tamaño de las fuentes. Las opciones seleccionadas aquí no entran en vigor hasta que cierre y reinicie IBM SPSS Modeler.

- **Aspecto.** Le permite seleccionar un esquema de colores y un diseño de pantalla estándar. Puede seleccionar:
 - SPSS Standard** (valor predeterminado), un diseño común en todos los productos de IBM SPSS.
 - SPSS Classic**, un diseño familiar para los usuarios de versiones anteriores de IBM SPSS Modeler.
 - Windows**, un diseño de Windows que puede resultar útil cuando se aumenta el contraste en el lienzo de rutas y en las paletas.
- **Tamaño de fuente predeterminado para nodos.** Especifique un tamaño de fuente para utilizarlo en las paletas de los nodos y para los nodos mostrados en el lienzo de rutas.

Note: Puede establecer el tamaño de los iconos de nodo para una ruta en el panel Diseño de la pestaña Opciones del cuadro de diálogo de propiedades de ruta. En el menú principal, elija:

Herramientas > Propiedades de ruta > Opciones > Diseño.

Colores personalizados. Esta tabla muestra los colores seleccionados actualmente utilizados para varios elementos de representación. Para cada uno de los elementos de la tabla, puede cambiar el color actual pulsando dos veces la fila correspondiente en la columna **Color** y seleccionando un color de la lista. Para especificar un color personalizado, desplácese hasta la parte inferior de la lista y pulse la entrada **Color**.

Orden de colores de categorías gráficas. Esta tabla indica los colores seleccionados actualmente para la representación de gráficos recién creados. El orden de los colores refleja el orden en el que se van a utilizar en el gráfico. Por ejemplo, si un campo nominal utilizado como una superposición de colores contiene cuatro valores exclusivos, sólo se utilizarán los cuatro primeros colores de la lista. Para cada uno de los elementos de la tabla, puede cambiar el color actual pulsando dos veces la fila correspondiente en la columna **Color** y seleccionando un color de la lista. Para especificar un color personalizado, desplácese hasta la parte inferior de la lista y pulse la entrada **Color**. Los cambios realizados no afectan a los gráficos creados anteriormente.

Pulse en **Valores predeterminados** para invertir la configuración del sistema predeterminada de esta pestaña.

Definición de opciones de visualización de sintaxis

Utilizando la pestaña Sintaxis del cuadro de diálogo Opciones de usuario, puede definir opciones para los atributos de font y mostrar colores en los scripts que cree en IBM SPSS Modeler.

Resaltado de la sintaxis. Esta tabla lista los colores seleccionados actualmente para distintos elementos de sintaxis, incluidos el font y la ventana en la que se visualizan. Para cada uno de los elementos listados en la tabla, puede cambiar el color pulsando la lista desplegable correspondiente en la fila y seleccionando un color en la lista. Además, para los elementos de font, puede elegir añadir negrita y cursiva para dar énfasis.

Vista previa. Esta tabla muestra una visualización de sintaxis de ejemplo que utiliza los colores y atributos de font que seleccione en la tabla **Resaltado de la sintaxis**. Esta vista previa se actualiza tan pronto como cambia alguna selección.

Pulse en **Valores predeterminados** para invertir la configuración del sistema predeterminada de esta pestaña.

Opciones de configuración de exportación de PMML

En la pestaña PMML, se puede controlar el modo en que IBM SPSS Modeler exporta modelos como lenguaje de códigos para modelos predictivos (PMML). Consulte el tema “Cómo importar y exportar modelos como PMML” en la página 140 para obtener más información.

Exportar PMML. Aquí se pueden configurar las variaciones de PMML que funcionan mejor con la aplicación objetivo.

- Seleccione **Con extensiones** para permitir las extensiones PMML en los casos especiales donde no hay PMML estándar equivalente. Tenga en cuenta que en la mayoría de los casos esto producirá el mismo resultado que PMML estándar.
- Seleccione **Como PMML estándar...** para exportar PMML que se adhiera lo máximo posible al PMML estándar.

Opciones PMML estándar. Cuando se selecciona la opción **Como PMML estándar...**, puede elegir una de las dos formas válidas de exportar modelos de regresión lineal y logística.

- Como modelos PMML <Regresión general>
- Como modelos PMML <Regresión>

Si desea obtener más información sobre PMML, consulte el sitio Web del grupo de minería de datos <http://www.dmg.org>.

Información de configuración de usuario

Información de usuario/autor. La información que introduzca aquí se puede visualizar en la pestaña Anotaciones de los nodos y otros objetos que cree.

Personalización de la paleta de nodos

Las rutas se generan mediante nodos. La paleta de nodos que se encuentra en la parte inferior de la ventana de IBM SPSS Modeler contiene todos los nodos posibles empleados para la generación de rutas. Consulte el tema “Paleta de nodos” en la página 16 para obtener más información.

Puede reorganizar la paleta de nodos de dos formas:

- Personalizar el gestor de paletas. Consulte el tema “Personalización del gestor de paletas” para obtener más información.
- Modifique cómo se representarán las pestañas de la paleta que contienen subpaletas en la paleta de nodos. Consulte el tema “Creación de una subpaleta” en la página 158 para obtener más información.

Personalización del gestor de paletas

El gestor de paletas se puede personalizar para acomodar el uso de IBM SPSS Modeler. Por ejemplo, si se analizan frecuentemente datos de serie temporal de una base de datos, es posible que desee asegurarse de que el nodo de origen Base de datos, el nodo Intervalos de tiempo, el nodo Serie temporal y el nodo de gráfico Gráfico de tiempo estén disponibles juntos en una pestaña de paleta exclusiva. El gestor de paletas permite realizar fácilmente estos ajustes creando sus propias pestañas de paleta en la paleta de nodos.

El gestor de paletas permite realizar diferentes tareas:

- Controlar las paletas que aparecen en la paleta de nodos bajo el lienzo de rutas.
- Cambiar el orden en que aparecen las paletas en la paleta de nodos.
- Crear y editar sus propias pestañas de paletas y cualquiera de sus subpaletas asociadas.
- Editar las selecciones del nodo predeterminado en su pestaña Favoritos.

Para acceder al Gestor de paletas, en el menú Herramientas, pulse en **Administrar paletas**.

Nombre de paleta. Aparecen todas las pestañas de paleta disponibles, con independencia de si aparecen en la paleta de nodos o no. Incluye cualquier pestaña de paleta que haya creado. Consulte el tema “Creación de una pestaña de paleta” en la página 157 para obtener más información.

Número de nodos. El número de nodos que aparece en cada pestaña de paleta. Un número alto significa que es posible que encuentre conveniente crear subpaletas para dividir los nodos de la pestaña. Consulte el tema “Creación de una subpaleta” en la página 158 para obtener más información.

¿Mostrar? Seleccione este campo para mostrar la pestaña paleta en la paleta de nodos. Consulte el tema “Visualización de las pestañas de paleta en la paleta de nodos” en la página 157 para obtener más información.

Subpaletas. Para seleccionar las subpaletas que se mostrarán en una pestaña de paleta, resalte el **Nombre de paleta** necesario y pulse en este botón para mostrar el cuadro de diálogo Subpaletas. Consulte el tema “Creación de una subpaleta” en la página 158 para obtener más información.

Restaurar las opciones predeterminadas. Para eliminar completamente todas las modificaciones y adiciones que haya realizado en las configuraciones de las paletas y en las subpaletas y volver a la configuración predeterminada de las paletas, pulse en este botón.

Creación de una pestaña de paleta

Para crear una pestaña de paleta personalizada:

1. En el menú Herramientas, abra el gestor de paletas.
2. A la derecha de la columna *¿Mostrar?*, pulse en el botón Añadir paleta; se abrirá el cuadro de diálogo Crear/editar paleta.
3. Introduzca un **nombre de paleta** exclusivo.
4. En el área **Nodos disponibles**, seleccione el nodo que se añadirá a la pestaña de paleta.
5. Pulse en el botón Añadir nodo con la flecha hacia la derecha para mover el nodo resaltado al área **Nodos seleccionados**. Repita este procedimiento hasta que haya terminado de añadir todos los nodos que desee.

Una vez haya añadido todos los nodos necesarios, puede modificar el orden en el que aparecen en la pestaña de paletas:

6. Utilice los botones de flechas sencillas para subir o bajar los nodos una fila.
7. Utilice los botones de flecha para mover un nodo a la parte superior o inferior de la lista.
8. Para eliminar un nodo de una paleta, resalte el nodo y pulse en el botón Eliminar a la derecha del área **Nodos seleccionados**.

Visualización de las pestañas de paleta en la paleta de nodos

Es posible que existan opciones de IBM SPSS Modeler disponibles que nunca utilice; en este caso, puede utilizar el gestor de paletas para ocultar las pestañas que contienen los nodos.

Para seleccionar las pestañas que se muestran en la paleta de nodos:

1. En el menú Herramientas, abra el gestor de paletas.
2. Mediante las casillas de verificación de la columna *¿Mostrar?*, seleccione si desea incluir u ocultar cada pestaña de paleta.

Para eliminar una pestaña de paleta de forma permanente de la paleta de nodos, resalte el nodo y pulse en el botón Eliminar a la derecha de la columna *¿Mostrar?* de columna. Una vez eliminada, una pestaña de paleta no se puede recuperar.

Note: No puede eliminar las pestañas de paletas predeterminadas de IBM SPSS Modeler, salvo la pestaña Favoritos.

Modificación del orden de visualización de la paleta de nodos

Una vez haya seleccionado las pestañas de paleta que desea visualizar, puede modificar el orden en el que aparecen en la paleta de nodos:

1. Utilice los botones de flechas sencillas para subir o bajar una fila una pestaña de paleta. Si las sube, las moverá a la izquierda de la paleta de nodos y viceversa.
2. Utilice los botones de flecha para mover una pestaña de paleta a la parte superior o inferior de la lista. Las que aparezcan en la parte superior de la lista se mostrarán a la izquierda de la paleta de nodos.

Visualización de subpaletas en una pestaña de paleta

De la misma forma que puede controlar las pestañas de paleta que se muestran en la paleta de nodos, puede controlar las subpaletas que estarán disponibles en la pestaña de la paleta padre.

Para seleccionar las subpaletas que se mostrarán en una pestaña de paletas:

1. En el menú Herramientas, abra el gestor de paletas.
2. Seleccione la paletas que necesite.
3. Pulse en el botón Subpaletas y se abrirá el cuadro de diálogo Subpaletas.

4. Mediante las casillas de verificación de la columna *¿Mostrar?*, seleccione si desea incluir cada subpaleta en la pestaña de paleta. La subpaleta **Todos** se muestra siempre y no se puede eliminar.
5. Para eliminar una subpaleta de la pestaña de paleta de forma permanente, resalte la subpaleta y pulse en el botón Eliminar a la derecha de *¿Mostrar?*.

Note: No puede eliminar las subpaletas predeterminadas que incluye la pestaña de paleta de modelado.

Modificación del orden de visualización de la pestaña de paleta

Una vez haya seleccionado las subpaletas que desea visualizar, puede modificar el orden en el que aparecen en la pestaña de la paleta padre:

1. Utilice los botones de flechas sencillas para subir o bajar una subpaleta una fila.
2. Utilice los botones de flecha para mover una subpaleta a la parte superior o inferior de la lista.

Las subpaletas que cree se mostrarán en la paleta de nodos cuando seleccione su pestaña de paleta padre. Consulte el tema “Modificación de la vista de la pestaña de paleta” para obtener más información.

Creación de una subpaleta

Como puede añadir cualquier nodo existente a las pestañas de paleta personalizada que cree, es posible que seleccione más nodos que se pueden visualizar más fácilmente en la pantalla sin desplazarse. Para evitar desplazarse, puede crear subpaletas en las que colocará los nodos que seleccione en la pestaña de paleta. Por ejemplo, si ha creado una pestaña de paleta que contiene los nodos que utiliza con mayor frecuencia para crear sus rutas, puede crear cuatro subpaletas que contengan las selecciones por nodo de origen, operaciones de campos, modelado y resultado.

Note: Sólo puede seleccionar los nodos de subpaletas que se hayan añadido en la pestaña de la paleta padre.

Para crear una subpaleta:

1. En el menú Herramientas, abra el gestor de paletas.
2. Seleccione la paleta a la que desea agregar subpaletas.
3. Pulse en el botón Subpaletas y se abrirá el cuadro de diálogo Subpaletas.
4. A la derecha de la columna *¿Mostrar?*, pulse en el botón Añadir subpaleta; se abrirá el cuadro de diálogo Crear/editar paleta secundaria.
5. Introduzca un **nombre de subpaleta** exclusivo.
6. En el área **Nodos disponibles**, seleccione el nodo que se añadirá a la subpaleta.
7. Pulse en el botón Añadir nodo con la flecha hacia la derecha para mover el nodo seleccionado al área **Nodos seleccionados**.
8. Cuando haya añadido los nodos necesarios, pulse en **Aceptar** para volver al cuadro de diálogo Subpaletas.

Las subpaletas que cree se mostrarán en la paleta de nodos cuando seleccione su pestaña de paleta padre. Consulte el tema “Modificación de la vista de la pestaña de paleta” para obtener más información.

Modificación de la vista de la pestaña de paleta

Debido al amplio número de nodos disponibles en IBM SPSS Modeler, es posible que no se puedan visualizar en pantallas más pequeñas sin tener que desplazarse a la izquierda o a la derecha de la paleta de nodos. Este hecho es especialmente patente en la pestaña de paleta de modelado. Para reducir el desplazamiento, puede seleccionar visualizar únicamente los nodos que se contienen en una subpaleta (si está disponible). Consulte el tema “Creación de una subpaleta” para obtener más información.

Para modificar los nodos que aparecen en una pestaña de paleta, seleccione la pestaña de paleta y, a continuación, en el menú de la izquierda, seleccione si se mostrarán todos los nodos o únicamente los de una subpaleta concreta.

Gestión de nodos CEMI

CEMI se ha desaprobado y se ha sustituido por CLEF, que ofrece un conjunto de características mucho más flexible y fácil de usar. Si desea obtener más información, consulte la *Guía del desarrollador de IBM SPSS Modeler 16 CLEF* incluida con este release.

Capítulo 13. Consideraciones de rendimiento para rutas y nodos

Puede diseñar sus rutas para aumentar al máximo el rendimiento organizando los nodos con la configuración más eficaz, activando el almacenamiento en caché del nodo cuando sea apropiado y prestando atención a otras consideraciones como se indica en esta sección.

Además de las consideraciones que se tratan aquí, normalmente se pueden obtener mejoras del rendimiento adicionales y más sustanciales haciendo un uso eficaz de la base de datos, especialmente a través de la optimización de SQL.

Orden de los nodos

Incluso cuando no está utilizando la optimización de SQL, el orden de los nodos en una ruta puede afectar al rendimiento. El propósito general es reducir al mínimo el procesamiento posterior; por lo tanto, cuando tenga nodos que reducen la cantidad de datos, colóquelos junto al inicio de la ruta. IBM SPSS Modeler Server puede aplicar algunas reglas de reordenación automáticamente durante la compilación para adelantar ciertos nodos cuando esté comprobado que es seguro. (Esta característica está activada de forma predeterminada. Consulte al administrador de su sistema para asegurarse de que está activada en su instalación.)

Al utilizar la optimización de SQL, desea maximizar su disponibilidad y eficacia. Como la optimización se detiene cuando la ruta contiene una operación que no se puede realizar en la base de datos, es mejor agrupar las operaciones con optimización de SQL al principio de la ruta. Esta estrategia dirige una mayor parte del proceso a la base de datos, lo que permite a IBM SPSS Modeler trabajar con menos datos.

Se pueden realizar las siguientes operaciones en la mayoría de bases de datos: Intente agruparlas al *principio* de la ruta:

- Fundir por clave (unión)
- Select
- Agregar
- Ordenar
- Ejemplo
- Append
- Operaciones distintas en el modo *incluir*, donde se seleccionan todos los campos
- Operaciones de rellenar
- Operaciones de derivación básicas que utilizan manipulación de cadena o aritmética estándar (dependiendo de qué operaciones admite la base de datos)
- Convertir a marcas

No se pueden realizar las siguientes operaciones en la mayoría de bases de datos. Se deberían colocar en la ruta *después* de las operaciones de la lista anterior:

- Operaciones en datos que no están en bases de datos, como archivos planos
- Fundir por orden
- Balance
- Operaciones de distinción en modo *descartar* o donde solamente se selecciona un subconjunto de campos como diferente
- Cualquier operación que requiera acceso a datos de registros que no sean el que se está procesando

- Derivaciones de campos de recuento y estado
- Operaciones de nodo Historial
- Operaciones que conllevan funciones "@" (series temporales)
- Modos de comprobación de tipo *Avisar* y *Abortar*
- Construcción de modelos, aplicación y análisis

Note: Los modelos generados por factores, árboles de decisión, conjuntos de reglas y regresión lineal pueden generar SQL y por ello se pueden devolver a la base de datos.

- Salida de datos a cualquier otro lugar que no sea la misma base de datos que está procesando los datos

Almacenamiento en caché de los nodos

Para optimizar la ejecución de la ruta, se puede configurar una **caché** en cualquier nodo no terminal. Cuando se configura una caché en un nodo, ésta se rellena con los datos que pasan a través del nodo la próxima vez que se ejecuta la ruta de datos. En adelante, los datos se leerán de la caché (que está almacenada en disco en un directorio temporal) en lugar del origen de datos.

El almacenamiento en caché es más útil tras una operación que exige mucho tiempo de ejecución, como la ordenación, fusión o agregación. Por ejemplo, supongamos que tiene un nodo de origen configurado para leer los datos de ventas desde una base de datos y un nodo Agregar que resume las ventas por ubicación. Se puede configurar una caché en el nodo Agregar en lugar de hacerlo en el nodo de origen, ya que se pretende que la caché almacene los datos agregados, no todo el conjunto de datos.

Note: El almacenamiento en caché en nodos de origen, que simplemente guarda una copia de los datos originales a medida que se leen en IBM SPSS Modeler, no mejorará el rendimiento en la mayoría de circunstancias.

Los nodos con el almacenamiento en caché activado se muestran con un pequeño icono de documento en la esquina superior derecha. Cuando los datos se almacenan en caché en el nodo, el icono del documento es verde.

Para activar una caché

1. En el lienzo de rutas, pulse con el botón derecho del ratón en el nodo y pulse en **Caché** en el menú.
2. En el submenú de caché, pulse en **Activar**.
3. Para desactivar la caché, pulse con el botón derecho del ratón en el nodo y pulse **Desactivar** del submenú de caché.

Almacenamiento en caché de nodos en una base de datos

Para las rutas ejecutadas en una base de datos, los datos se pueden almacenar en caché en medio de la ruta en una tabla temporal en la base de datos en lugar de en el sistema de archivos. Al combinarlo con la optimización de SQL, se puede mejorar considerablemente el rendimiento. Por ejemplo, el resultado de una ruta que fusiona varias tablas para crear una vista de minería de datos se puede guardar en caché y reutilizar cuando sea necesario. Al generar automáticamente SQL para todos los nodos posteriores en la ruta, el rendimiento se puede mejorar mucho más.

Si utiliza el almacenamiento en caché de la base de datos con cadenas con más de 255 caracteres, asegúrese de que hay un nodo Tipo anterior desde donde se lee el nodo de caché y los valores de campo, o bien, defina la longitud de la cadena mediante el parámetro `default_sql_string_length` en el archivo `options.cfg`. Al hacerlo, se asegura de que la columna correspondiente de la tabla temporal se define con la anchura correcta para acomodar las cadenas.

Para aprovechar el almacenamiento en caché en una base de datos, se debe activar el almacenamiento en caché de la base de datos y la optimización de SQL. Tenga en cuenta que la configuración de

optimización de Server reemplaza la de Client. Consulte el tema “Configuración de opciones de optimización de las rutas” en la página 44 para obtener más información.

Con el almacenamiento en caché de la base de datos activado, sólo tiene que pulsar con el botón derecho en cualquier nodo no terminal para almacenar en caché los datos en ese punto, y la caché se creará automáticamente de forma directa en la base de datos la próxima vez que se ejecute la ruta. Si no se activa el almacenamiento en caché de la base de datos o la optimización de SQL, la caché se escribirá en el sistema de archivos en lugar de en la base de datos.

Nota: las bases de datos siguientes admiten tablas temporales con el objetivo de almacenar en caché: DB2, Netezza, Oracle, SQL Server y Teradata. Otras bases de datos utilizarán una tabla normal para el almacenamiento en caché de la base de datos. El código SQL puede personalizarse para bases de datos específicas; póngase en contacto con la asistencia técnica para obtener ayuda.

Rendimiento: nodos de proceso

Ordenar. El nodo Ordenar debe leer el conjunto completo de datos de entrada antes de que se pueda ordenar. Los datos se almacenan en la memoria hasta un límite y el resto se vuelca en el disco. El algoritmo de ordenación es un algoritmo de combinación: los datos se leen en la memoria hasta el límite en cuestión y se ordenan mediante un algoritmo de ordenación rápida híbrido. Si todos los datos caben en la memoria, la ordenación se considera completada. Si no es así, se aplicará un algoritmo de ordenación de fusión. Los datos ordenados se escribirán en el archivo y el siguiente fragmento de datos se leerá en la memoria, se ordenará y se escribirá en el disco. Este proceso se repetirá hasta que se hayan leído todos los datos y, a continuación, se fusionarán los fragmentos ordenados. La fusión puede requerir repetidas pasadas sobre los datos almacenados en el disco. Durante el uso máximo, el nodo Ordenar tendrá dos copias completas del conjunto de datos en el disco: ordenada y sin ordenar.

El tiempo de ejecución global del algoritmo es del orden de $N \cdot \log(N)$, donde N es el número de registros. La ordenación en memoria es más rápida que la fusión desde disco, por lo que el tiempo de ejecución real se puede reducir asignando más memoria a la ordenación. El algoritmo se asigna a sí mismo una fracción de RAM física controlada por la opción de configuración de IBM SPSS Modeler Server *Multiplicador de uso de memoria*. Para incrementar la memoria utilizada para la ordenación, proporcione más memoria RAM física o aumente este valor. Tenga en cuenta que cuando la proporción de memoria utilizada excede el conjunto del proceso, de manera que parte de la memoria se pagina en el disco, el rendimiento empeora porque el modelo de acceso de memoria del algoritmo de ordenación en memoria es aleatorio y puede causar una paginación excesiva. Otros nodos aparte del nodo Ordenar también utilizan el algoritmo de ordenación, pero se aplican las mismas reglas de rendimiento.

Intervalos. El nodo Intervalos lee el conjunto de datos completo para calcular los límites de los intervalos antes de asignar los registros a los intervalos. El conjunto de datos se almacena en caché mientras se calculan los límites y se vuelve a explorar para la asignación. Cuando el método de intervalos es de *anchura fija* o *desviación estándar+media*, el conjunto de datos se almacena en caché directamente en el disco. Estos métodos tienen un tiempo de ejecución lineal y requieren suficiente espacio en disco para almacenar el conjunto completo de datos. Cuando el método de intervalos es *rangos* o *cuantiles*, el conjunto de datos se ordena utilizando el algoritmo de ordenación descrito anteriormente y el conjunto de datos ordenados se utiliza como caché. La ordenación proporciona a estos métodos un tiempo de ejecución de $M \cdot N \cdot \log(N)$, donde M es el número de campos en intervalos y N es el número de registros; requiere un espacio en disco equivalente al doble del tamaño del conjunto de datos.

La generación de un nodo Derivar basado en intervalos generados mejorará el rendimiento en las siguientes pasadas. Las operaciones de derivar son más rápidas que los intervalos.

Fundir por clave (unión). El nodo Fundir, cuando el método de fusión es *claves* (equivalente a una unión de base de datos), ordena cada uno de los conjuntos de datos de entrada por los campos clave. Esta parte del procedimiento tiene un tiempo de ejecución de $M \cdot N \cdot \log(N)$, donde M es el número de entradas y N es el número de registros de la entrada más grande; requiere suficiente espacio en disco para almacenar

todos los conjuntos de datos de entrada además de una segunda copia del conjunto de datos más grande. El tiempo de ejecución de la fusión en sí es proporcional al tamaño del conjunto de datos de salida, que depende de la frecuencia de las claves de coincidencia. En el peor de los casos, donde la salida es el producto cartesiano de las entradas, el tiempo de ejecución puede aproximarse a N^2 . Es poco frecuente; la mayoría de uniones tienen muchas menos claves de coincidencia. Si un conjunto de datos es relativamente mayor que los otros o si los datos entrantes ya se han ordenado por un campo clave, puede mejorar el rendimiento de este nodo utilizando la pestaña Optimización.

Agregar. Cuando no se ha activado la opción *Las claves son contiguas*, este nodo lee (pero no almacena) el conjunto de datos de entrada completo antes de generar una salida agregada. En las situaciones más extremas, en las que el tamaño de los datos agregados alcanza un límite (determinado por la opción de configuración de IBM SPSS Modeler Server *Multiplicador de uso de memoria*), el resto del conjunto de datos se ordena y se procesa como si se hubiera activado la opción *Las claves son contiguas*. Cuando se activa esta opción, no se almacena ningún dato porque los registros de salida agregados se generan a medida que se leen los datos de entrada.

Distinguir. El nodo Distinguir guarda todos los campos de clave exclusiva en el conjunto de entrada del conjunto de datos; si todos los campos son campos clave y todos los registros son exclusivos, guarda el conjunto de datos completo. De forma predeterminada, el nodo Distinguir clasifica los datos de los campos clave y selecciona (o descarta) el primer registro distinto de cada grupo. En conjuntos de datos más pequeños con un pequeño número de claves distintas, o los que se hayan ordenado previamente, puede seleccionar las opciones para mejorar la velocidad y eficacia de procesamiento.

Tipo. En algunos casos, el nodo Tipo almacena en caché los datos de entrada al leer los valores; la caché se utiliza para el procesamiento posterior de la ruta. La caché requiere espacio en disco suficiente para almacenar el conjunto de datos completo pero agiliza el procesamiento.

Evaluación. El nodo Evaluación debe ordenar los datos de entrada para calcular los cuantiles. La ordenación se repite para cada modelo evaluado porque las puntuaciones y el consiguiente orden de los registros son diferentes en cada caso. El tiempo de ejecución es $M*N*\log(N)$, donde M es el número de modelos y N es el número de registros.

Rendimiento: nodos de modelado

Nodo Red neuronal y Kohonen. Los algoritmos de entrenamiento de redes neuronales (incluyendo el algoritmo Kohonen) realizan muchas pasadas en los datos de entrenamiento. Los datos se almacenan en la memoria hasta un límite y el resto se vuelca en el disco. El acceso a los datos de entrenamiento desde el disco resulta caro porque el método de acceso es aleatorio, lo que puede provocar una excesiva actividad de disco. Puede desactivar el uso del almacenamiento en disco para estos algoritmos y obligar a que todos los datos se almacenen en memoria seleccionando la opción **Optimizar velocidad** en la pestaña Modelo del cuadro de diálogo del nodo. Observe que si la cantidad de memoria requerida para almacenar los datos es mayor que el conjunto del proceso de servidor, parte se paginará en el disco y el rendimiento sufrirá en consecuencia.

Cuando se active **Optimizar memoria**, un porcentaje de la RAM física se asigna al algoritmo según el valor de la opción de configuración de IBM SPSS Modeler Server de *Porcentaje límite de memoria de modelado*. Para utilizar más memoria para las redes neuronales de entrenamiento, proporcione más RAM o incremente el valor de esta opción, pero tenga en cuenta que un valor demasiado alto provocará paginación.

El tiempo de ejecución de los algoritmos de redes neuronales dependerá del nivel de precisión deseado. Puede controlar el tiempo de ejecución definiendo una condición de parada en el cuadro de diálogo del nodo.

K-medias. El algoritmo de clústeres de K-medias tiene las mismas opciones para controlar el uso de memoria que los algoritmos de redes neuronales. El rendimiento de datos almacenados en el disco es mejor, sin embargo, porque el acceso a los datos es secuencial.

Rendimiento: expresiones CLEM

Las funciones de secuencia CLEM (“funciones @”) que buscan hacia atrás en la ruta de datos deben almacenar suficientes datos como para satisfacer la búsqueda que más deba retroceder. Para las operaciones cuyo grado de búsqueda hacia atrás no tiene límites, se deben almacenar todos los valores del campo. Una operación sin límites es una operación donde el valor de desplazamiento no es un entero literal; por ejemplo @OFFSET(Ventas, Mes). El valor de desplazamiento es el nombre del campo *Mes*, cuyo valor no se conoce hasta que se ejecute. El servidor deberá guardar todos los valores del campo *Ventas* para garantizar la precisión de los resultados. Cuando se conoce un límite superior, deberá proporcionarlo como un argumento adicional; por ejemplo @OFFSET(Ventas, Mes, 12). Esta operación indica al servidor que no almacene más de los 12 valores más recientes de *Ventas*. Las funciones de secuencia, con límites o de otro tipo, casi siempre inhiben la generación de SQL.

Capítulo 14. Accesibilidad en IBM SPSS Modeler

Conceptos básicos sobre la accesibilidad de IBM SPSS Modeler

IBM SPSS Modeler proporciona soporte de accesibilidad a todos los usuarios, así como compatibilidad específica para usuarios con problemas de visión o funcionales. En esta sección se describen las características y los métodos de trabajo usando las mejoras de accesibilidad, como los lectores de pantallas y los atajos de teclado.

Tipos de soporte de accesibilidad

Si tiene algún tipo de deficiencia visual o depende del teclado para la manipulación, existe una amplia variedad de métodos alternativos que permiten utilizar este kit de herramientas de minería de datos. Por ejemplo, puede generar rutas, especificar opciones y leer resultados sin tener que utilizar el ratón. Los atajos de teclado disponibles se enumeran en los temas siguientes. Además, IBM SPSS Modeler ofrece un amplio soporte para lectores de pantallas como JAWS para Windows. También puede optimizar la combinación de colores para obtener más contraste. Estos tipos de soporte se describen en los siguientes temas.

Accesibilidad para personas con problemas de visión

Existen varias propiedades que se pueden especificar en IBM SPSS Modeler para mejorar la capacidad de utilización del software.

Opciones de representación

Puede seleccionar colores para la representación de los gráficos. También permite aplicar la configuración específica de Windows en el propio software. Esto puede ayudar a aumentar el contraste visual.

1. Para establecer opciones de representación, en el menú Herramientas, pulse en **Opciones de usuario**.
2. Pulse en la pestaña **Representación**. Las opciones de esta pestaña incluyen la combinación de colores del software, los colores de los gráficos y los tamaños de las fuentes de los nodos.

Utilización de sonidos para las notificaciones

Mediante la activación y desactivación de los sonidos puede controlar la forma en que recibirá los avisos de determinadas operaciones del software. Por ejemplo, puede activar el sonido en eventos como la creación y eliminación de nodos o la generación de nuevos resultados o modelos.

1. Para establecer opciones de notificación, en el menú Herramientas, pulse en **Opciones de usuario**.
2. Pulse en la pestaña **Notificaciones**.

Control de la apertura automática de nuevas ventanas

La pestaña Notificaciones del cuadro de diálogo Opciones de usuario también se utiliza para determinar si los nuevos resultados generados, como tablas o gráficos, se deben abrir en una ventana independiente. Puede que le resulte más fácil desactivar esta opción y abrir una ventana de resultados sólo cuando sea necesario.

1. Para establecer estas opciones, en el menú Herramientas, pulse en **Opciones de usuario**.
2. Pulse en la pestaña **Notificaciones**.
3. En el cuadro de diálogo, seleccione la opción **Nuevo resultado** de la lista del grupo **Notificaciones visuales**.
4. En **Abrir Ventana**, seleccione **Nunca**.

Tamaño del nodo

Los nodos pueden representarse mediante un tamaño estándar o pequeño. Si lo desea, puede ajustar estos tamaños para que se ajusten a sus necesidades.

1. Para establecer las opciones de tamaño del nodo, en el menú Archivo pulse en **Propiedades de ruta**.
2. Pulse en la pestaña **Diseño**.
3. En la lista **Tamaño de icono**, seleccione **Estándar**.

Accesibilidad para usuarios invidentes

El soporte para usuarios invidentes depende principalmente de la utilización de un lector de pantallas, como JAWS para Windows. Para optimizar el uso de un lector de pantallas con IBM SPSS Modeler, puede especificar varias configuraciones.

Opciones de representación

Los lectores de pantallas tienden a funcionar mejor cuando el contraste visual es mayor en la pantalla. Si ya cuenta con una configuración de Windows de alto contraste, puede elegir utilizar esta configuración de Windows para el propio software.

1. Para establecer opciones de representación, en el menú Herramientas, pulse en **Opciones de usuario**.
2. Pulse en la pestaña **Representación**.

Utilización de sonidos para las notificaciones

Mediante la activación y desactivación de los sonidos puede controlar la forma en que recibirá los avisos de determinadas operaciones del software. Por ejemplo, puede activar el sonido en eventos como la creación y eliminación de nodos o la generación de nuevos resultados o modelos.

1. Para establecer opciones de notificación, en el menú Herramientas, pulse en **Opciones de usuario**.
2. Pulse en la pestaña **Notificaciones**.

Control de la apertura automática de nuevas ventanas

La pestaña Notificaciones del cuadro de diálogo Opciones de usuario también se utiliza para determinar si los nuevos resultados generados se deben abrir en una ventana independiente. Puede que le resulte más fácil desactivar esta opción y abrir una ventana de resultados cuando lo necesite.

1. Para establecer estas opciones, en el menú Herramientas, pulse en **Opciones de usuario**.
2. Pulse en la pestaña **Notificaciones**.
3. En el cuadro de diálogo, seleccione la opción **Nuevo resultado** de la lista del grupo **Notificaciones visuales**.
4. En **Abrir Ventana**, seleccione **Nunca**.

Accesibilidad desde el teclado

Se puede acceder a la funcionalidad del producto desde el teclado. En el nivel más básico, puede pulsar Alt y la tecla adecuada para activar menús de ventana (por ejemplo Alt+A para acceder al menú Archivo) o pulsar el tabulador para desplazarse por los controles de los cuadros de diálogo. No obstante, existen ciertas cuestiones especiales relacionadas con cada una de las ventanas principales del producto y algunas sugerencias útiles para navegar por los cuadros de diálogo.

La presente sección tratará de los puntos principales de la accesibilidad mediante el teclado, desde la apertura de una ruta hasta el uso de cuadros de diálogo de nodos para trabajar con los resultados. Asimismo, se enumerarán los atajos de teclado que permiten moverse por el programa de un modo aún más eficaz.

Atajos para desplazarse por la ventana principal

La mayoría del trabajo de minería de datos se realiza en la ventana principal de IBM SPSS Modeler. El área principal se denomina **lienzo de rutas** y se utiliza para generar y ejecutar rutas de datos. La parte inferior de la ventana contiene **paletas de nodos**, que incluyen todos los nodos disponibles. Las paletas se organizan en pestañas que corresponden al tipo de operación de minería de datos de cada grupo de nodos. Por ejemplo, los nodos utilizados para introducir datos en IBM SPSS Modeler se agrupan en la pestaña Orígenes, mientras que los nodos utilizados para derivar, filtrar o clasificar campos se agrupan en la pestaña Operaciones con campos.

La parte derecha de la ventana contiene varias herramientas para administrar rutas, resultados y proyectos. La mitad superior derecha contiene los **gestores** e incluye tres pestañas que se utilizan para administrar rutas, resultados y modelos generados. Se puede acceder a estos objetos seleccionando la pestaña y un objeto de la lista. La mitad inferior derecha contiene el **panel de proyectos**, que le permite organizar el trabajo en proyectos. En esta área hay dos pestañas que reflejan dos vistas diferentes de un proyecto. La **vista Clases** ordena los objetos del proyecto por tipo, mientras que la **vista CRISP-DM** ordena los objetos según la fase de minería de datos relevante, como Preparación de los datos o Modelado. Estos distintos aspectos de la ventana de IBM SPSS Modeler se describen en el sistema de ayuda y el Manual del usuario.

A continuación se detalla una tabla de atajos utilizados para desplazarse por la ventana principal de IBM SPSS Modeler y generar rutas. Los atajos de los cuadros de diálogo y los resultados se enumeran en los temas siguientes. Recuerde que estos atajos sólo están disponibles en la ventana principal.

Tabla 32. Atajos de la ventana principal

Tecla de acceso directo	Función
Ctrl+F5	Se desplaza a las paletas de nodo.
Ctrl+F6	Se desplaza al lienzo de rutas.
Ctrl+F7	Se desplaza al panel de gestores.
Ctrl+F8	Se desplaza al panel de proyectos.

Tabla 33. Atajos de nodo y ruta

Tecla de acceso directo	Función
Ctrl+N	Crea un nuevo lienzo de rutas vacío.
Ctrl+O	Muestra el cuadro de diálogo Abrir, desde el que puede seleccionar y abrir una ruta existente.
Ctrl+teclas numéricas	Se desplaza a la pestaña correspondiente en una ventana o panel. Por ejemplo, en un panel o una ventana con pestañas, Ctrl+1 se desplaza a la primera pestaña comenzando por la izquierda; Ctrl+2, a la segunda, etc.
Ctrl+Flecha hacia abajo	Se utiliza en la paleta de nodos para desplazarse de una pestaña de paleta al primer nodo bajo esa pestaña.
Ctrl+Flecha hacia arriba	Se utiliza en la paleta de nodos para desplazarse de un nodo a su pestaña de paleta.
Intro	Cuando se selecciona un nodo en la paleta de nodos (incluyendo modelos refinados de la paleta de modelos generados), esta tecla añade el nodo al lienzo de rutas. Si pulsa en Intro cuando un nodo ya está seleccionado en el lienzo, abrirá el cuadro de diálogo de dicho nodo.
Ctrl+Intro	Cuando se selecciona un nodo en la paleta, añade el nodo al lienzo de rutas sin seleccionarlo, a la vez que se desplaza al primer nodo de la paleta.
Alt+Intro	Cuando se selecciona un nodo en la paleta, añade el nodo al lienzo de rutas y lo selecciona, a la vez que se desplaza al primer nodo de la paleta.

Tabla 33. Atajos de nodo y ruta (continuación)

Tecla de acceso directo	Función
Mayús+Barra espaciadora	Cuando un nodo o comentario está centrado en la paleta, cambia entre seleccionar y eliminar la selección del nodo o comentario. Si se seleccionan también otros nodos o comentarios, esto hace que se anule la selección.
Ctrl+Mayús+Barra espaciadora	Cuando un nodo o comentario está centrado en la ruta o un nodo o comentario está centrado en la paleta, cambia entre seleccionar y eliminar la selección del nodo o comentario. Esto no afecta a ningún otro nodo o comentario seleccionado.
Flecha izquierda/derecha	Si el lienzo de rutas está activo, desplaza toda la ruta horizontalmente en la pantalla. Si una pestaña de paleta está activa, muestra pestañas sucesivamente. Si un nodo de paleta está activo, se desplaza entre nodos en la paleta.
Flecha hacia arriba/hacia abajo	Si el lienzo de rutas está activo, desplaza toda la ruta verticalmente en la pantalla. Si un nodo de paleta está activo, se desplaza entre nodos en la paleta. Si una subpaleta está activa, se desplaza entre otras subpaletas de esta pestaña de paleta.
Alt+Flecha izquierda/derecha	Mueve los nodos y comentarios seleccionados en el lienzo de rutas horizontalmente en la dirección de la flecha.
Alt+Flecha hacia arriba/hacia abajo	Mueve los nodos y comentarios seleccionados en el lienzo de rutas verticalmente en la dirección de la flecha.
Ctrl+A	Selecciona todos los nodos de una ruta.
Ctrl+Q	Cuando un nodo está activo, selecciona ese nodo y todos los que se encuentran a continuación en la ruta y elimina la selección de todos los nodos anteriores en la ruta.
Ctrl+W	Cuando un nodo seleccionado está activo, elimina su selección y la de todos los nodos seleccionados que se encuentran a continuación en la ruta.
Ctrl+Alt+D	Duplica un nodo seleccionado.
Ctrl+Alt+L	Cuando se selecciona un nugget de modelo en la ruta, abre el cuadro de diálogo Insertar para permitirle cargar un modelo guardado desde un archivo .nod en la ruta.
Ctrl+Alt+R	Muestra la pestaña Anotaciones para un nodo seleccionado, lo que le permite cambiar el nombre del nodo.
Ctrl+Alt+U	Crea un nodo de origen Datos Usuario.
Ctrl+Alt+C	Activa/desactiva la caché para un nodo.
Ctrl+Alt+F	Vacía la caché de un nodo.
Pestaña	En el lienzo de rutas, muestra sucesivamente todos los nodos y comentarios de origen en la ruta actual. En una paleta de nodos, se desplaza entre nodos en la paleta. En una subpaleta seleccionada, se desplaza al primer nodo de la subpaleta.
Mayús+Tabulador	Realiza la misma operación que el tabulador, pero en orden inverso.
Ctrl+Tabulador	Con el panel de gestores o el de proyectos seleccionado, pasa a activar el lienzo de rutas. Con una paleta de nodos seleccionada, pasa a activar un nodo y su pestaña de paleta.
Cualquier tecla alfabética	Con un nodo de la ruta actual seleccionado, activa y muestra el siguiente nodo cuyo nombre comience por la tecla pulsada.
F1	Abre el sistema de ayuda en un tema relevante para el elemento activado.

Tabla 33. Atajos de nodo y ruta (continuación)

Tecla de acceso directo	Función
F2	Inicia el proceso de conexión para un nodo seleccionado en el lienzo. Utilice el tabulador para desplazarse al nodo que desee en el lienzo y pulse Mayús+barra espaciadora para finalizar la conexión.
F3	Elimina todas las conexiones del nodo seleccionado del lienzo.
F6	Pasa a activar el panel de gestores, el de proyectos y las paletas de nodo.
F10	Abre el menú Archivo.
Mayús+F10	Abre el menú emergente del nodo o ruta.
Suprimir	Elimina un nodo seleccionado del lienzo.
Esc	Cierra un menú emergente o cuadro de diálogo.
Ctrl+Alt+X	Expande un supernodo.
Ctrl+Alt+Z	Se acerca a un supernodo.
Ctrl+Alt+Mayús+Z	Se aleja del supernodo
Ctrl+E	Con el lienzo de rutas activo, ejecuta la ruta actual.

En IBM SPSS Modeler también se utilizan varias teclas de acceso directo estándar, como Ctrl+C para copiar. Consulte el tema “Uso de teclas de acceso directo” en la página 22 para obtener más información.

Atajos para cuadros de diálogo y tablas

Hay varias teclas de acceso directo y de lector de pantallas que son útiles cuando se trabaja con cuadros de diálogo, tablas y tablas en cuadros de diálogo. A continuación se detalla una lista completa de los atajos especiales del teclado y el lector de pantallas.

Tabla 34. Atajos de cuadro de diálogo y generador de expresiones

Tecla de acceso directo	Función
Alt+4	Se utiliza para anular todos los cuadros de diálogo o ventanas de resultados abiertos. Los resultados se pueden recuperar de la pestaña Resultados del panel de gestores.
Ctrl+Fin	Con cualquier control del generador de expresiones activo, moverá el punto de inserción al final de la expresión.
Ctrl+1	En el generador de expresiones, activa el control de edición de expresiones.
Ctrl+2	En el generador de expresiones, activa la lista de funciones.
Ctrl+3	En el generador de expresiones, activa la lista de campos.

Atajos de tabla

Los atajos de tabla se utilizan para las tablas de resultados y los controles de tabla de los cuadros de diálogo de nodos como Tipo, Filtro o Fundir. Normalmente, se utilizará el Tabulador para desplazarse por las casillas de la tabla y Ctrl+Tabulador para salir del control de tabla. *Note:* puede suceder que un lector de pantallas no comience a leer de forma inmediata el contenido de una casilla. Al pulsar en las teclas de flecha una o dos veces se restablecerá el software y se iniciará el discurso.

Tabla 35. Atajos de tabla

Tecla de acceso directo	Función
Ctrl+W	En las tablas, lee la descripción corta de la fila seleccionada (roW, en inglés). Por ejemplo, "Los 2 valores de la fila seleccionada son sexo, marca, m/f, etc."

Tabla 35. Atajos de tabla (continuación)

Tecla de acceso directo	Función
Ctrl+Alt+W	En las tablas, lee la descripción larga de la fila seleccionada (roW , en inglés). Por ejemplo, "Los 2 valores de la fila seleccionada son campo = sexo, tipo = marca, sexo = m/f, etc."
Ctrl+D	En las tablas, lee la Descripción corta del área seleccionada. Por ejemplo, "La selección es de una fila y seis columnas."
Ctrl+Alt+D	En las tablas, ofrece la Descripción larga del área seleccionada. Por ejemplo, "La selección es de una fila y seis columnas. Las columnas seleccionadas son Campo, Tipo y Perdido. La fila seleccionada es 1."
Ctrl+T	En las tablas, ofrece una descripción corta de las columnas seleccionadas. Por ejemplo, "Campos, Tipo y Perdido."
Ctrl+Alt+T	En las tablas, ofrece una descripción larga de las columnas seleccionadas. Por ejemplo, "Las columnas seleccionadas son Campos, Tipo y Perdido."
Ctrl+R	En las tablas, indica el número de Registros existentes.
Ctrl+Alt+R	En las tablas, indica el número de Registros existentes y los nombres de las columnas.
Ctrl+I	En las tablas, lee la Información o el contenido de la casilla de tabla activa.
Ctrl+Alt+I	En las tablas, lee la descripción larga de la Información (nombre de columna y contenido) de la casilla de tabla activa.
Ctrl+G	En las tablas, ofrece información corta de selección General.
Ctrl+Alt+G	En las tablas, ofrece información larga de selección General.
Ctrl+Q	En las tablas, ofrece un desplazamiento rápido (Quick , en inglés) por las casillas de la tabla. Ctrl+Q lee descripciones largas, como "Sexo=Mujer", mientras se desplaza por la tabla mediante las teclas de flecha. Al pulsar Ctrl+Q de nuevo, se mostrarán las descripciones cortas (contenido de las casillas).

Atajos para comentarios

Si trabaja con comentarios en pantalla, puede utilizar los siguientes atajos.

Tabla 36. Accesos directos de comentario

Tecla de acceso directo	Función
Alt+C	Activa y desactiva la característica de mostrar/ocultar comentarios.
Alt+M	Inserta un nuevo comentario si los comentarios se visualizan; muestra comentarios si están ocultos actualmente.
Pestaña	En el lienzo de rutas, muestra sucesivamente todos los nodos y comentarios de origen en la ruta actual.
Intro	Si un comentario está resaltado, indica el inicio de la edición.
Alt+Intro o Ctrl+Tabulador	Finaliza la edición y guarda los cambios de edición.
Esc	Cancela la edición. Se perderán los cambios realizados durante la edición.
Alt+Mayús+Flecha hacia arriba	Reduce la altura del área de texto en una casilla de cuadrícula (o un píxel) si el ajuste a la cuadrícula está activado (o desactivado).
Alt+Mayús+Flecha hacia abajo	Aumenta la altura del área de texto en una casilla de cuadrícula (o un píxel) si el ajuste a la cuadrícula está activado (o desactivado).

Tabla 36. Accesos directos de comentario (continuación)

Tecla de acceso directo	Función
Alt+Mayús+Flecha izquierda	Reduce la anchura del área de texto en una casilla de cuadrícula (o un píxel) si el ajuste a la cuadrícula está activado (o desactivado).
Alt+Mayús+Flecha derecha	Aumenta la anchura del área de texto en una casilla de cuadrícula (o un píxel) si el ajuste a la cuadrícula está activado (o desactivado).

Atajos para Cluster Viewer y Model Viewer

Las teclas de acceso directo están disponible para navegar en las ventanas de Cluster Viewer y Model Viewer.

Tabla 37. Accesos directos generales - Cluster Viewer y Model Viewer

Tecla de acceso directo	Función
Pestaña	Se desplaza al siguiente control de pantalla.
Mayús+Tabulador	Se desplaza al control de pantalla anterior.
Flecha hacia abajo	Si una lista desplegable está activa, abre la lista o se desplaza al siguiente elemento de la lista. Si un menú está activo, se desplaza al siguiente elemento del menú. Si un gráfico en miniatura está seleccionado, cambia al siguiente del conjunto (o al primero si es el último el que está seleccionado).
Flecha hacia arriba	Si hay una lista desplegable abierta, se desplaza al elemento anterior de la lista. Si un menú está activo, se desplaza al elemento anterior del menú. Si un gráfico en miniatura está seleccionado, cambia al anterior del conjunto (o al último si es el primero el que está seleccionado).
Intro	Cierra una lista desplegable abierta o hace una selección en un menú abierto.
F6	Cambia la activación entre los paneles izquierdo y derecho de la ventana.
Flechas izquierda y derecha	Si una pestaña está activa, se desplaza a la pestaña siguiente o anterior. Si un menú está activo, se desplaza al menú siguiente o anterior.
Alt+letra	Selecciona el botón o el menú que tiene esta letra subrayada en su nombre.
Esc	Cierra un menú o lista desplegable abierto.

Sólo para Cluster Viewer

Cluster Viewer tiene una visión de clústeres que contiene una cuadrícula de clústeres por características.

Para seleccionar la vista de clústeres en vez de la vista de resumen de modelos:

1. Pulse la tecla Tabulador repetidamente hasta que se seleccione el botón **Ver**.
2. Pulse la flecha hacia abajo dos veces para seleccionar **Clústeres**.

Desde aquí puede seleccionar una celda individual dentro de la cuadrícula:

3. Pulse Tabulador repetidamente hasta que llegue al último icono de la barra de herramientas de visualización.



Figura 17. Mostrar el icono de árbol de visualización

4. Pulse Tabulador una vez más, la barra espaciadora y, a continuación, la tecla de flecha.

Los siguientes atajos de teclado están ahora disponibles:

Tabla 38. Accesos directos de Cluster Viewer

Tecla de acceso directo	Función
Tecla de flecha	Se desplaza por las celdas individuales de la cuadrícula. La visualización de la distribución de celdas del panel derecho cambia a medida que se desplaza la activación.
Ctrl+, (coma)	Selecciona o anula la selección de la columna completa de la cuadrícula en la que una celda tiene la activación. Para añadir una columna a la selección, utilice las teclas de flecha para navegar hasta una celda en esa columna y pulse Ctrl+, de nuevo.
Pestaña	Desplaza el control de la cuadrícula al siguiente control de pantalla.
Mayús+Tabulador	Desplaza el control de la cuadrícula y vuelve al control de pantalla anterior.
F2	Abre el modo de edición (sólo para etiquetas y celdas de descripción).
Intro	Guarda los cambios de edición y sale del modo de edición ((sólo para etiquetas y celdas de descripción).
Esc	Sale del modo de edición sin guardar los cambios (sólo para etiquetas y celdas de descripción).

Ejemplo de teclas de acceso directo: crear rutas

Para que el proceso de generación de rutas sea más claro para los usuarios que dependen del teclado o de un lector de pantallas, a continuación se ofrece un ejemplo de creación de rutas sin utilizar el ratón. En este ejemplo, creará una ruta que contiene un nodo Archivo var., un nodo Derivar y un nodo Histograma siguiendo estos pasos:

1. **Inicie IBM SPSS Modeler.** Cuando se inicia IBM SPSS Modeler por primera vez, activa la pestaña Favoritos de la paleta de nodos.
2. **Ctrl+Flecha hacia abajo.** Se desplaza del contenido de la pestaña al cuerpo de ésta.
3. **Flecha derecha.** Se desplaza al nodo Archivo var.
4. **Barra espaciadora.** Selecciona el nodo Archivo var.
5. **Ctrl+Intro.** Añade el nodo Archivo de variable al lienzo de rutas. Esta combinación de teclas también mantiene la selección en el nodo Archivo var. de forma que el siguiente nodo que se añada se conecta a éste.
6. **Tabulador.** Devuelve la activación a la paleta de nodos.
7. **Flecha derecha 4 veces.** Se desplaza al nodo Derivar.
8. **Barra espaciadora.** Selecciona el nodo Derivar.
9. **Alt+Intro.** Añade el nodo Derivar al lienzo y desplaza la selección al nodo Derivar. Este nodo ya está listo para conectarse al siguiente nodo que se añada.
10. **Tabulador.** Devuelve la activación a la paleta de nodos.
11. **Flecha derecha 5 veces.** Se desplaza al nodo Histograma en la paleta.

12. **Barra espaciadora.** Selecciona el nodo Histograma.
13. **Intro.** Añade el nodo al lienzo y activa el lienzo de rutas.

Pasa al siguiente ejemplo, o guarda la ruta si desea probar el siguiente ejemplo más adelante.

Ejemplo de teclas de acceso directo: editar nodos

En este ejemplo, utilizará la ruta creada en el ejemplo anterior. La ruta consiste en un nodo Archivo var., un nodo Derivar y un nodo Histograma. Las instrucciones comienzan con el tercer nodo de la ruta activado, el nodo Histograma.

1. **Ctrl+Flecha izquierda 2 veces.** Vuelve a activar el nodo Archivo variable.
2. **Intro.** Abre el cuadro de diálogo Archivo var. Pasa por el campo Archivo e introduce una ruta de archivo de texto y nombre para seleccionar ese archivo. Pulse Ctrl+Tabulador para navegar por la parte inferior del cuadro de diálogo, y pase con el tabulador hasta el botón Aceptar y pulse Intro para cerrar el cuadro de diálogo.
3. **Ctrl+Flecha derecha.** Se activa el segundo nodo: Derivar.
4. **Intro.** Abre el cuadro de diálogo del nodo Derivar. Pulse el tabulador para seleccionar los campos y especificar las condiciones de derivación. Pulse Ctrl+Tabulador para navegar hasta el botón Aceptar y pulse Intro para cerrar el cuadro de diálogo.
5. **Ctrl+Flecha derecha.** Se activa el tercer nodo: Histograma.
6. **Intro.** Abre el cuadro de diálogo del nodo Histograma. Pulse el tabulador para seleccionar los campos y especificar las opciones de gráfico. Para las listas desplegadas, pulse la flecha hacia abajo para abrir la lista y resaltar un elemento de lista, a continuación, pulse Intro para seleccionar el elemento de lista. Pulse el tabulador para llegar al botón Aceptar y pulse Intro para cerrar el cuadro de diálogo.

En este punto, puede añadir nodos adicionales o ejecutar la ruta actual. Recuerde las siguientes sugerencias cuando esté generando rutas:

- Cuando se conectan nodos manualmente, utilice F2 para crear el punto inicial de una conexión, pulse el tabulador para pasar al punto final y utilice a continuación Mayús+Barra espaciadora para finalizar la conexión.
- Utilice F3 para destruir todas las conexiones de un nodo seleccionado en el lienzo.
- Una vez que haya creado una ruta, utilice Ctrl+E para ejecutar la ruta actual.

Existe una lista completa de teclas de acceso directo a su disposición. Consulte el tema “Atajos para desplazarse por la ventana principal” en la página 169 para obtener más información.

Utilización de un lector de pantallas

El mercado ofrece diversos modelos de lectores de pantallas. IBM SPSS Modeler se ha configurado para ser compatible con JAWS para Windows a través de Java Access Bridge, que se instala junto con IBM SPSS Modeler. Si ya tiene JAWS instalado, basta con iniciar JAWS antes de iniciar IBM SPSS Modeler para utilizar este producto.

Debido a la naturaleza de la exclusiva representación gráfica del proceso de minería de datos de IBM SPSS Modeler, los gráficos y los diagramas se utilizan visualmente de forma óptima. No obstante, es posible comprender y tomar decisiones en función de resultados y modelos visualizados textualmente mediante un lector de pantallas.

Note: Con máquinas cliente de 64 bits, algunas características de tecnología asistivas no funcionan. Esto se debe a que el puente de acceso Java no está diseñado para el funcionamiento de 64 bits.

Utilización del archivo de diccionario de IBM SPSS Modeler

Existe un archivo de diccionario de IBM SPSS Modeler (*Awt.JDF*) disponible que incluir con JAWS. Para utilizar este archivo:

1. Desplácese al subdirectorio */accessibility* de la carpeta de instalación de IBM SPSS Modeler y copie el archivo de diccionario (*Awt.JDF*).
2. A continuación, péguelo en el directorio que contiene los scripts de JAWS.

Puede que ya exista un archivo denominado *Awt.JDF* en su equipo si está ejecutando otras aplicaciones de JAVA. En ese caso, tal vez no pueda utilizar este archivo de diccionario si no lo modifica manualmente.

Utilización de un lector de pantallas con resultados en formato HTML

Al visualizar los resultados en formato HTML en IBM SPSS Modeler mediante un lector de pantallas, puede que surjan algunas dificultades. Se pueden ver afectados varios tipos de resultados, como:

- Resultados visualizados en la pestaña Avanzado para los nodos Regresión, Logística y PCA/Factorial
- Resultados del nodo Informe

En cada una de estas ventanas o cuadros de diálogo, hay una opción de la barra de herramientas que se puede utilizar para abrir los resultados en el explorador predeterminado, que ofrece soporte de lector de pantallas estándar. Puede utilizar entonces el lector de pantallas para visualizar la información de resultados.

Accesibilidad en la ventana Árbol interactivo

La visualización estándar de un modelo de árbol de decisión en la ventana Árbol interactivo puede causar problemas a las personas que lean los resultados en pantalla. Para acceder a una versión más inteligible, en los menús de Árbol interactivo pulse:

Ver > Ventana Accesible

Muestra una vista similar al mapa de árbol estándar, pero que JAWS puede leer correctamente. Puede desplazarse hacia arriba, abajo, derecha o izquierda usando las teclas de flecha. A medida que explora la ventana de acceso, el foco de la ventana Árbol interactivo se desplaza consecuentemente. Utilice la barra espaciadora para cambiar la selección o pulse Ctrl+barra espaciadora para ampliar la selección actual.

Sugerencias de utilización

Existen algunas sugerencias que hacen que el entorno de IBM SPSS Modeler resulte más accesible. A continuación se detallan sugerencias generales para trabajar en IBM SPSS Modeler.

- **Cierre de los cuadros de texto extendidos.** Utilice Ctrl+Tabulador para salir de los cuadros de texto extendidos. *Note:* Ctrl+Tabulador también se utiliza para salir de los controles de tabla.
- **Utilización del Tabulador en lugar de las teclas de flecha.** Al seleccionar las opciones de un cuadro de diálogo, utilice el Tabulador para desplazarse por los botones de opción. En este contexto no funcionarán las teclas de flecha.
- **Listas desplegadas.** En una lista desplegable de un cuadro de diálogo puede utilizar tanto la tecla Escape como la barra espaciadora para seleccionar un elemento y cerrar la lista. También puede utilizar la tecla Escape para cerrar las listas desplegadas que no se cierran cuando se desplaza a otro control con el Tabulador.
- **Estado de ejecución.** Al ejecutar una ruta en una base de datos grande, JAWS puede retardar la lectura del estado de la ruta. Pulse la tecla Ctrl de forma periódica para actualizar el informe de estado.
- **Utilización de las paletas de nodos.** Al entrar por primera vez en una pestaña de las paletas de nodos, JAWS leerá a veces "groupbox" en lugar del nombre del nodo. En ese caso, puede utilizar Ctrl+Flecha derecha y, a continuación, Ctrl+Flecha izquierda para restablecer el lector de pantallas y escuchar el nombre del nodo.
- **Lectura de menús.** En ocasiones, cuando se abre un menú por primera vez, puede que JAWS no lea el primer elemento del menú. Si sospecha que puede haber ocurrido esto, utilice la flecha hacia abajo y, a continuación, la flecha hacia arriba para escuchar el primer elemento del menú.

- **Menús en cascada.** JAWS no lee el primer nivel de un menú en cascada. Si nota una pausa en el discurso al desplazarse por el menú, pulse la tecla de flecha derecha para escuchar los elementos hijo del menú.

Además, si ha instalado IBM SPSS Modeler Text Analytics, las siguientes sugerencias pueden conseguir que la interfaz de trabajo interactiva resulte más accesible.

- **Cuadros de diálogo de edición.** Tal vez necesite pulsar el Tabulador para cambiar el foco al primer control al entrar en un cuadro de diálogo.
- **Cierre de los cuadros de texto extendidos.** Utilice Ctrl+Tabulador para salir de los cuadros de texto extendidos y pasar al siguiente control. *Note:* Ctrl+Tabulador también se utiliza para salir de los controles de tabla.
- **Escribir la primera letra para buscar un elemento en una lista de árbol.** Al buscar un elemento en el panel de categorías, el panel de resultados extraídos o el árbol de bibliotecas, puede escribir la primera letra de un elemento si el panel está enfocado. Al hacerlo, se seleccionará la siguiente instancia de un elemento que comience por la letra que ha escrito.
- **Listas desplegables.** En una lista desplegable de cuadros de diálogo, puede utilizar la barra espaciadora para seleccionar un elemento y cerrar la lista.

En los siguientes temas se describen con detalle otras sugerencias de utilización.

Interferencias con otro software

Al probar IBM SPSS Modeler con lectores de pantallas, como JAWS, nuestro equipo de desarrollo descubrió que la utilización de un servidor de gestión de sistemas (SMS, del inglés 'Systems Management Server) en su empresa puede interferir en la capacidad de JAWS para leer aplicaciones basadas en Java, como IBM SPSS Modeler. Al desactivar SMS se corrige esta situación. Visite el sitio Web de Microsoft si desea obtener más información sobre SMS.

JAWS y Java

Distintas versiones de JAWS ofrecen diversos niveles de soporte para aplicaciones de software basadas en Java. Aunque IBM SPSS Modeler funciona con todas las versiones recientes de JAWS, algunas pueden producir pequeños problemas al utilizarlas con sistemas basados en Java. Visite el sitio Web de JAWS para Windows en <http://www.FreedomScientific.com>.

Utilización de gráficos en IBM SPSS Modeler

Las visualizaciones de información (como los histogramas, diagramas de evaluación, gráficos múltiples y diagramas de dispersión) son difíciles de interpretar para un lector de pantallas. Recuerde, sin embargo, que las distribuciones y los gráficos de malla se pueden visualizar utilizando un resumen textual disponible desde la ventana de resultados.

Capítulo 15. Compatibilidad con Unicode

Compatibilidad con Unicode en IBM SPSS Modeler

IBM SPSS Modeler es totalmente compatible con Unicode, tanto en IBM SPSS Modeler como IBM SPSS Modeler Server. Esto posibilita el intercambio de datos con otras aplicaciones que admiten Unicode, incluidas las bases de datos multilingües, sin pérdida de información que pudiera haberse originado por la conversión desde o a una combinación de codificación específicas de la configuración regional.

- IBM SPSS Modeler almacena los datos Unicode internamente y puede leer y escribir datos multilingües almacenados como Unicode en bases de datos sin peligro de pérdida.
- IBM SPSS Modeler puede leer y escribir archivo de texto codificados con UTF-8. La importación y exportación de archivos de texto se configura de forma predeterminada con codificación regional, pero admite UTF-8 como alternativa. Esta configuración se puede especificar en los nodos de importación y exportación de archivos, pero también se puede cambiar la configuración predeterminada en el cuadro de diálogo de propiedades de la ruta. Consulte el tema “Configuración de opciones generales de las rutas” en la página 42 para obtener más información.
- Los archivos de datos de texto, Estadísticas y SAS almacenados con codificación regional se convierten en UTF-8 al importarse y volverse a exportar. Al escribir en cualquier archivo, si hay caracteres Unicode que no existen en el conjunto de caracteres de la configuración regional, esto se sustituye previa aparición de una advertencia. Esto debería ocurrir únicamente en el caso en que los datos se hayan importado de un origen de datos compatible con Unicode (una base de datos o un archivo de texto UTF-8) que contenga caracteres de un conjunto de caracteres de una o varias configuraciones regionales diferentes.
- Las imágenes de IBM SPSS Modeler Solution Publisher tienen codificación UTF-8 y son totalmente intercambiables entre plataformas y configuraciones regionales.

Acerca de Unicode

El propósito del estándar Unicode es proporcionar una forma coherente de codificar texto multilingüe para que se pueda visualizar correctamente a través de diferentes fronteras, configuraciones locales y aplicaciones. El estándar Unicode, ahora en su versión 4.0.1, define un conjunto de caracteres formado por un superconjunto de todos los conjuntos de caracteres de uso común en el mundo actual en el que se asigna a cada carácter un nombre y punto de código exclusivo. Los caracteres y sus puntos de codificación son idénticos a los del conjunto de caracteres universales (UCS, del inglés Universal Character Set) definidos por ISO-10646. Si desea obtener más información, consulte la página de inicio de Unicode.

Avisos

Esta información se ha desarrollado para los productos y servicios ofrecidos en todo el mundo.

Es posible que IBM no ofrezca los productos, servicios o características tratados en este documento en otros países. Consulte al representante local de IBM para obtener información sobre los productos y servicios actualmente disponibles en su zona. Las referencias a un programa, producto o servicio de IBM no pretenden afirmar ni implicar que solo se pueda utilizar el producto, programa o servicio de IBM. También se puede utilizar otro producto, programa o servicio que tenga la misma función y no infrinja el derecho de propiedad intelectual de IBM. Sin embargo, es responsabilidad del usuario evaluar y comprobar el funcionamiento de todo producto, programa o servicio que no sea de IBM.

IBM puede tener patentes o solicitudes de patentes pendientes que cubran la materia descrita en este documento. Este documento no le otorga ninguna licencia para estas patentes. Puede enviar preguntas acerca de las licencias, por escrito, a:

IBM Director of Licensing
IBM Corporation
North Castle
Drive
Armonk, NY 10504-1785
EE.UU.

Para consultas sobre licencias relativas a la información de doble byte (DBCS), póngase en contacto con el Departamento de Propiedad Intelectual de IBM en su país o envíe las consultas, por escrito, a:

Intellectual Property Licensing
Legal and Intellectual Property Law
IBM Japan Ltd.
1623-14, Shimotsuruma, Yamato-shi
Kanagawa 242-8502, Japón

El párrafo siguiente no se aplica al Reino Unido ni a ningún otro país en que dichas disposiciones entren en contradicción con la legislación local: INTERNATIONAL BUSINESS MACHINES CORPORATION PROPORCIONA ESTA PUBLICACIÓN "TAL CUAL" SIN GARANTÍA DE NINGÚN TIPO, NI EXPLÍCITA NI IMPLÍCITA, INCLUYENDO, PERO NO LIMITÁNDOSE, A LAS GARANTÍAS IMPLÍCITAS DE NO VULNERABILIDAD, COMERCIALIZACIÓN O ADECUACIÓN A UN PROPÓSITO DETERMINADO. Algunos estados no permiten la renuncia a expresar o a garantías implícitas en determinadas transacciones , por lo tanto , esta declaración no se aplique a usted.

Esta información puede incluir imprecisiones técnicas o errores tipográficos. Periódicamente, se efectúan cambios en la información aquí y estos cambios se incorporarán en nuevas ediciones de la publicación. IBM puede realizar mejoras y/o cambios en el producto o productos y/o en los programas que se describen en esta publicación, en cualquier momento y sin previo aviso.

Cualquier referencia a sitios Web que no sean de IBM en esta información solamente es ofrecida por comodidad y de ningún modo sirve como aprobación de esos sitios Web. Los materiales de esos sitios web no forman parte de los materiales de este producto de IBM y la utilización de esos sitios web será responsabilidad del usuario.

IBM puede utilizar o distribuir cualquier información que se le proporcione en la forma que considere adecuada, sin incurrir por ello en ninguna obligación para con el remitente.

Los licenciatarios de este programa que deseen obtener información sobre el mismo con el fin de permitir: (i) el intercambio de información entre programas creados independientemente y otros programas (incluido éste) y (ii) el uso mutuo de información que se haya intercambiado, deben ponerse en contacto con:

Tel. 901 100 400
Atención: Licensing
200 W. Madison St.
Chicago, IL; 60606
Estados Unidos

Esta información estará disponible, bajo las condiciones adecuadas, incluyendo en algunos casos el pago de una cuota.

El programa bajo licencia descrito en este documento y todo el material bajo licencia disponible los proporciona IBM bajo los términos del Acuerdo de cliente de IBM, el Acuerdo internacional de programas bajo licencia de IBM o cualquier acuerdo equivalente entre las partes.

Cualquier dato de rendimiento mencionado aquí ha sido determinado en un entorno controlado. Por lo tanto, los resultados obtenidos en otros entornos operativos pueden variar de forma significativa. Es posible que algunas mediciones se hayan realizado en sistemas en desarrollo y no existe ninguna garantía de que estas medidas sean las mismas en los sistemas comerciales. Además, es posible que algunas mediciones hayan sido estimadas a través de extrapolación. Los resultados reales pueden variar. Los usuarios de este documento deben consultar los datos que corresponden a su entorno específico.

Se ha obtenido información acerca de productos que no son de IBM de los proveedores de esos productos, de sus publicaciones anunciadas o de otros orígenes disponibles públicamente. IBM no ha probado dichos productos y no puede confirmar la precisión de su rendimiento, la compatibilidad ni contemplar ninguna otra reclamación relacionada con los productos que no son de IBM. Las preguntas acerca de las aptitudes de productos que no sean de IBM deben dirigirse a los proveedores de dichos productos.

Todas las declaraciones sobre el futuro del rumbo y la intención de IBM están sujetas a cambio o retirada sin previo aviso y representan únicamente metas y objetivos.

Esta información contiene ejemplos de datos e informes utilizados en operaciones comerciales diarias. Para ilustrarlos lo máximo posible, los ejemplos incluyen los nombres de las personas, empresas, marcas y productos. Todos esos nombres son ficticios y cualquier parecido con los nombres y direcciones utilizados por una empresa real es pura coincidencia.

Si está viendo esta información en copia electrónica, es posible que las fotografías y las ilustraciones en color no aparezcan.

Marcas comerciales

IBM, el logotipo de IBM e ibm.com son marcas registradas o marcas comerciales de International Business Machines Corp., registradas en muchas jurisdicciones en todo el mundo. Puede que otros productos o nombres de servicio sean marcas registradas de IBM u otras compañías. Hay disponible una lista actual de marcas registradas de IBM en la web en "Información de marca registrada y copyright" en www.ibm.com/legal/copytrade.shtml.

Intel, el logotipo de Intel, Intel Inside, el logotipo de Intel Inside, Intel Centrino, el logotipo de Intel Centrino, Celeron, Intel Xeon, Intel SpeedStep, Itanium y Pentium son marcas comerciales o marcas registradas de Intel Corporation o sus filiales en Estados Unidos y otros países.

Linux es una marca registrada de Linus Torvalds en Estados Unidos, otros países o ambos.

Microsoft, Windows, Windows NT, y el logotipo de Windows son marcas comerciales de Microsoft Corporation en Estados Unidos, otros países o ambos.

UNIX es una marca registrada de The Open Group en Estados Unidos y otros países.

Java y todas las marcas comerciales y logotipos basados en Java son marcas comerciales o marcas registradas de Oracle y/o sus afiliados.

Otros productos y nombres de servicio pueden ser marcas comerciales de IBM u otras empresas.

Índice

A

- aadvertencias
 - opciones de configuración 152
- accesibilidad 167, 177
 - características de IBM SPSS Modeler 167
 - ejemplo 174, 175
 - sugerencias de IBM SPSS Modeler 176
- activar nodos 37
- actualización
 - nodos de origen 42
- actualización de modelos 133, 136
- adaptación de rutas a la vista 22
- adición
 - a un proyecto 144
- admisión de fecha 43
- advertencias 47
- ajuste del tamaño 21
- almacenamiento
 - estados 61
 - nodos 61
 - objetos de resultados 61
 - rutas 61
 - varios objetos 61
- almacenar objetos en el repositorio de IBM SPSS Collaboration and Deployment Services 122
- análisis basado en árboles
 - aplicaciones típicas 25
- Análisis Champion/Challenger 120, 133
- anotación
 - nodos 55, 60
 - rutas 55, 60
- anotaciones
 - carpeta 146
 - conversión en comentarios 59
 - proyecto 146
- añadir conexiones de IBM SPSS Modeler Server 13
- apertura
 - estados 62
 - modelos 62
 - nodos 62
 - proyectos 144
 - resultados 62
 - rutas 62
- aplicaciones 25
- aplicaciones de minería de datos 26
- Aplicaciones predictivas 133
- aplicaciones típicas 25
- aprendizaje de las máquinas 25
- árboles de decisión
 - accesibilidad 176
- archivo de diccionario 175
- archivos datos de texto
 - codificación 179
- archivos de copia de seguridad de rutas
 - restauración 61
- archivos de estadísticas
 - codificación 179

- archivos de PowerPoint 144
- archivos de registro
 - presentación del SQL generado 46
- archivos de resultados
 - almacenamiento 61
- archivos de SAS
 - codificación 179
- asignación de nombres a nodos y rutas 60
- atajos
 - teclado 22, 168, 169, 171, 172, 173
 - uso general 65
- atajos de teclado 168, 169, 171, 172, 173
- atributo 25
- automatización 71

B

- barra de herramientas 19
- base de datos
 - funciones 79
- bloquear objetos de repositorio de IBM SPSS Collaboration and Deployment Services 129
- bloqueo de nodos 41
- botón central del ratón
 - simulación 22, 36
- buscar objetos en el repositorio de IBM SPSS Collaboration and Deployment Services 127
- búsqueda
 - de nodos en una ruta 52
- búsqueda de conexiones en COP 13

C

- cadena 85, 86
 - coincidencia 75
 - manipulación de expresiones CLEM 75
 - sustitución 75
- cambio del nombre
 - nodos 60
 - rutas 53
- campos 25, 85, 87
 - en expresiones CLEM 80
 - presentación de valores 80
- campos de correlaciones 63
- campos de plantilla 65
- campos de tiempo
 - conversión 109
- campos esenciales 63, 64
- campos obligatorios 65
- cancelación de correlación de campos 63
- carácter de barra inclinada invertida en expresiones CLEM 86
- caracteres 85, 86
 - caracteres especiales
 - eliminación de cadenas 75

- carga
 - estados 62
 - nodos 62
- carpetas, repositorio de IBM SPSS Collaboration and Deployment Services 129, 130
- caso 25
- cifras decimales
 - formatos de presentación 44
- clases 19, 143, 144
- CLEM 79
 - comprobación de expresiones 81
 - ejemplos 71
 - expresiones 74, 85
 - funciones 79
 - generación de expresiones 79
 - introducción 24, 71
 - lenguaje 85
 - tipos de datos 85, 86, 87
- cliente
 - directorio predeterminado 152
- codificación 42, 179
 - codificación de textos 42
 - codificación predeterminada de la ruta 42
 - codificación UTF-8 42, 179
- colores
 - configuración 154
- colores de script
 - configuración 155
- coma 42
- comentarios
 - atajos de teclado 172, 173
 - en nodos y rutas 55
 - incluir todos en una ruta 59
- compatibilidad con Unicode. 179
- comprobación de expresiones CLEM 81
- concatenación de cadenas 92
- condiciones 74
- conexiones
 - a IBM SPSS Modeler Server 12, 13
 - al repositorio de IBM SPSS Collaboration and Deployment Services 121
 - clúster de servidores 13
- configuración regional
 - opciones 151
- conformidad 508 167
- conjuntos 42
- conjuntos de categorías múltiples
 - en expresiones CLEM 78
- conjuntos de dicotomía múltiple
 - en expresiones CLEM 78
- conjuntos de reglas
 - evaluación 42
- conjuntos de respuestas múltiples
 - en expresiones CLEM 78
- conjuntos de varias respuestas
 - en expresiones CLEM 80
- contraseña
 - IBM SPSS Modeler Server 12

- convenciones 91
- Coordinator of Processes 13
- COP 13
- copiar 19
- correlación de datos 64
- corrientes
 - adición a proyectos 144
 - adición de comentarios 55
 - adición de nodos 36, 38
 - almacenamiento 61
 - anotación 55, 60
 - archivos de copia de seguridad 61
 - cambio del nombre 53, 60
 - carga 62
 - conexión de nodos 36
 - desactivación de nodos 37
 - ejecución 54
 - generación 35
 - omisión de nodos 37
 - opciones 41, 42, 43, 44, 46
 - visualización de tiempos de ejecución 48
- cortar 19
- count_non_nulls function 93
- creación de paleta personalizada 157
 - creación de una subpaleta 158
- CRISP-DM 19, 143
 - vista de proyectos 143
- cuadro de diálogo de arranque 154
- cuadro de diálogo de bienvenida 154

D

- datos
 - vista previa 41
- datos con ruido 26
- desactivar nodos 37, 38
- desbloquear objetos de repositorio de IBM SPSS Collaboration and Deployment Services 129
- descripciones de ruta 53, 54
- deshacer 19
- desplazamiento
 - opciones de configuración 46
- despliegue 120
- despliegue de casos 133
- detener ejecución 19
- directorio
 - valor predeterminado 152
- directorio temporal 14
- distribución chi-cuadrado
 - funciones de probabilidad 98
- distribución f
 - funciones de probabilidad 98
- distribución normal
 - funciones de probabilidad 98
- distribución t
 - funciones de probabilidad 98
- documentación 3
- DTD 140

E

- ejecución de rutas 54
- ejemplos
 - conceptos básicos 5

- ejemplos (*continuación*)
 - Manual de aplicaciones 3
- ejemplos de aplicaciones 3
- enteros 85
- escenarios 132
 - definido 120
 - opciones de despliegue 133
- espacio en blanco
 - eliminación de cadenas 75, 100
- espacios
 - eliminación de cadenas 75, 100
- estados
 - almacenamiento 61
 - carga 62
- etiquetas
 - presentación 42
 - resumen 140
 - value 140
- etiquetas, objeto de repositorio de IBM SPSS Collaboration and Deployment Services 132
- etiquetas de versión, objeto de repositorio de IBM SPSS Collaboration and Deployment Services 132
- exportación
 - descripciones de ruta 54
- exportar
 - PMML 140
- expresiones 85
- expresiones CLEM
 - rendimiento 165
- extracción de conocimiento 25

F

- factorial 176
- fechas
 - conversión 109
 - gestión 109
- formato de presentación de las monedas 44
- formatos de fecha 43, 87, 88
- formatos de hora 43, 87, 88
- formatos de presentación
 - científica 44
 - cifras decimales 44
 - moneda 44
 - números 44
 - símbolo de agrupación 44
- formatos de presentación de los números 44
- fuentes 154, 155
- función @BLANK 69, 92, 115
- función @DIFF 109
- función @FIELD 69, 115
- función @FIELDS_BETWEEN 69, 77, 115
- función @FIELDS_MATCHING 69, 77, 115
- función @INDEX 109
- función @LAST_NON_BLANK 109, 115
- función @MAX 109
- función @MEAN 109
- función @MIN 109
- Función @MULTI_RESPONSE_SET 78, 115
- función @NULL 69, 92, 115
- función @OFFSET 109

- función @OFFSET (*continuación*)
 - consideraciones sobre el rendimiento 165
- función @PARTITION_FIELD 115
- función @PREDICTED 115
- función @SDEV 109
- función @SINCE 109
- función @TARGET 115
- función @TESTING_PARTITION 115
- función @THIS 109
- función @TODAY 105
- función @TRAINING_PARTITION 115
- función
 - @VALIDATION_PARTITION 115
- función abs 97
- función allbutfirst 100
- función allbutlast 100
- función alphabefore 100
- función arccos 98
- función arccosh 98
- función arcsin 98
- función arcsinh 98
- función arctan 98
- función arctan2 98
- función arctanh 98
- función cdf_chisq 98
- función cdf_f 98
- función cdf_normal 98
- función cdf_t 98
- función cos 98
- función cosh 98
- función count_equal 77, 93
- función count_greater_than 77, 93
- función count_less_than 77, 93
- función count_not_equal 77, 93
- función count_nulls 69, 77, 93
- función count_substring 100
- función date_before 93
- Función datetime_date 92
- función DIFF 109
- función div 97
- función endstring 100
- función exponential 97
- Función first_index 78, 93
- Función first_non_null 78, 93
- Función first_non_null_index 78, 93
- función fracof 97
- función hasendstring 100
- función hasmidstring 100
- función hasstartstring 100
- función hassubstring 100
- función INDEX 109
- función integer_bitcount 99
- función integer_leastbit 99
- función integer_length 99
- función intof 97
- Función is_date 92
- Función is_datetime 92
- Función is_integer 92
- Función is_number 92
- Función is_real 92
- Función is_string 92
- Función is_time 92
- Función is_timestamp 92
- función isalphacode 100
- función isendstring 100
- función islowercode 100

- función ismidstring 100
- función isnumbercode 100
- función isstartstring 100
- función issubstring 100
- función issubstring_count 100
- función issubstring_lim 100
- función isuppercode 100
- Función last_index 78, 93
- función LAST_NON_BLANK 109
- Función last_non_null 78, 93
- Función last_non_null_index 78, 93
- función length 100
- función locchar 100
- función locchar_back 100
- función log 97
- función log10 97
- función lowertoupper 100
- función matches 100
- función max 93
- función MAX 109
- Función max_index 78, 93
- función max_n 77, 93
- función MEAN 109
- función mean_n 77, 97
- función member 93
- función min 93
- función MIN 109
- Función min_index 78, 93
- función min_n 77, 93
- función mod 97
- función negate 97
- función OFFSET 109
- función oneof 100
- función pi 98
- función power (exponential) 97
- función random 100
- función random0 100
- función rem 97
- función replace 100
- función replicate 100
- función round 97
- función SDEV 109
- función sdev_n 77, 97
- función sign 97
- función sin 98
- función SINCE 109
- función sinh 98
- función skipchar 100
- función skipchar_back 100
- función soundex 105
- función soundex_difference 105
- función sqrt 97
- función startstring 100
- función stripchar 100
- función strmember 100
- función subscrs 100
- función substring 100
- función substring_between 100
- función SUM 109
- función sum_n 77, 97
- función tan 98
- función tanh 98
- función tesbit 99
- función THIS 109
- función time_before 93
- función to_date 92, 105
- función to_datetime 105

- Función to_datetime 92
- Función to_integer 92
- Función to_number 92
- Función to_real 92
- Función to_string 92
- función to_time 92, 105
- función to_timestamp 92, 105
- función trim 100
- función trim_start 100
- función trimend 100
- función undef 115
- función unicode_char 100
- función unicode_value 100
- función uppertolower 100
- Función value_at 78, 93
- funciones 87, 88, 91, 92, 109
 - @BLANK 69
 - @FIELD 79, 115
 - @GLOBAL_MAX 114
 - @GLOBAL_MEAN 114
 - @GLOBAL_MIN 114
 - @GLOBAL_SDEV 114
 - @GLOBAL_SUM 114
 - @PARTITION 115
 - @PREDICTED 79, 115
 - @TARGET 79, 115
- base de datos 79
- ejemplos 71
- en expresiones CLEM 79
- funciones definidas por el usuario (UDF) 79
- gestión de valores perdidos 69
- funciones a nivel de bit 99
- funciones CLEM
 - valores perdidos 69
- Funciones CLEM
 - aleatorio 100
 - cadena 100
 - comparación 93
 - fecha y hora 105
 - globales 114
 - información 92
 - probabilidad 98
 - secuencia 109
 - trigonómicas 98
- funciones de cadena 100
- funciones de CLEM
 - a nivel de bit 99
 - de conversión 92
 - funciones especiales 115
 - lista de funciones disponibles 90
 - lógicas 96
 - numéricas 97
 - vacíos y nulos 115
- funciones de comparación 93
- funciones de conversión 92
- funciones de distribución 98
- funciones de fecha 87, 88
 - date_before 93, 105
 - date_days_difference 105
 - date_in_days 105
 - date_in_months 105
 - date_in_weeks 105
 - date_in_years 105
 - date_months_difference 105
 - date_weeks_difference 105
 - date_years_difference 105

- funciones de fecha (*continuación*)
 - función @TODAY 105
- funciones de fecha y hora
 - datetime_date 105
 - datetime_day 105
 - datetime_day_name 105
 - datetime_day_short_name 105
 - datetime_hour 105
 - datetime_in_seconds 105
 - datetime_minute 105
 - datetime_month 105
 - datetime_month_name 105
 - datetime_month_short_name 105
 - datetime_now datetime_second 105
 - datetime_time 105
 - datetime_timestamp 105
 - datetime_weekday 105
 - datetime_year 105
- funciones de hora 87, 88
 - time_before 93, 105
 - time_hours_difference 105
 - time_in_hours 105
 - time_in_mins 105
 - time_in_secs 105
 - time_mins_difference 105
 - time_secs_difference 105
- funciones de hora y fecha 87, 88
- funciones de información 92
- funciones de probabilidad 98
- funciones de secuencia 109
- funciones definidas por el usuario (UDF) 79
- funciones especiales 115
- funciones globales 114
- funciones if, then, else 96
- funciones lógicas 96
- funciones numéricas 97
- funciones trigonométricas 98

G

- generación de SQL
 - presentación preliminar 46
 - registro 46
- Generador de expresiones 171
 - acceso 79
 - conceptos básicos 79
 - utilización 79
- gestores 17
- grados
 - unidades de medida 44
- gráficos
 - adición a proyectos 144
 - almacenamiento de resultados 61

H

- herramienta de correlación de datos 63

I

- IBM InfoSphere Warehouse (ISW)
 - Exportar PMML 140
- IBM SPSS Collaboration and Deployment Services 120

- IBM SPSS Collaboration and Deployment Services Enterprise View 120, 133
- IBM SPSS Modeler 1, 15
 - características de accesibilidad 167
 - conceptos básicos 11, 151
 - documentación 3
 - ejecución desde la línea de comandos 11
 - opciones 151
 - primeros pasos 11
 - sugerencias y atajos 65
- IBM SPSS Modeler Advantage 120, 132
- IBM SPSS Modeler Server 1
 - contraseña 12
 - ID de usuario 12
 - nombre de dominio (Windows) 12
 - nombre de host 12, 13
 - número de puerto 12, 13
- iconos
 - opciones de configuración 22, 46
- ID de usuario
 - IBM SPSS Modeler Server 12
- idioma
 - opciones 151
- importar
 - PMML 140
- impresión 23
 - rutas 22, 38
- incluir todos los comentarios de una ruta 59
- información sobre herramientas
 - anotación de nodos 60
- informes
 - adición a proyectos 144
 - almacenamiento de resultados 61
 - configuración de propiedades 147
 - generación 147
- iniciar sesión en IBM SPSS Modeler Server 12
- inicio de sesión único 12
- inicio de sesión único, repositorio de IBM SPSS Collaboration and Deployment Services 119, 121
- introducción 85
 - IBM SPSS Modeler 11, 151

J

- Java 177
- JAWS 167, 175, 177

L

- lectores de pantallas 169, 171, 175, 176
 - ejemplo 174, 175
- lienzo 15
- lienzo de rutas
 - configuración 46
- línea de comandos
 - iniciar IBM SPSS Modeler 11
- listas 85, 86

M

- memoria
 - administración 151

- memoria caché
 - activación 151
 - almacenamiento 39
 - configuración de una caché 38
 - vaciado 39, 42
- mensajes
 - presentación del SQL generado 46
- mensajes de error 47
- minería de datos 25
 - ejemplos de aplicaciones 34
 - estrategia 27
- minimizar 21
- modelado
 - rama 55
- modelo del proceso CRISP-DM 27, 28
- modelos 55
 - actualización 136
 - adición a proyectos 144
 - almacenar en el repositorio de IBM SPSS Collaboration and Deployment Services 126
 - exportación 155
 - sustitución 152
- modelos PMML
 - regresión lineal 155
 - regresión logística 155

N

- navegación
 - atajos de teclado 168
- nodo Agregar
 - rendimiento 163
- nodo Auditoría de datos
 - uso en la exploración 25
- Nodo Auditoría de datos
 - uso en la minería de datos 26
- nodo Calidad
 - valores perdidos 68
- nodo de archivo caché
 - carga 62
- nodo de creación de regla
 - carga 62
- nodo de relleno
 - valores perdidos 69
- nodo de tipo
 - valores perdidos 69
- nodo Distinguir
 - rendimiento 163
- nodo Enterprise View 133
- nodo Evaluación
 - rendimiento 163
- nodo Fundir
 - rendimiento 163
- nodo Intervalos
 - rendimiento 163
- Nodo K-medias
 - conjuntos grandes 42
 - rendimiento 164
- Nodo Kohonen
 - conjuntos grandes 42
 - rendimiento 164
- nodo Ordenar
 - rendimiento 163
- nodo Red neuronal
 - conjuntos grandes 42
 - rendimiento 164

- nodo Selección de características
 - valores perdidos 68
- nodo Tipo
 - rendimiento 163
- nodos 11
 - activación 37
 - adición 36, 38
 - adición a proyectos 144
 - adición de comentarios a 55
 - almacenamiento 61
 - almacenar en el repositorio de IBM SPSS Collaboration and Deployment Services 125
 - anotación 55, 60
 - bloqueo 41
 - búsqueda de 52
 - carga 62
 - conexión en una ruta 36
 - creación de paleta personalizada 157
 - creación de una subpaleta
 - personalizada 158
 - desactivación 37, 38
 - desactivación en una ruta 37
 - duplicado 38
 - edición 38
 - eliminación 36
 - eliminación de conexiones 38
 - eliminación de la paleta 157
 - introducción 36
 - omisión en una ruta 37
 - opciones de configuración 38
 - orden de 161
 - personalización de la pestaña de paleta 158
 - rendimiento 163, 164
 - tiempos de ejecución 48
 - vista previa de datos 41
 - visualización de la paleta 157
- nodos de modelado 36
 - personalización de la pestaña de paleta de modelado 158
 - rendimiento 164
- nodos de origen 36
 - actualización 42
 - correlación de datos 63
- nodos de proceso 36
 - rendimiento 163
- nodos de resultados 36
- nodos terminales 36
- nombre de dominio (Windows)
 - IBM SPSS Modeler Server 12
- nombre de host
 - IBM SPSS Modeler Server 12, 13
- nombres de nodos 60
- nombres de rutas 60
- notación científica
 - formato de presentación 44
- notificaciones
 - opciones de configuración 152
- nuevas características 7
- nugget 55
 - definido 17
- nuggets de modelo 55
- nulos 76
- número de puerto
 - IBM SPSS Modeler Server 12, 13
- números 76, 85, 86

O

- objetos
 - propiedades 147
- objetos de resultados
 - almacenar en el repositorio de IBM SPSS Collaboration and Deployment Services 125
- opciones 151
 - de IBM SPSS Modeler 151
 - display 154
 - PMML 155
 - propiedades de ruta 41, 42, 43, 44, 46, 47, 48
 - sintaxis 155
 - usuario 152
- opciones de despliegue
 - escenarios 133
- opciones de usuario 152
- operador and 96
- operador equals 93
- operador mayor que 93
- operador menor que 93
- operador not 96
- operador not equal 93
- operador or 96
- operadores
 - en expresiones CLEM 79
 - unión de cadenas 92

P

- palabras clave
 - anotación de nodos 60
- paleta de modelos generados 17
- paleta modelos 126
- paletas 15
 - personalización 156
- parámetros
 - en expresiones CLEM 80
 - generación de modelos 135
 - puntuación 135
 - ruta 49, 50
 - sesión 49, 50
 - solicitudes en tiempo de ejecución 49
 - tipo 50
 - uso en escenarios 135
- parámetros de ruta 49, 50
- parámetros de sesión 49, 50
- pegar 19
- personalización de la pestaña de paleta 158
- plantillas 63
- PMML
 - exportación de modelos 140
 - importación de modelos 140
 - opciones de exportación 155
- prioridad 88
- prioridad del operador 88
- programación visual 15
- propiedades
 - carpeta de proyecto 146
 - de rutas de datos 41
 - fases de informe 147
- propiedades de objeto, repositorio de IBM SPSS Collaboration and Deployment Services 131

- propiedades de ruta
 - Analytic Server 47
- proyectos 19, 143
 - adición de objetos 144
 - almacenar en el repositorio de IBM SPSS Collaboration and Deployment Services 124
 - anotación 146
 - cierre 147
 - configuración de propiedades 146
 - creación de nuevos 144
 - del repositorio de IBM SPSS Collaboration and Deployment Services 145
 - establecimiento de una carpeta predeterminada 143
 - generación 144
 - generación de informes 147
 - propiedades de carpeta 146
 - propiedades de objeto 147
 - Vista Clases 144
 - vista CRISP-DM 143
- punto 42
- puntuación
 - rama 55, 135, 137

R

- radianes
 - unidades de medida 44
- ramas, modelado y puntuación 55, 135, 137
- ratón
 - utilizar en IBM SPSS Modeler 22, 36
- reales 85, 86
- recuperar objetos del repositorio de IBM SPSS Collaboration and Deployment Services 126
- registros 25
 - valores perdidos 68
- regresión lineal
 - exportar como PMML 155
- regresión logística 176
- exportar como PMML 155
- regression 176
- rendimiento
 - de nodos de modelado 164
 - de nodos de proceso 163
 - expresiones CLEM 165
- Repositorio de IBM SPSS Collaboration and Deployment Services 119, 120
 - almacenamiento de objetos 122
 - bloqueo y desbloqueo de objetos 129
 - búsqueda en 127
 - carpetas 129, 130
 - conexión con 121
 - eliminación de objetos y versiones 129
 - explorar 121
 - inicio de sesión único 121
 - propiedades de objeto 131
 - recuperación de objetos 126
 - transferencia de proyectos a 145
- resultado en formato HTML
 - lector de pantallas 176
- resultados 17
- ruta 15

- rutas 11
 - adaptación a la vista 22
 - adición a proyectos 144
 - almacenar en el repositorio de IBM SPSS Collaboration and Deployment Services 124
 - opciones de despliegue 133
- rutas de datos
 - generación 35

S

- scripts 24, 71
- selección de paleta de nodos 157
- servidor
 - adición de conexiones 13
 - búsqueda de servidores en COP 13
 - directorio predeterminado 152
 - inicio de sesión 12
- símbolo de agrupación
 - formatos de presentación de los números 42
- símbolo decimal
 - formatos de presentación de los números 42
- sistema
 - opciones 151
- solicitudes, tiempo de ejecución 49
- subpaleta
 - creación 158
 - eliminación de la pestaña de paleta 157
 - visualización en pestaña de paleta 157
- sugerencias
 - para accesibilidad 176
 - uso general 65
- sustitución de modelos 152

T

- tablas 171
 - adición a proyectos 144
 - almacenamiento de resultados 61
- teclas de aceleración 22
- tiempos de ejecución, visualización 48
- tipo de despliegue 133
- tipos de datos 73
 - en parámetros 50
- tratamiento de espacios en blanco
 - funciones de CLEM 115

V

- vacíos 68, 76
- valor predeterminado
 - fase del proyecto 143
- valores 73
 - adición a expresiones CLEM 80
 - presentación desde una auditoría de datos 80
- valores de fecha/hora 76
- valores globales
 - en expresiones CLEM 80
- valores perdidos 68, 76

- valores que faltan
 - en registros 68
 - expresiones CLEM 69
- variables 25
- varias sesiones de IBM SPSS Modeler 14
- Ventana Árbol interactivo
 - accesibilidad 176
- ventana principal 15
- vista previa
 - datos de nodo 41

Z

- zoom 19



Impreso en España