

*Руководство пользователя
IBM SPSS Modeler Entity
Analytics 18.0*

IBM

Примечание

Прежде чем использовать эту информацию и продукт, описанный в ней, прочтите сведения в разделе “Уведомления” на стр. 57.

Информация о продукте

Это издание применимо к версии 18, выпуск 0, модификация 0 IBM SPSS Modeler и ко всем последующим версиям и модификациям до тех пор, пока в новых изданиях не будет указано иное.

Содержание

Предисловие	v
------------------------------	----------

Глава 1. Аналитика объекта.	1
О программе Entity Analytics	1
Entity Analytics и прогнозная аналитика.	2

Глава 2. Entity Analytics и IBM SPSS Modeler.	5
--	----------

Использование Entity Analytics с IBM SPSS Modeler	5
Этап 1: Чтение исходных данных в SPSS Modeler.	6
Этап 2: Создать репозиторий	6
Этап 3: Соединить SPSS Modeler с этим репозиторием	7
Этап 4: Отобразить входные поля на характеристики репозитория	7
Этап 5: Экспорт данных в репозиторий и распознавание объектов.	7
Этап 6: Анализ установленных соответствий	9
Этап 7: Распознавание новых записей по репозиторию	10
Этап 8: Генерирование предупреждений	11

Глава 3. Задачи Entity Analytics.	13
--	-----------

Об этих задачах	13
Конфигурирование репозитория объектов (узел Экспорт EA)	13
Репозиторий объектов	13
Соединение с источником данных	14
Создание репозитория	14
Отображения входных полей на характеристики (узел Экспорт EA)	16
Вывод отображений полей (узел Экспорт EA)	18
Конфигурирование репозитория объектов	18
Просмотр отображений источников данных	19
Управление характеристиками репозитория	19
Добавление или изменение характеристики	21
Анонимизация характеристик репозитория	22
Управление типами объектов.	22
Задание порога для соответствия объектов	25
Повторное использование конфигурации репозитория	25
Сохранение изменений конфигурации	25
Закрытие окна конфигурации.	25
Анализ установленных соответствий (узел источника Entity Analytics (EA))	26
Выбор источника данных	26
Переименование полей данных	27
Задание информации о типе для полей данных	27
Добавление узлов к потоку	27
Сравнение новых наблюдений с репозиторием (узел потоков EA)	28

Отображение полей ввода на характеристики (узел потоков EA)	29
Вывод отображений полей и источников данных (узел потоков EA)	30
Вывод из узла потоков EA	31
Использование IBM SPSS Modeler Entity Analytics с другими продуктами IBM SPSS	31
Административные задачи	32
Конфигурирование назначений портов	32
Управление регистрационными данными администратора управления для базы данных репозитория	33
Перемещение репозитория в другой каталог хранения	33
Задание свойств потока для полей даты/времени и отметки времени.	34
Настройка параметров срока ожидания	34
Выполнение IBM SPSS Modeler Entity Analytics с клиентом SPSS Modeler и сервером SPSS Modeler в той же самой системе Windows	34
Очистка репозитория объектов	35
Удаление ненужных источников данных из репозитория	35
Удаление репозитория объектов.	35
Удаление репозитория, если к нему не удастся подключиться	36

Глава 4. Entity Analytics в действии	37
---	-----------

Об этом примере.	37
Исходная модель	37
Добавление Entity Analytics	40
Получение данных источника в репозиторий.	40
Чтение установленных соответствий	41
Сравнение вывода Entity Analytics с исходной моделью	47
Итог.	51

Приложение. Свойства сценариев для IBM SPSS Modeler Entity Analytics.	53
--	-----------

Сценарии в IBM SPSS Modeler Entity Analytics	53
Общие свойства	53
Свойства entityanalytics_exportnode	53
Свойства entityanalytics_sourcenode	54
Свойства entityanalytics_processnode	54

Уведомления.	57
-------------------------------	-----------

Товарные знаки	58
Правила и условия для документации продукта.	59

Индекс	61
-------------------------	-----------

Предисловие

IBM® SPSS Modeler - это инструментальная среда исследования данных IBM Corp., рассчитанная на работу с предприятием. SPSS Modeler помогает организациям улучшить взаимосвязи с клиентами и отдельными лицами, обеспечивая глубокое понимание данных. Организации используют приобретенные с помощью SPSS Modeler глубокие знания для сохранения выгодных заказчиков, обнаружения возможностей дополнительных покупок, привлечения новых клиентов, обнаружения ошибок, сокращения рисков и улучшений в обеспечении государственных служб.

Наглядный интерфейс SPSS Modeler дает пользователям возможность применить свой конкретный опыт в бизнесе, что способствует разработке более мощных предсказывающих моделей и сокращает время принятия решения. SPSS Modeler предлагает много способов моделирования, таких как алгоритмы предсказания, классификации, сегментации и ассоциативного обнаружения. Когда моделей IBM SPSS Modeler Solution Publisher поддерживает их распространение на уровне организации для принимающих решение сотрудников или для применения к базе данных.

О бизнес аналитике IBM

Программное обеспечение IBM для бизнес аналитики предоставляет полную, последовательную и точную информацию, которая повышает эффективность ведения бизнеса. Полный набор программного обеспечения для business intelligence, прогностической аналитики, управления финансовой эффективностью и стратегией и аналитических приложений позволяет ясно видеть текущую ситуацию, а также делать прогнозы, позволяющие предпринимать практические действия. В сочетании с решениями для конкретных отраслей, проверенной практикой и услугами бизнес аналитика IBM позволяет организациям любых размеров достигать наивысшей производительности, уверенно автоматизировать процессы принятия решений и добиться лучших результатов.

Как составная часть этого набора, программное обеспечение IBM SPSS Predictive Analytics помогает организациям предсказывать будущие события и предпринимать практические действия непосредственно на основе этих предсказаний. Коммерческие, правительственные и научные организации всего мира, полагаются на технологию IBM SPSS, обеспечивающую конкурентное преимущество в привлечении, удержании клиентов и повышения отдачи от них при уменьшении доли ошибочных решений и сокращении рисков. Включая программное обеспечение IBM SPSS в свои ежедневные операции, организации могут прогнозировать будущие события, направлять и автоматизировать решения для соответствия бизнес-целям и достигать ощутимых конкурентных преимуществ. Чтобы получить дальнейшую информацию или связаться с представителем, зайдите на <http://www.ibm.com/spss>.

Техническая поддержка

Техническая поддержка предоставляется клиентам, оплачивающим обновительные взносы. Пользователи могут обращаться в службу технической поддержки, если у них возникают какие-либо проблемы с использованием или установкой программного обеспечения IBM Corp.. К службе технической поддержки можно вызывать через сайт IBM Corp. по адресу <http://www.ibm.com/support>. При обращении за поддержкой будьте готовы назвать себя и организацию, в которой вы работаете.

Глава 1. Аналитика объекта

О программе Entity Analytics

IBM SPSS Modeler Entity Analytics добавляет дополнительное измерение к прогностической аналитике IBM SPSS Modeler. Прогностическая аналитика пытается предсказать будущее поведение данных из прошлого, а объектная аналитика направлена на улучшение связности и согласованности текущих данных посредством устранения конфликтов идентичности в самих записях. Идентичность может относиться к индивидууму, организации, а также к любому другому объекту, для которого возможна неоднозначность. Разрешение идентичности может оказаться крайне необходимым для ряда полей, в том числе для управления отношениями с клиентами, обнаружения мошенничества, противодействия отмыванию денег или для национальной и международной безопасности.

Представьте, что у вас есть следующие записи о клиентах из двух различных источников, и вы не уверены, описан в них один и тот же человек, или разные люди.

Источник 1

Номер записи: 70001
Имя: Jon Smith
Адрес: 123 Main Street
Номер налогоплательщика: 555-00-1111
Водительское удостоверение: 0001133107
Кредитная карта: 10229127

Источник 2

Номер записи: 9103
Имя: JOHNATHAN Smith
Дата рождения: 06/17/1934
Телефон: 555-1212
Кредитная карта: 10229128
Email: jls@mail.com
IP-адрес: 9.50.18.77

Точных совпадений данных между двумя записями нет. Тем не менее, если ввести третий источник, мы найдём некоторые общие атрибуты.

Источник 3

Номер записи: 6251
Имя: Jon Smith
Телефон: 555-1212
Водительское удостоверение: 0001133107
Кредитная карта: 10229132

Номер водительского удостоверения связывает записи из источника 1 и источника 3, тогда как номер телефона связывает записи из источников 2 и 3, Таким образом, мы можем быть уверены, что все три источника описывают одного и того же человека.

Но что, если распознавание не так легко сделать? У нас может быть очень мало данных, на основе которых мы принимаем наше решение. Рассмотрим следующие две записи.

Источник 4

Номер записи: S45286
Имя: John T Smith Jr
Адрес: 456 Main Street
Телефон: 703-555-2000
Дата рождения: 03/12/1984

Номер записи: S45287
Имя: John T Smith
Адрес: 456 Main Street
Телефон: 703-555-2000
Водительское удостоверение: 009900991

Очевидно, что это не тот же самый мистер Смит из предыдущих записей - различий достаточно, чтобы мы могли сделать такой вывод. Однако проблема остается. Две различные записи из одного источника, похоже, относятся к одному и тому же человеку. Являются ли они дубликатами? Мы не можем быть уверены, пока не найдём ещё запись, возможно, из другого источника, которая даст нам больше информации.

Источник 5

Номер записи: 769582-2
Имя: John T Smith Sr
Адрес: 456 Main Street
Телефон: 703-555-2000
Водительское удостоверение: 009900991
Дата рождения: 06/25/1959

Эта запись решает проблему. Две записи в источнике 4 - это не дубликаты, на самом деле они относятся к отцу и сыну с одинаковыми именами, живущим по одному и тому же адресу и использующими один и тот же телефонный номер. Вручную поиск одной записи, которая устанавливает требуемые соотношения, может занять несколько недель. Автоматизированная система Entity Analytics резко сокращает время поиска.

Entity Analytics и прогнозная аналитика

Если бы все ваши данные состояли из единственного источника записей, полных и однозначных, для IBM SPSS Modeler было бы просто разрешать любые конфликты идентификации. Используя только прогнозную аналитику, можно было бы считать свои данные в IBM SPSS Modeler, выполнить обработку и получить надежные результаты.

Но в реальном мире картина несколько иная. Данные обычно далеко не полны, часто неоднозначны и рассеяны по многим различным источникам данных; различных атрибутов много, и лишь некоторые поля перекрываются. Часть ценности Entity Analytics состоит в сборе данных из всех различных источников в единую централизованную область хранения, называемую **репозиторием**. Затем система Entity Analytics исследует подробности данных для разрешения конфликтов, добавляя уникальный идентификатор к записям, которые описывают одного и того же человека или организацию.

В следующей таблице показаны различия между двумя типами аналитики.

Таблица 1. Различия между прогнозной аналитикой и Entity Analytics.

Характеристика	Прогнозная аналитика	Аналитика объекта
Типы данных обучения	Основаны на относительно малых наборах и диапазонах чисел	Могут обрабатывать большие наборы (поля без типов), такие как имена и адреса
Размер данных обучения	Обычно большие наборы (поля без типов) игнорируются	Все используемые данные

Таблица 1. Различия между прогнозной аналитикой и Entity Analytics (продолжение).

Характеристика	Прогнозная аналитика	Аналитика объекта
Обобщение	Алгоритм выполняет обобщение на основе данных обучения для построения краткой модели	Данные представляются в структурах, подходящих для распознавания объектов и обнаружения связей
Обнаружение мошенничества	Записи отмечаются как потенциально мошеннические, если у них есть типичные характеристики мошенничества	Записи отмечаются как потенциально мошеннические, если они связаны с известными мошенническими записями или относятся к тем же самым людям, но с иными идентификационными данными

Глава 2. Entity Analytics и IBM SPSS Modeler

Использование Entity Analytics с IBM SPSS Modeler

Вы подозреваете, что возможны проблемы идентификации с вашими данными. Например, отдельные люди могут упоминаться в данных по несколько раз, разные люди могут быть записаны как один человек, люди могут быть пропущены. Как IBM SPSS Modeler Entity Analytics может помочь с этими проблемами? Ниже описана предлагаемая процедура, хотя вам, возможно, нужно будет изменить ее для удовлетворения вашим конкретным требованиям.

- Прочитать исходные данные в IBM SPSS Modeler
- Создать репозиторий для хранения этих данных
- Соединить IBM SPSS Modeler с этим репозиторием
- Отобразить поля данных на характеристики репозитория
- Экспортировать данные в репозиторий и установить соответствия
- Проанализировать установленные соответствия
- Распознать новые записи по репозиторию
- Сгенерировать все необходимые предупреждения (в пакетном режиме или в реальном времени)

В этот момент надо знать кое-что о том, как работает IBM SPSS Modeler. IBM SPSS Modeler - это удобный для пользователя инструмент на основе графического представления потока данных через множество узлов. Каждый узел представляет определенный этап рабочего потока.

Сам IBM SPSS Modeler предлагает широкий диапазон узлов, покрывающих все стандартные функции исследования данных. IBM SPSS Modeler Entity Analytics добавляет узлы специально для аналитики объектов. Это узел Экспорт EA, узел источника Entity Analytics (EA) и узел обработки потоков EA.

Этот процесс показан на следующем рисунке.

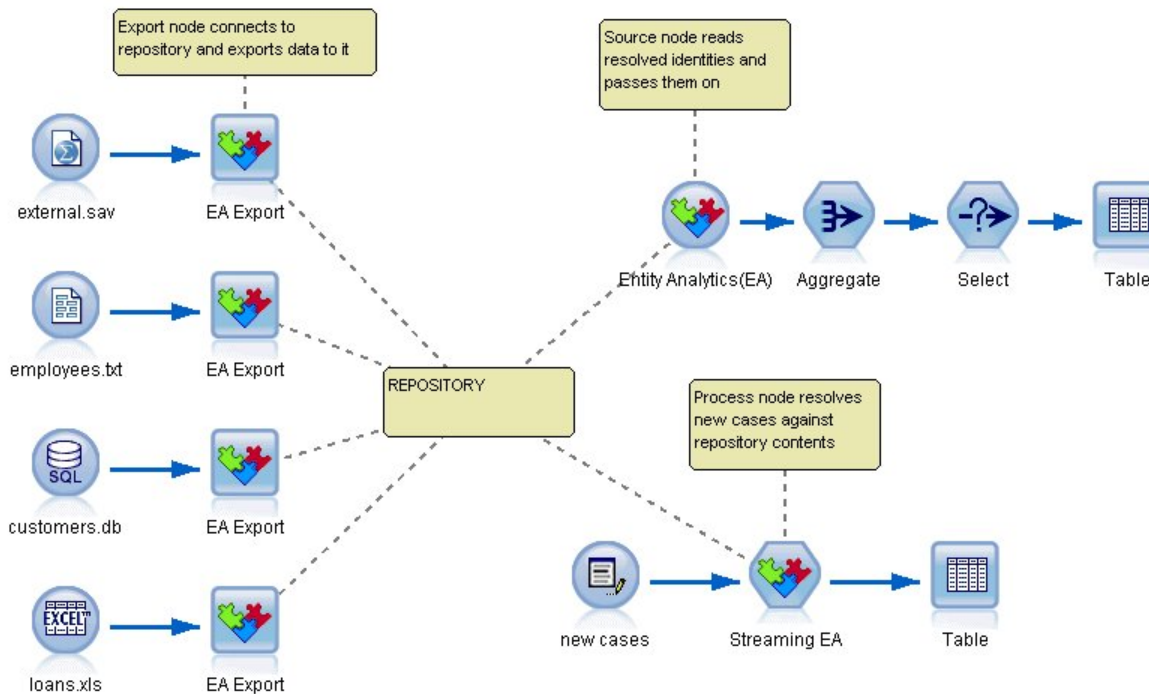


Рисунок 1. Процесс Entity Analytics

Этап 1: Чтение исходных данные в SPSS Modeler

Ваша первая задача - прочитать данные источника в SPSS Modeler с помощью одного или нескольких узлов источников, обозначаемых в SPSS Modeler кружками.

Данные могут быть в любом формате, поддерживаемом SPSS Modeler, например, в текстовом файле, в таблице базы данных, электронной таблице, файле XML и т.п., но каждый различный формат требует соответствующего узла источника SPSS Modeler. В нашем примере это узел источника База данных.

Каждый файл источника данных должен содержать одно поле, которое однозначно идентифицирует каждую запись. Если источника данных не содержит такого поля, его можно добавить в SPSS Modeler.

Дополнительную информацию смотрите в разделе “Добавление уникального идентификатора записи” на стр. 14.

Дополнительную информацию смотрите в разделе “Соединение с источником данных” на стр. 14.

Примечание: Данные с нелатинскими символами не поддерживаются. Если данные состоят из смеси латинских (как в Западной Европе, например) и нелатинских наборов символов, будут разрешаться только записи с латинскими символами.

Этап 2: Создать репозиторий

Главная точка всей вашей работы с Entity Analytics - репозиторий, центральная область хранения, где собираются вместе все ваши записи данных.

Чтобы создать репозиторий, вы начинаете с соединения источника данных с узлом Экспорт EA, показанным как значок квадрата.

Из узла экспорта можно создать новый репозиторий (или выбрать существующий), готовый для приема экспортируемых данных.

Процесс создания репозитория полностью описан ниже. Дополнительную информацию смотрите в разделе “Конфигурирование репозитория объектов (узел Экспорт EA)” на стр. 13.

Примечание: Если вы работаете в режиме удаленного сервера, надо создать репозиторий на компьютере сервера Modeler (то есть при создании репозитория вы должны быть соединены с сервером Modeler с клиента Modeler, чтобы репозиторий EA был создан на компьютере сервера).

Сконфигурировав репозиторий, вы поддерживаете его содержимое различными способами. Дополнительную информацию смотрите в разделе “Конфигурирование репозитория объектов” на стр. 18.

Этап 3: Соединить SPSS Modeler с этим репозиторием

Создав репозиторий, вы соединяете его с потоком SPSS Modeler.

Дополнительную информацию смотрите в разделе “Опции репозитория объектов” на стр. 15.

Этап 4: Отобразить входные поля на характеристики репозитория

Источники данных могут содержать много различных видов информации об объектах. Некоторые типы информации характерны для большинства источников данных, в то время как другие могут относиться только к конкретному источнику данных. В репозитории объектов эти различные типы информации называют **характеристиками**. Репозиторий предоставляет набор стандартных характеристик; можно также создавать собственные характеристики.

Характеристика репозитория - это отдельный тип информации, который может быть использован вместе с источником данных объекта. Некоторые характеристики (например, Имя, Фамилия, Дата рождения и тому подобные) могут использоваться во многих источниках данных, тогда как другие характеристики могут относиться только к конкретному источнику данных. Характеристика обычно эквивалентна либо полю в записи данных, либо столбцу в таблице базы данных.

Создав репозиторий и соединившись с ним, вы указываете одно из полей своих входных данных как поле **ключа уникальности**, используемое в последующем анализе. Вы также отображаете входные поля данных на соответствующие характеристики в репозитории. Отображение на predeterminedную характеристику указывает репозиторию в целом, какие поля сравнивать и, что важнее, каким образом. Узел экспорта EA содержит таблицу отображений, где можно создавать отображения.

Дополнительную информацию смотрите в разделе “Отображения входных полей на характеристики (узел Экспорт EA)” на стр. 16.

Этап 5: Экспорт данных в репозиторий и распознавание объектов

Для каждого узла источника данных требуется свой собственный узел экспорта EA, поэтому если ваши данные распределены по многим различным источникам, ваш поток может содержать несколько источников данных, каждый из которых соединен с отдельным узлом экспорта EA. Дополнительную информацию смотрите в разделе “Использование Entity Analytics с IBM SPSS Modeler” на стр. 5.

Если у вас несколько источников данных, вы можете решить читать записи из одного, некоторых или всех источников данных. Система Entity Analytics анализирует выбранные вами записи и добавляет к каждой из них поле идентификатора с именем \$EA_ID. Там, где две или большее число ранее неоднозначных записей теперь могут быть распознаны, к этим записям добавляются идентификаторы, уникальные в пределах репозитория. Система также добавляет поле указания источника данных, откуда происходит запись.

Вы соединяете каждый узел источника данных с его собственным узлом экспорта EA, отображаете входные поля на характеристики репозитория, и затем запускаете поток, чтобы экспортировать данные из SPSS

Modeler в репозиторий и разрешить конфликты идентификационных данных в ходе одной операции. Чтобы показать, как это работает, предположим, что у вас есть следующие записи в четырех различных источниках данных.

Внешние данные

Таблица 2. Внешние данные

Имя	Телефон	Кредитный риск
Mike	555-1234	560
Joe	555-4567	780

Сотрудники

Таблица 3. Сотрудники

Имя	Адрес	Телефон
Michael	1234 5th Street	555-1234
Fred	543 1st Avenue	555-9876

Заказчики

Таблица 4. Заказчики

Имя	Адрес	Сбережения
Susan	1234 5th Street	\$1234
Joe	777 Oak Street	\$5

Кредиты

Таблица 5. Кредиты

Имя	Адрес	Телефон	Кредит
Sue	1234 5th Street	555-1234	\$10000
Joseph	777 Oak Street	555-4567	\$50000

Как мы видим, вы экспортируете каждый источник данных в репозиторий по очереди. По мере этого репозиторий выполняет распознавание объектов. В репозитории в каждую запись добавляется поле идентификатора (с именем *\$EA-ID*) и поле индикатора источника (с именем *\$EA-SRC*), которое показывает источник данных для этой записи. В нашем примере после экспорта всех четырех источников данных содержимое репозитория будет выглядеть примерно так:

Таблица 6. Пример содержимого репозитория после этапа экспорта.

<i>\$EA-ID</i>	<i>\$EA-SRC</i>	Имя	Телефон	Адрес	Кредитный риск	Сбережения	Кредит
1	Сотрудники	Michael	555-1234	1234 5th St			
1	Внешние данные	Mike	555-1234		560		
2	Заказчики	Joe		777 Oak St		\$5	
2	Внешние данные	Joe	555-4567		780		
2	Кредиты	Joseph	555-4567	777 Oak St			\$50000

Таблица 6. Пример содержимого репозитория после этапа экспорта (продолжение).

SEA-ID	SEA-SRC	Имя	Телефон	Адрес	Кредитный риск	Сбережения	Кредит
3	Сотрудники	Fred	555-9876	543 1st Ave			
4	Заказчики	Susan		1234 5th St		\$1234	
4	Кредиты	Sue	555-1234	1234 5th St			\$10000

Система Entity Analytics определила, что *Mike* в наборе данных *Внешние данные* - это тот же человек, что и *Michael* в наборе данных *Сотрудники*, на основании того, что у них один и тот же номер телефона, и назначила ему ID 1.

Случай *Joe* в наборе данных *Внешние данные* несколько сложнее. Тот ли это человек, что *Joe* в наборе данных *Заказчики*? Только по этим двум источникам это невозможно определить, однако у нас есть третий источник, *Кредиты*, а в нем запись *Joseph*. Теперь у нас есть соответствие: номер телефона в записи *Joseph* совпадает с номером телефона в записи *Joe* в наборе данных *Внешние данные*. На этом основании система решает, что это один и тот же человек, и дает ему идентификатор 2.

Для записи *Fred* нет никаких соответствий, и он получает ID 3. *Susan* из набора данных *Заказчики* идентифицируется как *Sue* в наборе данных *Кредиты*, так как у них один тот же адрес, и ей присваивается ID 4.

Примечание: В целях иллюстрации мы приводим пример оптимистического соответствия. Вы можете выбрать более пессимистический набор правил, чтобы совпадение имени и номера телефона или адреса само по себе не квалифицировалось как точное совпадение, и таким записям не присваивался бы один тот же идентификатор.

Этап 6: Анализ установленных соответствий

Разрешив конфликты идентификационных данных в репозитории, можно выполнить дальнейший анализ и обработку результатов. Например, если вы подозреваете возможные мошеннические действия на основании существования двойных записей для одних и тех же объектов, вы могли бы хотеть составить отчет, перечисляющий такие дубликаты.

Начните с создания узла источника Entity Analytics (EA) и его соединения с репозиторием.

Общие выходные данные узла состоят из следующих полей.

- Поле идентификатора, добавленное системой (*SEA-ID* в примере Этапа 5)
- Поле указания источника, добавленное системой (*SEA-SRC* в примере Этапа 5)
- Поле ключа, которое вы указали на Этапе 4

Кроме того, если вас интересуют взаимосвязи, генерируется следующий вывод: Дополнительную информацию смотрите в разделе “Выбор источника данных” на стр. 26.

- Степень разделения между объектами (*SEA-DEGREE*)
- Поле родительского объекта (*SEA-PARENT*)
- Поле дочернего объекта (*SEA-CHILD*)
- Правило, идентифицирующее взаимосвязь (*SEA-RULE*)

Для просмотра вывода в SPSS Modeler можно присоединить узел вывода SPSS Modeler, такой как узел Таблица или узел Отчет, и запустить эту часть потока. Если надо сделать сводку вывода, который может быть очень большим, можно добавить узлы операций над записями, такие как Агрегация или Выбор.

Узел источника Entity Analytics (EA) полностью описан ниже. Дополнительную информацию смотрите в разделе “Анализ установленных соответствий (узел источника Entity Analytics (EA))” на стр. 26.

Этап 7: Распознавание новых записей по репозиторию

Вы распознали объекты всех записей во всех ваших источниках данных. Но что будет, если вы хотите сравнить новый набор записей, чтобы посмотреть, как они соотносятся с тем, что вы уже знаете, для лучшей оценки? Здесь используется узел потоков EA.

Прежде всего, вы добавляете новый узел источника данных SPSS Modeler для чтения новых данных в поток. Затем вы соединяете этот узел с узлом потоков EA. Для просмотра вывода вы, как и раньше, добавляете узел Таблица.

При выполнении этой части потока узел потоков EA читает каждую новую запись и сравнивает ее с содержимым репозитория. Если он находит соответствие записям в репозитории, узел потоков EA выводит все соответствующие записи вместе с новой записью, к которой он добавляет поля ID и указания источника. Если никаких соответствий не найдено, узел обработки выводит только новую запись с добавленными полями ID и указания источника.

Чтобы проиллюстрировать это, предположим, что содержимое репозитория в настоящее время состоит из вывода узла источника Entity Analytics (EA). Смотрите раздел Табл. 6 на стр. 8.

Теперь мы получаем следующие новые записи. Относятся ли к кому-либо, кого мы уже знаем?

Таблица 7. Новые записи для оценки.

Имя	Адрес	Телефон	Кредит
Suzan	1234 5th Street	555-1234	\$100000
Mark	888 9th Ave	555-9999	\$60000

Сравнивая новые данные с существующим содержимым репозитория, узел потоков EA сопоставляет первую из новых записей с человеком с идентификатором 4 в существующих записях. Однако для второй новой записи никакого соответствия найти не удастся, таким образом, ей присваивается новый уникальный идентификатор 5.

Узел потоков EA добавляет поля идентификатора и указания источника и выводит новые записи вместе со всеми соответствующими им записями. Мы получим следующий вывод.

Таблица 8. Вывод узла потоков EA.

SEA-ID	SEA-SRC	Имя	Телефон	Адрес	Кредитный риск	Сбережения	Кредит
4	Заказчики	Susan		1234 5th St		\$1234	
4	Кредит	Sue	555-1234	1234 5th St			\$10000
4	Новые кредиты	Suzan	555-1234	1234 5th Street			\$100000
5	Новые кредиты	Mark	555-9999	888 9th Ave			\$60000

Этот вывод можно агрегировать, используя идентификатор Entity Analytics в качестве ключа агрегации, и передать другим узлам ниже по потоку для последующей обработки.

Узел потоков EA полностью описан ниже.

Этап 8: Генерирование предупреждений

И снова потенциально подозрительные действия могут стать очевидными. В данном случае человек с идентификатором 4 уже взял кредит на 10000 долларов и теперь, под немного иным именем, претендует на еще один, в десять раз больший. Конечно, это может быть приемлемым и не обязательно связано с мошенничеством. Однако если такие действия рассматриваются как подозрительные согласно вашим бизнес-правилам, стоит это отметить.

Вы могли, например, присоединить и запустить узел таблицы или узел отчета SPSS Modeler, напечатать содержимое его окна вывода, чтобы кто-нибудь прочитал его и составил предупреждения вручную. Но можно и передать вывод узла потоков EA в модель оценки рисков, которую вы ранее создали в IBM SPSS Modeler, и вычислить набор оценок, более строго отражающих ваши бизнес-правила. Еще одна возможность - экспортировать вывод в базу данных или на какой-либо другой носитель для последующей обработки. При использовании IBM SPSS Modeler у вас есть широкий выбор действий, соответствующих вашим конкретным требованиям.

Глава 3. Задачи Entity Analytics

Об этих задачах

В этом разделе описываются следующие задачи Entity Analytics.

- Создание репозитория объектов
- Конфигурирование репозитория объектов
- Анализ установленных соответствий
- Распознавание новых записей по репозиторию объектов
- Очистка репозитория объектов
- Удаление репозитория объектов
- Использование Entity Analytics с другими продуктами IBM SPSS
- Администрирование Entity Analytics

Конфигурирование репозитория объектов (узел Экспорт EA)

Процесс конфигурирования репозитория объектов состоит из следующих задач:

1. Соединение с источником данных. Дополнительную информацию смотрите в разделе “Соединение с источником данных” на стр. 14.
2. Создание репозитория. Дополнительную информацию смотрите в разделе “Создание репозитория” на стр. 14.
3. Отображение входных полей в источнике данных на характеристики в репозитории. Дополнительную информацию смотрите в разделе “Отображения входных полей на характеристики (узел Экспорт EA)” на стр. 16.

Сконфигурировав отображения, можно вывести их либо для текущего источника данных, либо для всех источников данных, известных в репозитории. Дополнительную информацию смотрите в разделе “Вывод отображений полей (узел Экспорт EA)” на стр. 18.

Примечание: Начиная с версии 16, SPSS Entity Analytics поддерживает репозитории в базах данных IBM DB2. Репозиторий специфичен для версии SPSS Modeler и не может быть импортирован из более ранней версии, поэтому если у вас есть существующий репозиторий и вы выполняете обновление до SPSS Entity Analytics версии 16, надо будет пересоздать этот репозиторий в новой базе данных DB2.

Репозиторий объектов

Репозиторий представляет собой центральное место хранения, выступая в качестве хранилища данных для всей информации об объектах. Поскольку репозиторий развивается, у него есть единственное состояние, поэтому у репозитория объектов нет понятия версии. Репозиторий содержит текущее состояние всех входных данных и может вырасти очень большим.

Вы можете поддерживать содержимое репозитория посредством простого в использовании графического интерфейса. Дополнительную информацию смотрите в разделе “Конфигурирование репозитория объектов” на стр. 18.

Важное замечание: Начиная с версии 16, SPSS Entity Analytics поддерживает репозитории в базах данных IBM DB2; предыдущие версии SPSS Entity Analytics поддерживали репозитории в IBM solidDB. Если у вас есть существующий репозиторий в SolidDB, при обновлении до SPSS Entity Analytics версии 16 или новее надо будет пересоздать этот репозиторий в новой базе данных DB2.

Соединение с источником данных

Вы начинаете работу с чтения данных источника в SPSS Modeler с помощью узла источника.

Чтобы соединиться с источником данных

1. На вкладке Источники на палитре узлов в нижней части главного окна SPSS Modeler дважды щёлкните по значку, соответствующему типу данных источника. Это действие добавляет узел источника на холст экрана.
2. На холсте экрана дважды щёлкните по значку, чтобы открыть его диалоговое окно.
3. В поле Файл введите положение и имя файла данных источника.
4. Заполните остальные поля диалогового окна, как требуется (выберите Справка, чтобы получить более полную информацию), затем нажмите кнопку ОК.
5. Если файл данных источника не содержит поля, которое однозначно идентифицирует каждую запись, добавьте его с помощью узла Производные. Дополнительную информацию смотрите в разделе “Добавление уникального идентификатора записи”.

Примечание: Данные с нелатинскими символами не поддерживаются. Если данные состоят из смеси латинских (как в Западной Европе, например) и нелатинских наборов символов, будут разрешаться только записи с латинскими символами.

Добавление уникального идентификатора записи

Каждый файл источника данных, который является входным для репозитория объектов, должен содержать одно поле, которое однозначно идентифицирует каждую запись. Если файл источника данных не содержит такого поля, его можно добавить с помощью узла Производные SPSS Modeler.

Чтобы добавить уникальный идентификатор записи к файлу источника данных

1. На холсте экрана щёлкните по узлу источника, который вы добавили в предыдущей задаче.
2. На вкладке **Опции поля** на палитре узлов дважды щёлкните по значку **Производные**, чтобы прикрепить узел Производные к узлу источника.
3. На холсте экрана дважды щёлкните по узлу Производные, чтобы открыть его диалоговое окно.
4. В поле **Производные** замените имя по умолчанию на значимое имя (например, **ID**) для добавляемого поля идентификатора.
5. Убедитесь, что в поле **Вывести как** задано **Формула**
6. Задайте для **Типа поля** значение Количественное.
7. В текстовом окне **Формула** введите @INDEX и нажмите кнопку **ОК**.

Создание репозитория

Вам надо создать репозиторий, чтобы хранить все входные данные.

Примечание: Если вы работаете в режиме удаленного сервера, надо создать репозиторий на компьютере сервера Modeler (то есть при создании репозитория вы должны быть соединены с сервером Modeler с клиента Modeler, чтобы репозиторий EA был создан на компьютере сервера).

Чтобы создать репозиторий

1. На вкладке Экспорт палитры узлов SPSS Modeler поместите узел Экспорт EA на холст потока.
Примечание: Если вы создаёте репозиторий в первый раз, используйте узел Экспорт EA и соедините его с узлом источника SPSS Modeler, содержащим данные, которые вы хотите ввести в репозиторий (или с узлом Производные, если вы добавляете узел, чтобы получить поле уникального идентификатора).
Чтобы соединить узлы, выполните следующие действия:
 - a. Щёлкните правой кнопкой мыши по узлу источника SPSS Modeler.
 - b. Выберите Соединить.
 - c. Щёлкните по узлу Экспорт EA.

2. Дважды щёлкните по узлу Экспорт EA, чтобы открыть его диалоговое окно.
3. Щелкните по списку **Репозиторий объектов**.
4. Нажмите кнопку **<Просмотр...>**, чтобы вывести диалоговое окно Репозитории объектов.
5. В диалоговом окне Репозитории объектов щёлкните по полю Имя репозитория.
6. Выберите **<Создать новый репозиторий...>**, чтобы вызвать мастер по созданию репозитория.

Мастер по созданию репозитория

Шаг 1

Здесь вы выбираете, создавать ли локальный репозиторий, используя продукт IBM DB2, поставляемый в комплекте с IBM SPSS Modeler Entity Analytics, или использовать для этого репозитория внешнюю базу данных.

Создать локальный репозиторий. Задайте имя пользователя и пароль администратора для базы данных IBM DB2, в которой будет находиться создаваемый репозиторий. Подтвердите пароль и нажмите кнопку **Далее**.

Примечание: В имени пользователя нельзя использовать ни дефис, ни символ подчеркивания.

Для задач администрирования репозитория на узлах Entity Analytics, таких как создание или уничтожение репозитория, требуются дополнительные разрешения. В UNIX пользователь, вошедший на IBM SPSS Modeler Server, должен быть членом группы db2iadm1. В Windows пользователь, вошедший на IBM SPSS Modeler Server, должен быть членом группы DB2ADMNS, чтобы выполнять задачи администрирования для репозитория.

Если в дальнейшем потребуется изменить регистрационные данные администратора, это можно сделать с помощью редактора командной строки для базы данных. Дополнительную информацию смотрите в разделе “Управление регистрационными данными администратора управления для базы данных репозитория” на стр. 33.

Примечание: Возможно только одно сочетание имени пользователя и пароля. Для входа в репозиторий все пользователи используют одно и то же имя пользователя и пароль.

Добавить внешний репозиторий. Выберите эту опцию, если вы хотите использовать для репозитория внешнюю базу данных. Введите положение файла .ini базы данных в поле **Выберите файл .ini репозитория** и нажмите кнопку **Далее**.

Шаг 2

Имя нового репозитория. Введите уникальное имя для нового репозитория.

Импортировать конфигурацию из. (Только локальный репозиторий) Если вы хотите, чтобы конфигурация основывалась на конфигурации существующего репозитория, выберите здесь репозиторий, в противном случае выберите **По умолчанию**. Дополнительную информацию смотрите в разделе “Конфигурирование репозитория объектов” на стр. 18.

При использовании существующего репозитория введите подробности соединения, если они отличаются от тех, что вы ввели на предыдущем экране.

Нажмите кнопку **ОК**, чтобы создать новый репозиторий и вывести диалоговое окно Экземпляры распознавания объектов, из которого вы можете установить соединение с репозиторием.

Опции репозитория объектов

Диалоговое окно Репозитории объектов содержит много опций для создания, соединения, конфигурирования и обслуживания репозитория объектов.

Соединиться с репозиторием. Используйте эти опции, чтобы создать новый репозиторий объектов или соединиться с существующим.

- **Имя репозитория.** Содержит текущий репозиторий объектов, если он есть. Чтобы выбрать другой репозиторий, если их несколько, выберите нужный из списка.

Для создания нового репозитория выберите **<Создать новый репозиторий...>**. При этом запускается мастер, который поможет выполнить процесс создания.

- **Имя пользователя.** Введите допустимое имя пользователя для выбранного репозитория.
- **Пароль.** Пароль для этого имени пользователя.
- **Соединить.** Нажмите эту кнопку для соединения с текущим репозиторием.

Управление репозиторием. В этой таблице приводится список источников данных, загруженных в текущий репозиторий (тот, с которым вы соединены), с количеством записей в каждом источнике данных.

- **Обновить.** Обновляет информацию об источнике данных и его размере в таблице, например, если вы добавили новый источник данных или изменили размер существующего.
- **Очистить все.** Удаляет все исходные данные из репозитория, но сохраняет все подробности конфигурации. Эту опцию можно использовать, если информация конфигурации все еще полезна, но вы хотите удалить все записи данных из репозитория. Дополнительную информацию смотрите в разделе “Очистка репозитория объектов” на стр. 35.
- **Удалить ненужные.** Удаляет выделенные исходные данные из репозитория, но сохраняет все подробности конфигурации. Дополнительную информацию смотрите в разделе “Удаление ненужных источников данных из репозитория” на стр. 35.
- **Переименовать источник.** Открывает диалоговое окно, в котором можно изменить имя выделенного источника данных.

Примечание: Эта операция переименовывает источник данных внутри репозитория; вам надо будет повторно выбрать этот источник данных под новым именем, если он используется в существующем узле экспорта или узле потока.

Уничтожить весь репозиторий. Полностью уничтожает содержимое текущего репозитория и подробности его конфигурации. Дополнительную информацию смотрите в разделе “Удаление репозитория объектов” на стр. 35.

Конфигурировать репозиторий. Выводит окно, где можно сконфигурировать текущий репозиторий. Дополнительную информацию смотрите в разделе “Конфигурирование репозитория объектов” на стр. 18.

Отображения входных полей на характеристики (узел Экспорт EA)

Репозиторий предоставляет набор стандартных предопределенных функций. Различные источники данных могут использовать различные имена полей (например, **Address1** или **Address Line 1**) для типов информации, которые соответствуют одной и той же характеристике. Чтобы избежать дублирования, надо отобразить входные поля источника данных на конкретные характеристики репозитория. Отображать нужно не каждое поле в наборе данных, а только те, которые, скорее всего, соответствуют той же самой характеристике в других наборах данных.

Если источник данных использует поля, соответствующие иным типам информации, помимо предопределенных в репозитории, можно создавать новые характеристики из окна конфигурации репозитория. Дополнительную информацию смотрите в разделе “Конфигурирование репозитория объектов” на стр. 18.

Чтобы отобразить входные поля на характеристики

1. Присоедините узел Экспорт EA Export к узлу источника данных на холсте экрана. Каждый из используемых вами узлов источников данных надо присоединить к своему собственному узлу Экспорт EA.

2. Откройте узел Экспорт EA, чтобы вывести вкладку Входные поля с опциями для отображения входных полей. Дополнительную информацию смотрите в разделе “Опции ввода в репозиторий для отображения”.
3. В узле Экспорт EA выберите вкладку Репозиторий, чтобы просмотреть назначения отображений либо для текущего источника данных, либо для всех источников данных, если вы используете несколько источников.
4. Чтобы сохранить набор назначений отображений (например, для использования с узлом экспорта для другого источника данных), нажмите кнопку **Экспорт отображения**.

Закончив отображение узла первого источника данных, повторите процесс для остальных узлов источников данных, которые вы хотите использовать.

Опции ввода в репозиторий для отображения

Вкладка Входные поля содержит опции для отображения полей источника данных на характеристики репозитория, готовые для экспорта в репозиторий. Задайте назначения отображений на этой вкладке, если хотите, щёлкните по вкладке Репозиторий, чтобы посмотреть отображение для других источников данных, затем нажмите кнопку **Выполнить** для экспорта данных в репозиторий.

Если вы уже сохранили набор отображений в файл XML, можно использовать их, нажав кнопку **Импортировать отображение**.

Режим. Оставьте выбор по умолчанию **Добавить в репозиторий**, если вы хотите добавить записи исходных файлов к существующему содержимому репозитория. Если вы хотите очистить содержимое репозитория, но сохранить информацию о конфигурации перед добавлением записей источника, выберите **Очистить репозиторий перед экспортом**.

Репозиторий объектов. Содержит текущий репозиторий объектов, если он есть. Чтобы выбрать другой репозиторий, если их несколько, выберите нужный из списка. Чтобы создать новый репозиторий, выберите **<Просмотр...>** для вывода окна, в котором можно создать репозиторий. Дополнительную информацию смотрите в разделе “Опции репозитория объектов” на стр. 15.

Отобразить на тип объектов. Список типов объектов (то есть наборов характеристик), определенных в репозитории. Выберите один тип объектов из списка, либо выберите **<Добавить новый тип объектов...>**, чтобы вывести окно конфигурации репозитория, где можно определить новый тип объектов. Дополнительную информацию смотрите в разделе “Конфигурирование репозитория объектов” на стр. 18.

Тег источника. Список тегов указывает источники данных, известных в настоящее время для репозитория. Выберите один из списка или выберите **<Добавить новый тег источника...>**, чтобы создать тег для нового источника данных.

Ключ уникальности. (обязательное) Поле ввода для использования идентификаторов уникальности для записей данных.

Таблица отображений. В этой таблице можно отобразить каждое входное поле на соответствующую характеристику в репозитории. Если в выбранном типе объектов нет подходящей характеристики, здесь можно создать новую характеристику.

- **Поле.** Набор входных полей в выбранном источнике данных. У каждого поля есть значок, обозначающий уровень измерения (то есть тип данных) для данного поля.
- **Отображено на характеристику.** Чтобы отобразить поле на характеристику, дважды щёлкните по этому столбцу (или нажмите клавишу пробела) на строке поля и выберите характеристику из списка. Если подходящая характеристика недоступна, выберите **<Добавить новую характеристику...>**, чтобы вывести окно конфигурации репозитория, где можно определить новую характеристику для этого типа объектов. Дополнительную информацию смотрите в разделе “Конфигурирование репозитория объектов” на стр. 18.
- **Использование.** Указывает контекст конкретного поля, где возможно несколько контекстов, например, домашний и рабочий номера телефонов. Есть предопределенные типы использования для характеристик

ADDRESS и PHONE, и вы можете создать ваши собственные типы использования для всех характеристик. Чтобы задать использование, отличное от использования по умолчанию (**Автоматически**), щелкните по этому столбцу на нужной строке и выберите либо один из существующих типов использования (если они есть), либо нажмите кнопку **<Добавить использование...>**, чтобы создать новое использование. Дополнительную информацию смотрите в разделе “Управление типами объектов” на стр. 22.

Импорт отображений. Импортирует ранее экспортированный набор отображений полей на характеристики из внешнего файла XML. Это может быть полезно, если у вас есть различные источники данных с одними и теми же требованиями к отображению, что позволяет избежать переопределения тех же самых отображений для различных источников.

Экспорт отображения. Экспортирует во внешний файл XML набор отображений полей на характеристики, указанный в таблице отображений.

Вывод отображений полей (узел Экспорт EA)

На вкладке Репозиторий нажмите кнопку **Обновить**, чтобы увидеть, у каких характеристик репозитория есть отображенные на них входные поля. Это можно посмотреть либо для текущего источника данных (управляемого узлом источника, прикрепленным к этому узлу экспорта), либо для всех источников данных.

Показать входные поля для. Выберите опцию вывода отображений либо для текущего источника данных, либо для всех источников данных, известных в репозитории.

Обновить. Обновляет вывод для выбранной входной опции.

Элементы. Список всех характеристик, у которых есть отображения на выведенные источники данных. Неотображенные характеристики не выводятся.

<Источник данных>. В каждом столбце перечисляются отображенные поля в конкретном источнике данных для каждой характеристики, для которой определено отображение.

Конфигурирование репозитория объектов

Вы можете управлять содержимым репозитория из окна Конфигурация репозитория, которое обеспечивает простой в использовании графический интерфейс для всего репозитория.

Если вы используете несколько репозитория с одинаковыми или похожими конфигурациями, можно задать базовую конфигурацию и экспортировать её в файл, который затем можно импортировать в другие репозитории. Дополнительную информацию смотрите в разделе “Повторное использование конфигурации репозитория” на стр. 25.

Примечание: Начиная с версии 16, SPSS Entity Analytics поддерживает репозитории в базах данных IBM DB2. Репозиторий специфичен для версии SPSS Modeler и не может быть импортирован из более ранней версии, поэтому если у вас есть существующий репозиторий и вы выполняете обновление до SPSS Entity Analytics версии 16, надо будет пересоздать этот репозиторий в новой базе данных DB2.

ОСТОРОЖНО:

Если изменяете и сохраняете конфигурацию репозитория, который уже содержит данные, вам могут предложить очистить содержимое репозитория и перезагрузить данные. Это позволит избежать ситуации, когда репозиторий останется в несогласованном состоянии.

Чтобы задать конфигурацию репозитория

1. Откройте любой узел Entity Analytics.
2. Щелкните по списку **Репозиторий объектов**.
3. Нажмите кнопку **<Просмотр...>**, чтобы вывести диалоговое окно Экземпляры распознавания объектов.
4. В диалоговом окне Экземпляры распознавания объектов щелкните по списку **Имя репозитория**.

5. Выберите репозиторий, для которого вы хотите задать конфигурацию.
6. Если вы ещё не соединены, введите имя администратора и пароль и нажмите кнопку **Соединить**.
7. Когда появится кнопка **Конфигурировать репозиторий**, нажмите её, чтобы вывести окно Конфигурация репозитория.
8. Создайте элементы конфигурации, как описано в следующих разделах.

Панель навигации в левой части окна Конфигурация репозитория содержит структуру дерева, откуда вы можете управлять различными характеристиками репозитория.

Таблица 9. Основные элементы окна Конфигурация репозитория.

Раздел	Описание	
Источники данных	Содержит отображения всех источников данных на различные характеристики репозитория.	Дополнительную информацию смотрите в разделе “Просмотр отображений источников данных”.
Показатели	Создать новую характеристику или же дублировать, изменить или удалить существующую.	Дополнительную информацию смотрите в разделе “Управление характеристиками репозитория”.
Типы объектов	Создать новый тип объекта или управлять существующими (дублировать, переименовывать, добавлять или удалять его характеристики, удалять).	Дополнительную информацию смотрите в разделе “Управление типами объектов” на стр. 22.
Правила распознавания	Задать порог для соответствия объектов.	Дополнительную информацию смотрите в разделе “Задание порога для соответствия объектов” на стр. 25.

Просмотр отображений источников данных

В разделе Источники данных окна Конфигурация репозитория запись Все источники выводит (только для чтения) отображения всех источников данных на различные характеристики репозитория.

Нажмите кнопку **Обновить**, чтобы обновить список, если новые источники данных уже добавлены в репозиторий.

Примечание: Здесь нельзя добавить источник данных в репозиторий. Источники данных могут быть добавлены только путём создания узла источника SPSS Modeler и соединением его с узлом Экспорт Entity Analytics. Дополнительную информацию смотрите в разделе “Соединение с источником данных” на стр. 14.

Управление характеристиками репозитория

Характеристика репозитория - это отдельный тип информации, который может быть использован вместе с источником данных объекта. Некоторые характеристики (например, Имя, Фамилия, Дата рождения и тому подобные) могут использоваться во многих источниках данных, тогда как другие характеристики могут относиться только к конкретному источнику данных. Характеристика может состоять из одной или нескольких частей; каждая часть характеристики обычно эквивалентна либо полю в записи данных, либо столбцу в таблице базы данных.

В разделе Характеристики окна Конфигурация репозитория запись Все характеристики дает средства для управления всеми характеристиками репозитория. Можно:

- Создать новую характеристику
- Дублировать существующую характеристику (например, создать новую характеристику на основе существующей)
- Изменить существующую характеристику
- Удалить существующую характеристику

Инструкции по этим задачам даны ниже в настоящем разделе.

Список характеристик содержит все характеристики, которые определены в данном репозитории. Столбцы в этом списке показывают различные свойства, которые какая-либо характеристика может иметь.

Характеристика. Имя характеристики. Значок в виде замка, следующий за именем характеристики, означает, что данная характеристика заблокирована. Заблокированные характеристики нельзя удалять и дублировать, и единственное изменение в них, которое можно сохранять - это изменение атрибута анонимизации.

Частота. Указывает, у скольких объектов может быть одно и то же значение данной характеристики. Допустимые значения - **One** (один, как у номера паспорта), **Few** (мало, как например, у адреса) или **Many** (много, как например, у даты рождения)

Исключительность. Указывает, что у объекта обычно будет только одна из характеристик данного типа. Например, для даты рождения или номера национального паспорта здесь было бы значение **Yes**, в то время как у адреса или номера кредитной карты - значение **No** (у человека может быть несколько адресов или несколько кредитных карт).

Стабильность. Указывает значение стабильности этой характеристики (то есть является ли ее изменение за время жизни объекта *маловероятным*). Например, у даты рождения здесь было бы значение **Yes**, поскольку она никогда не меняется, а у характеристики адреса - значение **No**, так как адрес за время жизни вполне может измениться, то есть он менее стабилен. *Примечание:* Обычно пол человека за время жизни не меняется, однако его часто неверно указывают в недостаточно надежных данных, поэтому в конфигурации по умолчанию для него задается значение **No**.

Анонимизировать. Указывает, что характеристика анонимизирована. Возможны значения **Да** и **Нет**. Дополнительную информацию смотрите в теме “Анонимизация характеристик репозитория” на стр. 22.

Чтобы создать новую характеристику

1. Выполните одно из следующих действий:
 - Нажмите кнопку Создать новую характеристику (верхняя кнопка в правой части экрана).
 - Щёлкните правой кнопкой мыши по **Все характеристики** на навигационной панели в левой части экрана и выберите **Новая характеристика**.
2. Заполните диалоговое окно Добавить/изменить характеристику. Дополнительную информацию смотрите в разделе “Добавление или изменение характеристики” на стр. 21.

Чтобы дублировать существующую характеристику

1. В столбце **Характеристика** таблицы в правой части экрана выберите характеристику, которую вы хотите дублировать.
2. Нажмите кнопку Дублировать выбранную характеристику (вторая кнопка в правой части экрана).
3. Заполните диалоговое окно Добавить/изменить характеристику. Дополнительную информацию смотрите в разделе “Добавление или изменение характеристики” на стр. 21.

Чтобы изменить существующую характеристику

ПРЕДОСТЕРЕЖЕНИЕ: Если вы редактируете, удаляете или анонимизируете характеристику или элемент характеристики, когда репозиторий уже содержит данные, позже надо выполнить очистку репозитория и перезагрузить данные. Это позволит избежать ситуации, когда репозиторий находится в несогласованном состоянии.

1. В столбце **Характеристика** таблицы в правой части экрана выберите характеристику, которую хотите изменить. *Примечание:* Изменять можно только те характеристики, которые вы сами создали, но не системные характеристики.
2. Нажмите кнопку Изменить выбранную характеристику (третья кнопка в правой части экрана).

3. Заполните диалоговое окно **Добавить/изменить характеристику**. Дополнительную информацию смотрите в разделе “Добавление или изменение характеристики”.

Чтобы удалить существующую характеристику

ПРЕДОСТЕРЕЖЕНИЕ: Если вы редактируете, удаляете или анонимизируете характеристику или элемент характеристики, когда репозиторий уже содержит данные, позже надо выполнить очистку репозитория и перезагрузить данные. Это позволит избежать ситуации, когда репозиторий находится в несогласованном состоянии.

1. В столбце **Характеристика** таблицы в правой части экрана выберите характеристику, которую хотите удалить. *Примечание:* Удалять можно только те характеристики, которые вы сами создали, но не системные характеристики.
2. Выполните одно из следующих действий:
 - Нажмите кнопку **Удалить** выбранную характеристику (нижняя кнопка в правой части экрана).
 - Щёлкните правой кнопкой мыши по **Все характеристики** на навигационной панели в левой части экрана и выберите **Удалить**.
3. Нажмите кнопку **Продолжить**, чтобы подтвердить удаление характеристики.

ОСТОРОЖНО:

Отменить удаление характеристики нельзя.

Добавление или изменение характеристики

ПРЕДОСТЕРЕЖЕНИЕ: Если вы редактируете, удаляете или анонимизируете характеристику или элемент характеристики, когда репозиторий уже содержит данные, позже надо выполнить очистку репозитория и перезагрузить данные. Это позволит избежать ситуации, когда репозиторий находится в несогласованном состоянии.

В диалоговом окне **Добавить/изменить характеристику** можно создать новую характеристику репозитория или же дублировать или изменить существующую.

Примечание: Если существующая характеристика заблокирована, ее подробности нельзя редактировать в этом диалоговом окне.

Тип характеристики. Метка, указывающая тип информации, с которым связана данная характеристика. Эта метка образует первую часть идентификатора характеристики.

Описание. Краткое текстовое описание типа характеристики, только для информации.

Частота. Указывает, у скольких объектов может быть одно и то же значение данной характеристики. Допустимые значения - **One** (один, как у номера паспорта), **Few** (мало, как например, у адреса) или **Many** (много, как например, у даты рождения)

Исключительность. Указывает, что у объекта обычно будет только одна из характеристик данного типа. Например, для даты рождения или номера национального паспорта здесь было бы значение **Yes**, в то время как у адреса или номера кредитной карты - значение **No** (у человека может быть несколько адресов или несколько кредитных карт).

Стабильность. Указывает значение стабильности этой характеристики (то есть является ли ее изменение за время жизни объекта *маловероятным*). Например, у даты рождения здесь было бы значение **Yes**, поскольку она никогда не меняется, а у характеристики адреса - значение **No**, так как адрес за время жизни вполне может измениться, то есть он менее стабилен. *Примечание:* Обычно пол человека за время жизни не меняется, однако его часто неверно указывают в недостаточно надежных данных, поэтому в конфигурации по умолчанию для него задается значение **No**.

Таблица элементов. Список элементов, из которых состоит данная характеристика.

- **Элемент.** Имя элемента.
- **Описание.** Краткое описание того, что содержит элемент.
- **Тип данных.** Тип данных, который можно использовать для данного элемента. Доступные типы - строка, целое, действительное число и дата.

Кнопка Добавить новый элемент. Добавляет новую строку в таблицу элементов, чтобы определить новый элемент.

Кнопка Удалить элемент. Удаляет выбранную строку из таблицы элементов. Отменить эту операцию нельзя.

ПРЕДОСТЕРЕЖЕНИЕ: Если вы редактируете, удаляете или анонимизируете характеристику или элемент характеристики, когда репозиторий уже содержит данные, позже надо выполнить очистку репозитория и перезагрузить данные. Это позволит избежать ситуации, когда репозиторий находится в несогласованном состоянии.

Анонимизировать. Для защиты данных вы можете выбрать анонимизацию данных, добавляемых в репозиторий; чтобы включить анонимизацию для характеристики, выберите **Да**. Дополнительную информацию смотрите в разделе “Анонимизация характеристик репозитория”.

Анонимизация характеристик репозитория

Как часть защиты данных, вы, возможно, захотите анонимизировать данные при добавлении в репозиторий, чтобы сократить риск ненамеренного раскрытия информации персональной идентификации.

Когда анонимизированные данные экспортируются в репозиторий, требуется такой метод анонимизации, который все же позволит выполнять распознавание объектов для анонимизированных данных. Например, если две записи данных для подробностей кредитных карт одного человека анонимизированы как "anon_s21" и "anon_s9271", их взаимосвязь теряется, однако с помощью внутренней, фоновой связи между этими записями система все же может понять, что одно имя представляет собой краткую форму другого.

Фоновые связи и идентификаторы, которые позволяет связывать анонимизированные данные, генерируются при создании репозитория; они уникальны для этого репозитория. Зашифрованные данные хранятся внутренним образом и затем читаются, когда поток соединяется с репозиторием.

Когда вы конфигурируете репозиторий, можно задать, будет ли анонимизирована каждая конкретная характеристика или нет. Если характеристика анонимизируется, анонимизируются все ее элементы, независимо от типов их использования. Дополнительную информацию смотрите в разделе “Добавление или изменение характеристики” на стр. 21.

Примечание: Не следует анонимизировать все поля для SPSS Entity Analytics, иначе вы не сможете идентифицировать данные при их возврате. Мы рекомендуем оставить хотя бы одно поле (даже если это просто поле номера строки) не анонимизированным, чтобы можно было управлять обратным слиянием данных с их исходными формами.

В столбце в списке возможностей в окне Конфигурация репозитория показано, для каких характеристик задана анонимизация. Возможны значения **Да** и **Нет**.

Примечание: Если существующий репозиторий содержал какие-либо данные до того, как выполняется анонимизация характеристики, сначала надо очистить репозиторий от данных, иначе сопоставление неанонимизированных и анонимизированных характеристик будет невозможно.

Управление типами объектов

Тип объекта - это именованный набор характеристик репозитория, логически связанных друг с другом. Например, тип объектов, предназначенный для работы с набором данных о клиенте, мог бы содержать такие характеристики, как Имя, Дата рождения, Пол, Адрес, Номер телефона и так далее.

Репозиторий IBM SPSS Modeler Entity Analytics поставляется со стандартным набором типов объектов, и вы можете добавить свои собственные типы.

В разделе Типы объектов окна Конфигурация репозитория перечисляются различные созданные типы объектов. Вы можете:

- Создать новый тип объектов
- Дублировать существующий тип объектов (например, чтобы создать новый тип объектов на основе уже существующего)
- Добавить характеристики к типу объектов
- Удалить характеристики из типа объектов
- Переименовать тип объектов
- Удалить тип объектов

Тип объектов. Имя выбранного типа объектов.

Характеристика. Список допустимых характеристик, из которых состоит данный тип объектов.

Тип использования. (Необязательно) Указывает различные контексты, в которых данная характеристика может быть использована. Щёлкните дважды по этому столбцу, чтобы добавить или изменить тип использования; разделяйте типы использования запятыми с пробелами. Значения, которые вы здесь задаёте, определяют значения, выводимые на узлах Экспорт EA или Поток EA, когда пользователь щёлкает по столбцу Использование для элемента на вкладке Входная информация. Дополнительную информацию смотрите в разделе “Опции ввода в репозиторий для отображения” на стр. 17.

Общая информация о типах использования:

- • Типы использования - это определенные метки.
- • Вы можете создать тип использования из практически любой текстовой записи, однако при этом нельзя вводить пробелы и недопустимые символы.
- • Все, что вы вводите, автоматически преобразуется в верхний регистр.
- • Типов использования может быть столько, сколько вам нужно.
- • Типы использования не обязаны называться содержательно, но если вы используете соглашение об именовании, осмысленном для вас и других пользователей, это поможет вам при последующем отображении.
- • Когда вы выполняете отображение, выводится предупреждение красным шрифтом, если вы отображаете некоторые элементы одного типа использования, а некоторые - другого.

Обычно, если вы пытаетесь отобразить два поля на один и тот же элемент характеристики, выводится ошибка. Типы использования - это способ отобразить два или несколько полей на один и тот же элемент характеристики с возможностью установить соответствия между ними.

Например, если вы определяете две отдельные характеристики - *HOMEADDRESS* (домашний адрес) и *WORKADDRESS* (рабочий адрес), между ними не должно быть соответствий. Если *HOMEADDRESS* у одного объекта совпадает с *WORKADDRESS* другого объекта, соответствие между ними не устанавливается, так как это разные характеристики. Однако если вы определяете одну характеристику с различными типами использования, при установлении соответствий можно учесть, что *ADDRESS.WORK* совпадает с *ADDRESS.HOME*.

Можно определить одни и те же типы использования для разных характеристик или задать для них разные типы, скажем *HM* и *WK* для телефонов и *HOME* и *WORK* для адресов. Это не имеет значения, поскольку мы не сопоставляем телефоны с адресами, однако если вы последовательны, это поможет идентифицировать и сгруппировать поля позже.

Когда несколько типов объектов попадают в один репозиторий, и вы используете одну и ту же характеристику, неважно, как вы назвали типы использования. Например, если вы определили *WK* и *HM* как типы использования *ADDRESS* для типа объектов *COMPANY*, это не мешает использовать *WORK* и *HOME* как типы использования *ADDRESS* для типа объектов *PERSON*.

Чтобы создать новый тип объектов

1. Щелкните правой кнопкой мыши по **Типы объектов** на панели навигации в левой части экрана.
2. Выберите **Новый тип объектов**.
3. Введите уникальное имя для типа объектов и нажмите кнопку ОК.
4. Добавьте характеристики к типу объектов (смотрите следующий раздел).

Чтобы добавить характеристики к типу объектов

1. Выберите тип объектов на навигационной панели в левой части экрана.
2. Нажмите кнопку **Добавить характеристику** (верхняя кнопка в правой части экрана).
3. Выберите из списка допустимых характеристик одну или несколько (используйте щелчок при нажатой клавише Ctrl, чтобы выбрать несколько характеристик) и затем нажмите кнопку ОК.

Чтобы удалить характеристики из типа объектов

1. Выберите тип объектов на навигационной панели в левой части экрана.
2. Выберите одну или несколько характеристик из таблицы добавленных характеристик в правой части экрана. Используйте щелчок при нажатой клавише Ctrl, чтобы выбрать несколько характеристик.
3. Нажмите кнопку **Отсоединить характеристику** (нижняя кнопка в правой части экрана).

Чтобы дублировать существующий тип объектов

1. На навигационной панели в левой части экрана щелкните правой кнопкой мыши по типу объектов, который вы хотите дублировать.
2. Выберите **Дублировать тип объектов**.
3. Введите уникальное имя для нового типа объектов, затем нажмите кнопку ОК.
4. Добавьте характеристики к типу объектов или удалите их из типа объектов, как вам требуется (смотрите инструкции выше).

Чтобы переименовать тип объектов

ПРЕДОСТЕРЕЖЕНИЕ: Если вы редактируете, удаляете или анонимизируете характеристику или элемент характеристики, когда репозиторий уже содержит данные, позже надо выполнить очистку репозитория и перезагрузить данные. Это позволит избежать ситуации, когда репозиторий находится в несогласованном состоянии.

1. На навигационной панели в левой части экрана щелкните правой кнопкой мыши по типу объектов, который вы хотите переименовать.
2. Выберите **Переименовать**.
3. Введите новое имя для типа объектов и нажмите кнопку ОК.

Чтобы удалить тип объектов

ПРЕДОСТЕРЕЖЕНИЕ: Если вы редактируете, удаляете или анонимизируете характеристику или элемент характеристики, когда репозиторий уже содержит данные, позже надо выполнить очистку репозитория и перезагрузить данные. Это позволит избежать ситуации, когда репозиторий находится в несогласованном состоянии.

1. На навигационной панели в левой части экрана щелкните правой кнопкой мыши по типу объектов, который вы хотите удалить.
2. Выберите **Удалить**.

3. Нажмите кнопку **ОК**, чтобы подтвердить удаление типа объектов.

ОСТОРОЖНО:

Отменить удаление типа объектов нельзя.

Задание порога для соответствия объектов

В разделе Правила распознавания окна Конфигурация репозитория вы выбираете порог, при котором устанавливается соответствие объектов.

Когда вы создаёте репозиторий, для соответствия задан порог по умолчанию.

Выберите **Задать агрессивное распознавание**, если в ваших записях не находится достаточно соответствий, чтобы выполнить распознавание объектов.

Выберите **Задать распознавание по умолчанию**, чтобы вернуться к порогу по умолчанию от одного из прочих вариантов.

Выберите **Задать консервативного распознавание**, если находится слишком много соответствий.

Чтобы построить репозиторий и для объектов, и для взаимосвязей, выберите **Включить взаимосвязи**.

Повторное использование конфигурации репозитория

Если вы уже задали конфигурацию и хотите использовать её для другого репозитория, можно экспортировать существующую конфигурацию в файл XML и импортировать полученный файл в другой репозиторий (репозиторий назначения). Это возможно только в пределах существующей установки. Например, нельзя перенести конфигурацию репозитория из одной версии IBM SPSS Modeler в другую или из базы данных одного типа в базу данных другого.

Чтобы повторно использовать существующую конфигурацию

1. Выведите окно Конфигурация репозитория для репозитория, конфигурацию которого вы хотите использовать. Дополнительную информацию смотрите в разделе “Конфигурирование репозитория объектов” на стр. 18.
2. Из меню в этом окне выберите **Конфигурация > Экспортировать конфигурацию**.
3. В диалоговом окне Сохранить как выберите имя и положение экспортируемого файла XML.
4. Выведите окно Конфигурация репозитория для репозитория назначения.
5. Из меню в этом окне выберите **Конфигурация > Импортировать конфигурацию**.
6. В диалоговом окне Открыть выберите имя и положение предыдущего экспортированного файла XML и нажмите кнопку **Открыть**.

Сохранение изменений конфигурации

Чтобы сохранить изменения конфигурации

В меню в окне Конфигурация репозитория выберите

Файл > Сохранить.

Заккрытие окна конфигурации

Чтобы выйти из окна конфигурации

В меню в окне Конфигурация репозитория выберите

Файл > Выход.

Если вы не сохранили изменения в конфигурации, нажмите кнопку **ОК**, чтобы сохранить изменения и выйти, или кнопку **Отмена**, чтобы выйти без сохранения изменений.

Анализ установленных соответствий (узел источника Entity Analytics (EA))

После того, как данные были экспортированы в репозиторий, можно использовать узел источника Entity Analytics (EA), чтобы передать установленные соответствия другим узлам IBM SPSS Modeler для дальнейшего анализа или обработки, такого как создание списка установленных соответствий.

Чтобы проанализировать установленные соответствия

1. Добавьте узел источника Entity Analytics (EA) к потоку.
2. Откройте узел Entity Analytics (EA).
3. На вкладке Данные выберите репозиторий объектов и один или несколько его входных источников данных (нажмите кнопку **Обновить**, чтобы обновить количество записей). Дополнительную информацию смотрите в разделе “Выбор источника данных”.
4. Добавьте ещё узлы к потоку, чтобы выполнить нужную вам обработку. Дополнительную информацию смотрите в разделе “Добавление узлов к потоку” на стр. 27.

Выбор источника данных

На вкладке Данные вы выбираете в репозитории по крайней мере один источник данных, которые собираетесь далее обрабатывать. Чтобы обновить количество записей в перечисленных источниках данных, нажмите кнопку **Обновить**.

Репозиторий объектов. Содержит текущий репозиторий объекта, если он есть. Чтобы выбрать другой репозиторий, если их несколько, выберите нужный из списка. Чтобы создать новый репозиторий, выберите **<Просмотр...>**, чтобы вывести окно, в котором можно создать репозиторий. Дополнительную информацию смотрите в разделе “Опции репозитория объектов” на стр. 15.

Включать записи из источников данных. В этой таблице перечисляются различные входные источники данных для репозитория вместе с числом записей в каждом источнике. Включите переключатель **Включить** для тех источников данных, которые вы хотите использовать для выполнения дальнейшего анализа или обработки. Чтобы выбрать или отменить выбор всех источников данных, нажмите кнопку **Включить все** или **Исключить все** соответственно.

Взаимосвязи. Выберите тип взаимосвязей для включения в репозиторий. Доступно только, если репозиторий сконфигурирован для включения взаимосвязей.

- **Без взаимосвязей.** Подробности взаимосвязей не используются.
- **Тесные взаимосвязи.** Выбирает только тесно связанные объекты. Близость взаимосвязей зависит от многих переменных, таких как свойства отображаемых характеристик, совместного использования характеристик и способа разрешения идентичности - консервативного или агрессивного.
- **Все взаимосвязи.** Выбирает все связанные объекты.

Максимальная степень разделения. Доступно только, если выбраны **Тесные взаимосвязи** или **Все взаимосвязи**. Выбрать количество степеней разделения, которые будут использованы в идентификации взаимосвязи. Например, если Анна и Боб не знают друг друга, но Джон знает их обоих, то Анна и Боб связаны двумя степенями разделения.

Вывод типа объектов. По умолчанию, если репозиторий содержит данные, здесь выводится тип первого объекта в списке объектов репозитория. Если в репозитории несколько типов объектов, при выборе здесь

типа характеристики, показанные на вкладке Фильтр, меняются на характеристики этого типа. Вы можете выбрать любой из типов объектов, используемых в репозитории.

Переименование полей данных

При помощи вкладки Фильтр можно переименовать любое из полей установленных соответствий Entity, которые передаются вниз по потоку для дальнейшей обработки. Вы можете захотеть переименовать поле установленного соответствия, например, чтобы поддерживать совместимость имен полей при слиянии нисходящего потока с другим набором данных.

Поля вместе со своими исходными именами приводятся ниже:

Таблица 10. Поля установленного соответствия

Поле	Описание
\$EA-ID	Идентификатор объекта
\$EA-SRC	Тег источника задаёт источник данных, откуда происходят записи
\$EA-KEY	Поле, обозначенное как ключ уникальности в файле источника данных

Примечание: Вкладку Фильтр можно использовать, чтобы отфильтровать (удалить) поля, но поступать здесь таким образом не стоит, так как поля установленного соответствия - это лишь необходимый минимум для обработки Entity Analytics.

Задание информации о типе для полей данных

На вкладке Типы можно посмотреть или изменить различные свойства полей установленных соответствий, которые передаются вниз по потоку для дальнейшей обработки.

Свойства, которые вы можете изменить, - те же самые, что и на вкладке Типы обычного узла Тип SPSS Modeler; они приведены ниже:

Таблица 11. Свойства типа для полей

Свойство	Описание
Измерение	Тип измерения (то есть тип данных), используется для описания характеристик данных в поле.
Значения	Задаёт опции для чтения значений данных из набора данных.
Пропущенные	Используется, чтобы задать способ обработки пропущенных значений для поля.
Проверка	Опции проверки для обеспечения соответствия значений полей заданным значениям или диапазонам.
Роль	Определяет, как будет использоваться поле, если данные передаются в узел моделирования или в слепок модели.

Добавление узлов к потоку

Вы можете добавлять различные узлы SPSS Modeler к потоку, чтобы провести анализ или операции обработки выходных данных из узла источника Entity Analytics (EA). Например, можно добавить один или несколько следующих узлов:

- Узел агрегации или особый узел, чтобы подготовить сводку выходных данных, объём которых может быть очень велик
- Узел выбора, чтобы выбрать подмножество выходных данных
- Узел таблицы, чтобы просмотреть выходные данные из узла источника Entity Analytics (EA)
- Узел отчета, чтобы вывести выходные данные в виде отчёта

- Узел экспорта SPSS Modeler, чтобы экспортировать выходные данные в ином формате, например, в формате электронной таблицы или базы данных

Дополнительную информацию смотрите в разделах об операциях записи, узлах вывода и экспорта в публикации *Руководство по узлам источников, обработки и вывода IBM SPSS Modeler*.

Сравнение новых наблюдений с репозиторием (узел потоков EA)

Когда вы уже выполнили распознавание некоторых объектов в репозитории, можно использовать узел потоков EA, чтобы сравнить новые записи, полученные позже, с содержимым репозитория. Этот узел обрабатывает записи из нового источника данных, сравнивает их с уже распознанными объектами в репозитории и передаёт все соответствующие записи для дальнейшей обработки. Можно задать установление точных соответствий или более свободных связей с существующими объектами.

Как и узел Экспорт EA, узел потоков EA принимает в качестве входного один узел источника SPSS Modeler. Однако у узла потоков EA есть такое отличие: В то время как узел Экспорт выводит записи для всех объектов, связанных с его входными записями, узел потоков EA выводит записи только для объектов, связанных с объектами, которые уже опознаны в репозитории. Дополнительную информацию смотрите в разделе “Вывод из узла потоков EA” на стр. 31.

Чтобы сравнить новые наблюдения с репозиторием

1. Соединитесь с источником данных, содержащим новые записи, которые вы хотите сравнить с существующими объектами. Дополнительную информацию смотрите в разделе “Соединение с источником данных” на стр. 14.
2. Из вкладки Операции записей присоедините узел потоков EA к узлу источника данных.
3. Дважды щёлкните по узлу Экспорт Entity Analytics, чтобы открыть диалоговое окно.
4. Щёлкните по списку **Репозиторий объектов**.
5. Нажмите кнопку **<Просмотр...>**, чтобы вывести диалоговое окно Репозитории объектов.
6. В диалоговом окне Репозитории объектов щёлкните по полю Имя репозитория.
7. Щёлкните по имени репозитория, который вы хотите использовать.
8. Введите имя пользователя и пароль для этого репозитория и нажмите кнопку **Соединить**. Когда соединение с репозиторием будет установлено, нажмите кнопку **ОК**.
9. В диалоговом окне Потоки EA выберите тип объектов, на который вы хотите выполнить отображение. Дополнительную информацию смотрите в разделе “Управление типами объектов” на стр. 22.
10. Отобразите поля ввода в источнике данных на характеристики в репозитории. Дополнительную информацию смотрите в разделе “Отображение полей ввода на характеристики (узел потоков EA)” на стр. 29.
11. Если хотите, можно изменять записи в репозитории в реальном времени при выполнении скоринга данных. Дополнительную информацию смотрите в разделе “Отображение полей ввода на характеристики (узел потоков EA)” на стр. 29.
12. Щёлкните по вкладке **Вывод**, чтобы увидеть подробности различных источников данных, введенных в репозиторий, и задайте критерии выбора для получения существующих объектов. Дополнительную информацию смотрите в разделе “Вывод отображений полей и источников данных (узел потоков EA)” на стр. 30.
13. Щёлкните по вкладке **Фильтр**, чтобы посмотреть подробности входных полей и характеристик, хранящихся в репозитории. Любые характеристики, не отображенные на узле, по умолчанию отфильтровываются, но если требуется, это можно изменить.
14. Когда узел будет сконфигурирован правильно, нажмите кнопку **ОК**.
15. Присоедините узел Таблица к узлу потоков EA и запустите поток.

В окне вывода узла Таблица появится список всех извлечённых объектов, которые соответствуют новым записям в источнике данных. К именам полей вывода добавляется префикс \$EA-. Дополнительную информацию смотрите в разделе “Вывод из узла потоков EA” на стр. 31.

Примечание: При выполнении узла потоков EA вы можете столкнуться с ошибкой вида **В модели данных сервера сервера обнаружено неправильное число полей**. Подобное может случиться, если вы изменили конфигурацию репозитория с момента создания узла потоков EA. Изменение конфигурации в подобных обстоятельствах может привести к изменению числа и имён полей, которые выводятся из узла. Чтобы решить эту проблему, откройте узел потоков EA и нажмите кнопку **Обновить**. При этом количество и имена полей вывода будут пересчитаны.

Отображение полей ввода на характеристики (узел потоков EA)

Вкладка Входные поля содержит опции для отображения входных полей для данного узла на характеристики репозитория. Задайте назначения отображений на этой вкладке или выберите вкладку **Просмотр**, чтобы просмотреть подробности всех источников данных в репозитории, затем нажмите кнопку **ОК**.

Если вы уже сохранили набор отображений в файл XML, можно использовать их, нажав кнопку **Импортировать отображение**.

Репозиторий объектов. Содержит текущий репозиторий объектов, если он есть. Чтобы выбрать другой репозиторий, если их несколько, выберите нужный из списка. Чтобы создать новый репозиторий, выберите **<Просмотр...>** для вывода окна, в котором можно создать репозиторий. Дополнительную информацию смотрите в разделе “Опции репозитория объектов” на стр. 15.

Отобразить на тип объектов. Список типов объектов (то есть наборов характеристик), определенных в репозитории. Выберите один тип объектов из списка, либо выберите **<Добавить новый тип объектов...>**, чтобы вывести окно конфигурации репозитория, где можно определить новый тип объектов. Дополнительную информацию смотрите в разделе “Конфигурирование репозитория объектов” на стр. 18.

Сохранение поисков. Если вы хотите изменять записи в репозитории в реальном времени при выполнении скоринга данных, выберите эту опцию.

Тег источника. Доступно, только если вы выбрали **Сохранение поисков**. Список тегов указывает источники данных, известных в настоящее время для репозитория. Выберите один из списка или выберите **<Добавить новый тег источника...>**, чтобы создать тег для нового источника данных.

Ключ уникальности. Доступно, только если вы выбрали **Сохранение поисков**. Поле ввода для использования идентификаторов уникальности для записей данных.

Таблица отображений. В этой таблице можно отобразить каждое входное поле на соответствующую характеристику в репозитории. Если в выбранном типе объектов нет подходящей характеристики, здесь можно создать новую характеристику.

- **Поле.** Набор входных полей в выбранном источнике данных. У каждого поля есть значок, обозначающий уровень измерения (то есть тип данных) для данного поля.
- **Отображено на возможность .** Чтобы отобразить поле на характеристику, дважды щёлкните по этому столбцу (или нажмите клавишу пробела) на строке поля и выберите характеристику из списка. Если подходящая характеристика недоступна, выберите **<Добавить новую характеристику...>**, чтобы вывести окно конфигурации репозитория, где можно определить новую характеристику для этого типа объектов. Дополнительную информацию смотрите в разделе “Конфигурирование репозитория объектов” на стр. 18.
- **Использование.** Указывает контекст конкретного поля, где возможно несколько контекстов, например, домашний и рабочий номера телефонов. Дополнительную информацию смотрите в разделе “Управление типами объектов” на стр. 22.

Импорт отображений. Импортирует ранее экспортированный набор отображений полей на характеристики из внешнего файла XML. Это может быть полезно, если у вас есть различные источники данных с одними и теми же требованиями к отображению, что позволяет избежать переопределения тех же самых отображений для различных источников.

Экспорт отображения. Экспортирует во внешний файл XML набор отображений полей на характеристики, указанный в таблице отображений.

Вывод отображений полей и источников данных (узел потоков EA)

На вкладке Вывод вы можете просмотреть подробности различных источников данных, которые были введены в репозиторий. Это источники данных, для которых обрабатывается входная информация этого узла с целью поиска и получения соответствующих объектов узла. Нажмите кнопку **Обновить**, чтобы обновить количество записей.

Включать соответствия из источников данных. В этой таблице перечисляются различные источники данных, доступные в репозитории, вместе с числом записей в каждом источнике.

Соответствия. Эти опции задают, насколько точно информация отображения полей на характеристики, которую вы задаёте на вкладке Входные поля, должна соответствовать записям-кандидатам (то есть всему содержимому репозитория). Чем строже критерий соответствия, тем меньше объектов будет получено.

Примечание: Если найдено более 20 соответствий, то возвращаются только первые 20.

- **Включать только точные соответствия.** Это наиболее строгий критерий соответствия, и он приводит к наименьшему числу отбираемых записей. Используйте эту опцию, когда вы хотите, чтобы возвращались только объекты с точными соответствиями.
- **Включить возможные соответствия.** Используйте этот параметр, когда вы хотите вернуть, как совпадающие объекты, так и объекты, у которых одинаковые идентификаторы (идентификаторы с характеристиками, для которых было задано значение частоты Один, например, совпадения номеров кредитных карт, идентификационных номеров налогоплательщика, и так далее).
- **Включить все соответствия.** Используйте эту опцию, когда хотите просмотреть наибольшее возможное число объектов в репозитории с общими характеристиками. Это самый широкий критерий соответствия, он приводит к наибольшему числу отобранных записей. Эта опция возвращает точные соответствия и объекты, у которых совпадают почти все характеристики (как правило, характеристики со значениями частот Один или Несколько). Например, здесь были бы отобраны оба объекта с одним и тем же идентификационным номером налогоплательщика, а также объекты с одинаковыми адресами.

Взаимосвязи. Доступно только, если репозиторий сконфигурирован для включения взаимосвязей. Выберите тип взаимосвязей, который будет включен в вывод.

- **Без взаимосвязей.** Подробности взаимосвязей не используются.
- **Тесные взаимосвязи.** Выбирает только тесно связанные объекты. Близость взаимосвязей зависит от многих переменных, таких как свойства отображаемых характеристик, совместного использования характеристик и способа разрешения идентичности - консервативного или агрессивного.
- **Все взаимосвязи.** Выбирает все связанные объекты.

Максимальный степень разделения. Доступно только, если выбраны **Тесные взаимосвязи** или **Все взаимосвязи**. Выбрать количество степеней разделения, которые будут использованы в идентификации взаимосвязи. Например, если Анна и Боб не знают друг друга, но Джон знает их обоих, то Анна и Боб связаны двумя степенями разделения.

Вывод типа объектов. По умолчанию, если репозиторий содержит данные, здесь выводится тип первого объекта в списке объектов репозитория. Если в репозитории несколько типов объектов, при выборе здесь типа характеристики, показанные на вкладке Фильтр, меняются на характеристики этого типа. Вы можете выбрать любой из типов объектов, используемых в репозитории.

Вывод из узла потоков EA

Вывод из узла потоков EA состоит из следующих полей для каждой извлекаемой записи.

Поле	Описание
<i>Поле1[, Поле2[, ... ПолеN]]</i>	Поля из источника данных, которые содержат новые записи.
\$EA-ID	идентификатор объекта для этой записи в данном репозитории.
\$EA-SRC	Тег источника, идентифицирующий источник данных происхождения этой записи.
\$EA-KEY	Значение ключа уникальности для этой записи в файле источника данных.
\$EA-SC	Поле, указывающее близость соответствия между этой записью и наблюдаемым объектом в данном репозитории; значение от 1,0 (слабое соответствие) до 10,0 (хорошее соответствие).
<i>\$EA-Характеристика1[, \$EA-Характеристика2[, ... \$EA-ХарактеристикаN]]</i>	Значения отображённых характеристик для этой записи в данном репозитории.

Если поля взаимосвязей включены в репозиторий, и степень разделения на вкладке Вывод больше нуля, вывод из узла потоков EA содержит также следующие поля для каждой извлекаемой записи.

Поле	Описание
\$EA-DEGREE	Степень разделения.
\$EA-PARENT	Идентификатор записи, от которой рассчитывается разделение.
\$EA-CHILD	Идентификатор записи, до которой рассчитывается разделение.
\$EA-RULE	

Использование IBM SPSS Modeler Entity Analytics с другими продуктами IBM SPSS

Программы установки могут позволить вам использовать IBM SPSS Modeler Entity Analytics со следующими продуктами:

- IBM SPSS Collaboration and Deployment Services
- IBM SPSS Modeler Batch для Windows
- IBM SPSS Modeler Solution Publisher

Эти программы установки надо запустить, прежде чем можно будет использовать возможности IBM SPSS Modeler Entity Analytics с данными продуктами. Более подробную информацию смотрите в руководстве *IBM SPSS Modeler Premium Installation*.

После установки надо при помощи клиента IBM SPSS Deployment Manager создать определение сервера репозитория Entity Analytics. Это требуется для использования потока IBM SPSS Modeler, который содержит узел Entity Analytics в задании IBM SPSS Collaboration and Deployment Services (другими словами, запускает потоки Entity Analytics в IBM SPSS Collaboration and Deployment Services). Определение сервера должно соответствовать имени репозитория в потоке; это определение служит для того, чтобы сообщить потоку, где найти репозиторий, и дать ему необходимую информацию о соединении.

Административные задачи

Для тех репозиториев, что созданы в Entity Analytics, создаётся новая служба баз данных, использующая продукт IBM DB2. Есть несколько административных задач, связанных с DB2; эти задачи обычно выполняет либо администратор базы данных, либо администратор системы, вот они:

- Конфигурирование назначений портов
- Управление регистрационными данными администратора управления для базы данных репозитория

Другие административные задачи, которые могут понадобиться, применяются ко всем репозиториям:

- Перемещение репозитория в другой каталог хранения
- Задание свойств потока для полей даты/времени и отметки времени
- Настройка параметров срока ожидания
- Выполнение IBM SPSS Modeler Entity Analytics с клиентом SPSS Modeler и сервером SPSS Modeler в той же самой системе Windows
- Очистка репозитория объектов
- Удаление репозитория объектов
- Удаление репозитория, если к нему не удастся подключиться

Конфигурирование назначений портов

Каждой службе баз данных DB2 надо выделить порт; этот порт нельзя выделять после этого ни одной из служб, выполняемых на данном компьютере. Службы баз данных находятся на том же самом компьютере, что и IBM SPSS Modeler Server (или, когда IBM SPSS Modeler используется без соединения с сервером IBM SPSS Modeler, на компьютере, который запускает IBM SPSS Modeler).

По умолчанию Entity Analytics назначает порты в диапазоне от 1320 до 1520, начиная с порта 1320 для первого созданного репозитория. В случае конфликта вы можете сконфигурировать назначение портов, изменив файл: *<путь установки сервера Modeler>/ext/bin/pasw.entityanalytics/ea.cfg* и задав соответствующие значения для параметров *min_port* и *max_port*. Содержимое этого файла по умолчанию показано ниже:

```
# конфигурация диапазона портов для Entity Analytics
#
#     этот диапазон портов определяет, какие порты могут использовать базы данных DB2
#     (созданные для хранения репозиториев Entity Analytics).
# Сконфигурируйте его, если диапазон портов по умолчанию
#     приводит к конфликту в вашей системе.
#
# default min_port = 1320
# default max_port = 1520
min_port, 1320
max_port, 1520
```

Управление регистрационными данными администратора управления для базы данных репозитория

Имя пользователя и пароль администратора для базы данных DB2, на которой находится репозиторий объектов, определяются при создании репозитория. Если вы знаете текущие регистрационные данные, можно изменить их с помощью редактора SQL DB2.

Чтобы запустить редактор SQL DB2

1. На компьютере клиента откройте окно командной строки.
2. Введите:

```
cd каталог_установки_Modeler\ext\bin\pasw.entityanalytics\DB2\bin
```

где *каталог_установки_Modeler* - это каталог, где установлен SPSS Modeler.
3. Введите:

```
sqlsql -c "C:\Documents and Settings\All Users\Application Data\IBM\SPSS\Modeler\версия\EA\repositories\имя_репозитория"
```

где *версия* - это номер версии установки SPSS Modeler, а *имя_репозитория* - это имя репозитория.
4. В ответ на приглашение введите имя и пароль администратора текущей базы данных, чтобы система вывела приглашение sqlsql>.

Чтобы изменить пароль администратора базы данных

1. В ответ на приглашение sqlsql> введите:

```
alter user имя_пользователя identified by пароль;  
commit work;
```

где *имя_пользователя* - это текущее имя администратора базы данных, а *пароль* - новый пароль.
2. Введите exit;, чтобы закрыть редактор.
3. Перезапустите клиент SPSS Modeler.

Информацию о других административных задачах для базы данных DB2 смотрите в документации для соответствующей версии IBM DB2 на сайте <http://publib.boulder.ibm.com/>.

Перемещение репозитория в другой каталог хранения

По умолчанию файлы репозитория хранятся в каталоге *EA*, который расположен:

- C:\Documents and Settings\All Users\ApplicationData\IBM\SPSS\Modeler*version*\EA (в системах Windows)
- *каталог_установки_Modeler*/ext/bin/pasw.entityanalytics/EA (в системах UNIX)

Поскольку файлы, хранящиеся в репозитории, могут становиться очень большими, вам может понадобиться переместить их на другой диск или в другой раздел, чтобы освободить место.

Чтобы переместить репозиторий в другой каталог

1. Выйдите из SPSS Modeler.
2. Переместите каталог *EA* из его исходного положения (которое указано выше) в новое положение. Например, в Windows вы могли бы переместить его в каталог *F:\data\EA*.
3. Измените файл *<путь_установки_сервера_Modeler>/ext/bin/pasw.entityanalytics/ea.cfg*, добавив в него следующую опцию:

```
repository_data_directory, новое_положение
```

где *новое_положение* - это каталог, куда вы перемещаете каталог *EA*, например, *F:\data\EA*.

Задание свойств потока для полей даты/времени и отметки времени

Если ваши исходные данные включают поля, содержащие данные даты/время или отметки времени, убедитесь, что соответствующие свойства потока задают формат, который распознаётся IBM SPSS Modeler Entity Analytics.

Чтобы задать формат в свойствах потока

1. В главном меню SPSS Modeler выберите:
Инструменты > Свойства потока > Опции.
2. Выберите **Дата/Время.**
3. Задайте для **Формата даты** значение **YYYY-MM-DD.**
4. Задайте для **Формата времени** значение **HH:MM:SS.**
5. Щелкните по **ОК.**

Настройка параметров срока ожидания

В медленных или перегруженных системах, если вы сталкиваетесь с ошибками при создании репозитория или при доступе к ним, может понадобиться увеличить значения сроков ожидания запуска и остановки механизма Entity Analytics или сервера баз данных Entity Analytics.

Чтобы настроить срок ожидания для механизма Entity Analytics

1. Выйдите из SPSS Modeler.
2. Измените файл *<путь установки сервера Modeler>/ext/bin/pasw.entityanalytics/ea.cfg*, увеличив значение для следующего параметра:

`timeout`, *значение*

где *значение* - это срок ожидания в секундах для механизма Analytics Engine (по умолчанию 60).

Чтобы настроить срок ожидания для сервера баз данных Entity Analytics (только для DB2)

1. Выйдите из SPSS Modeler.
2. Измените файл *<путь установки сервера Modeler>/ext/bin/pasw.entityanalytics/ea.cfg*, увеличив значение для следующего параметра:

`timeout`, *значение*

где *значение* - это срок ожидания в секундах для сервера баз данных DB2 Entity Analytics (по умолчанию 100).

Выполнение IBM SPSS Modeler Entity Analytics с клиентом SPSS Modeler и сервером SPSS Modeler в той же самой системе Windows

Если вы установили IBM SPSS Modeler Entity Analytics и на клиент SPSS Modeler, и на SPSS Modeler Server в той же системе Windows, по умолчанию клиент и сервер будут совместно использовать один и тот же репозиторий. Если вы хотите, чтобы они использовали разные репозитории, надо заменить файл конфигурации *ea.cfg* на **одной** из систем, чтобы она использовала другой диапазон портов и другую папку репозитория.

Примечание: Например, эту процедуру надо проделать, если вы используете 32-битный клиент SPSS Modeler и 64-битный клиент сервера SPSS Modeler (или наоборот).

1. Откройте файл *<путь установки [сервера] Modeler>/ext/bin/pasw.entityanalytics/ea.cfg* для редактирования.
2. Измените параметры `min_port` и `max_port`, чтобы использовать из другой системы другие порты.
Дополнительную информацию смотрите в разделе “Конфигурирование назначений портов” на стр. 32.
3. Измените параметр `repository_data_directory`, чтобы использовать из другой системы другой каталог.
4. Сохраните и закройте файл *ea.cfg*.

Очистка репозитория объектов

Если вы хотите очистить все содержимое репозитория от записей данных, но сохранить информацию о конфигурации, можно очистить репозиторий от данных.

Чтобы вычистить все данные из репозитория:

1. Откройте узел Entity Analytics.
2. Щелкните по списку **Репозиторий объектов**.
3. Нажмите кнопку **<Просмотр...>**, чтобы вывести диалоговое окно Экземпляры распознавания объектов.
4. В диалоговом окне Экземпляры распознавания объектов щелкните по списку **Имя репозитория**.
5. Выберите репозиторий, который вы хотите очистить.
6. Если вы ещё не соединены, введите имя администратора и пароль и нажмите кнопку **Соединить**.
7. Когда появится кнопка **Очистить все**, нажмите её.
8. В диалоговом окне Очистить источники данных нажмите кнопку **Очистить**, чтобы подтвердить очистку репозитория.

Удаление ненужных источников данных из репозитория

Если у вас есть источник данных, который больше не используется или не нужен в репозитории объектов, можно удалить этот источник данных из репозитория. Можно выбрать один или несколько источников данных для удаления.

Чтобы удалить выбранный источник данных из репозитория:

1. Откройте узел Entity Analytics.
2. Щелкните по списку **Репозиторий объектов**.
3. Нажмите кнопку **<Просмотр...>**, чтобы вывести диалоговое окно Экземпляры распознавания объектов.
4. В диалоговом окне Экземпляры распознавания объектов щелкните по списку **Имя репозитория**.
5. Выберите репозиторий, из которого вы хотите удалить источник данных.
6. Если вы ещё не соединены, введите имя администратора и пароль и нажмите кнопку **Соединить**.
7. В списке **Управление репозиторием** выберите источник данных для удаления. Если требуется, щелкните при нажатой клавише Ctrl по другим источникам данных, чтобы выбрать их для удаления.
8. Когда появится кнопка **Удалить ненужные**, нажмите её.
9. В диалоговом окне Удалить ненужные источники данных нажмите кнопку **Удалить**, чтобы подтвердить очистку репозитория.

Удаление репозитория объектов

Когда репозиторий больше не нужен, его можно удалить полностью.

ОСТОРОЖНО:

Эта операция выполняется, как сказано. Отменить эту операцию нельзя.

Если вы не уверены, используйте кнопку **Очистка**, чтобы удалить все данные источников. При этом конфигурация репозитория не удаляется. Дополнительную информацию смотрите в разделе “Очистка репозитория объектов”.

Примечание: Следующая процедура предполагает, что вы можете соединиться с репозиторием из SPSS Modeler и знаете имя пользователя и пароль администратора для базы данных, где расположен репозиторий. Если это не так и вы не можете подключиться к репозиторию, применяйте процедуру удаления. Дополнительную информацию смотрите в разделе “Удаление репозитория, если к нему не удается подключиться” на стр. 36.

Чтобы удалить репозиторий

1. Откройте узел Entity Analytics.
2. Щелкните по списку **Репозиторий объектов**.
3. Нажмите кнопку **<Просмотр...>**, чтобы вывести диалоговое окно Экземпляры распознавания объектов.
4. В диалоговом окне Экземпляры распознавания объектов щелкните по списку **Имя репозитория**.
5. Выберите репозиторий, который вы хотите удалить.
6. Если вы ещё не соединены, введите имя администратора и пароль и нажмите кнопку **Соединить**.
7. Когда появится кнопка **Удалить весь репозиторий**, нажмите её.
8. Нажмите кнопку **Удалить**, чтобы подтвердить удаление репозитория.
9. Нажмите кнопку **ОК**, чтобы подтвердить успешное удаление.

Удаление репозитория, если к нему не удается подключиться

Используйте следующую процедуру, если вы хотите удалить репозиторий объектов, но не можете подключиться к нему, либо из-за проблем соединения с SPSS Modeler, либо потому, что вы забыли имя пользователя или пароль.

Выполняйте эту процедуру на компьютере, где находится база данных репозитория.

Системы Windows

1. Откройте окно Командная строка.
2. Введите:

```
cd каталог_установки_modeler  
cd ext\bin\pasw.entityanalytics  
delete_repository.bat имя_репозитория
```

где *каталог_установки_modeler* - это каталог, где установлен SPSS Modeler, а *имя_репозитория* - имя репозитория.

Примечание: Имя репозитория регистрозависимо.

3. Перейдите к "Завершение процедуры" далее в этом разделе.

Системы UNIX

1. Откройте оболочку.
2. Введите:

```
cd каталог_установки_сервера_modeler  
cd ext/bin/pasw.entityanalytics  
./delete_repository.sh имя_репозитория
```

где *каталог_установки_сервера_modeler* - это каталог, где установлен SPSS Modeler Server, а *имя_репозитория* - имя репозитория.

Примечание: Имя репозитория регистрозависимо.

Завершение процедуры (все системы)

1. В ответ на подсказку подтвердите удаление репозитория, введя Y.
2. Удалите каталог, имя которого совпадает с именем удаленного репозитория. Если этот каталог не удается удалить, перезапустите компьютер и попробуйте еще раз.

Глава 4. Entity Analytics в действии

Об этом примере

В этом примере мы увидим, как добавление Entity Analytics может еще более улучшить результаты, которые можно получить от использования SPSS Modeler.

Этот пример использует поток *loan_entity_analytics.str*, в котором используется файл данных *loan_applications.csv*. Эти файлы доступны в каталоге *Demos* любой из установок SPSS Modeler, где установлена также SPSS Entity Analytics. Вы можете получить доступ к каталогу *Demos* из группы программ SPSS Modeler в меню Пуск Windows. Файл *loan_entity_analytics.str* находится в каталоге *Entity_Analytics*.

Примечание: Перед тем, как запустить этот пример потока, надо создать в вашей системе репозиторий. Сделайте это, прежде чем продолжить работу с примером. Дополнительную информацию смотрите в разделе “Создание репозитория” на стр. 14.

Давайте начнём со знакомой ситуации - руководители банка обеспокоены тем, какие клиенты могут оказаться неплатёжеспособными по кредитам, заявки на которые находятся на рассмотрении. IT-отдел банка давно использует SPSS Modeler, так что сотрудники отдела уже создали поток и построили прогнозную модель, исходя из существующих данных о 700 кредитах, которые банк уже выдал в прошлом. Эти кредиты либо уже были погашены, либо клиенты оказались неплатёжеспособными.

Исходная модель

Здесь мы покажем, как сотрудники банка строят модель и чему она их учит.



Рисунок 2. Исходный поток с моделированием узла

Данные *loan_applications.csv* содержат подробности о 150 клиентах, чьи заявки на кредиты находятся на рассмотрении, а также подробности о прошлых кредитах, что даёт в общей сложности 850 записей.

Не все поля этих данных полезны для построения прогноза, например, поля имён можно исключить. Узел Тип отфильтровывает бесполезные поля, задавая для них роль **Нет**. Для полей, которые влияют на прогноз, задается роль **Входное**, а для полей, значение которых модель предсказывает, задаётся роль **Назначение**.



Рисунок 3. Роли полей задаются в узле Тип

Поскольку модель должна делать свои прогнозы, основываясь исключительно на прошлых данных, поток включает в себя узел Выбор, который содержит только те кредиты, что *не* помечены, как находящиеся на рассмотрении; таким образом, исключаются 150 кредитов, находящихся на рассмотрении.

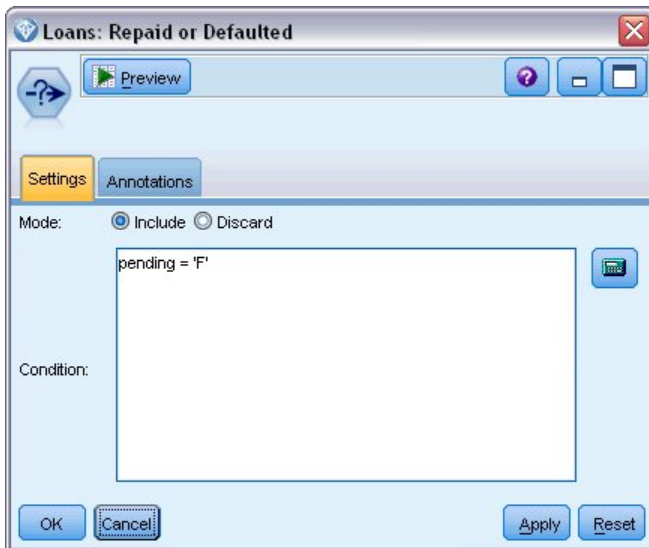


Рисунок 4. Исключение кредитов, находящихся на рассмотрении

Так как кредиты, находящиеся на рассмотрении, исключаются, узлу моделирования передаются только 700 оставшихся кредитов, среди которых есть, как погашенные, так и непогашенные в срок кредиты. Банк мог бы использовать один из нескольких алгоритмов SPSS Modeler, чтобы построить хорошую модель. В

данном случае случае они используют узел Дерево классификации и регрессии, который будет использован для построения модели для предсказания возможных неплательщиков, основанную на прошлом поведении клиентов банка.

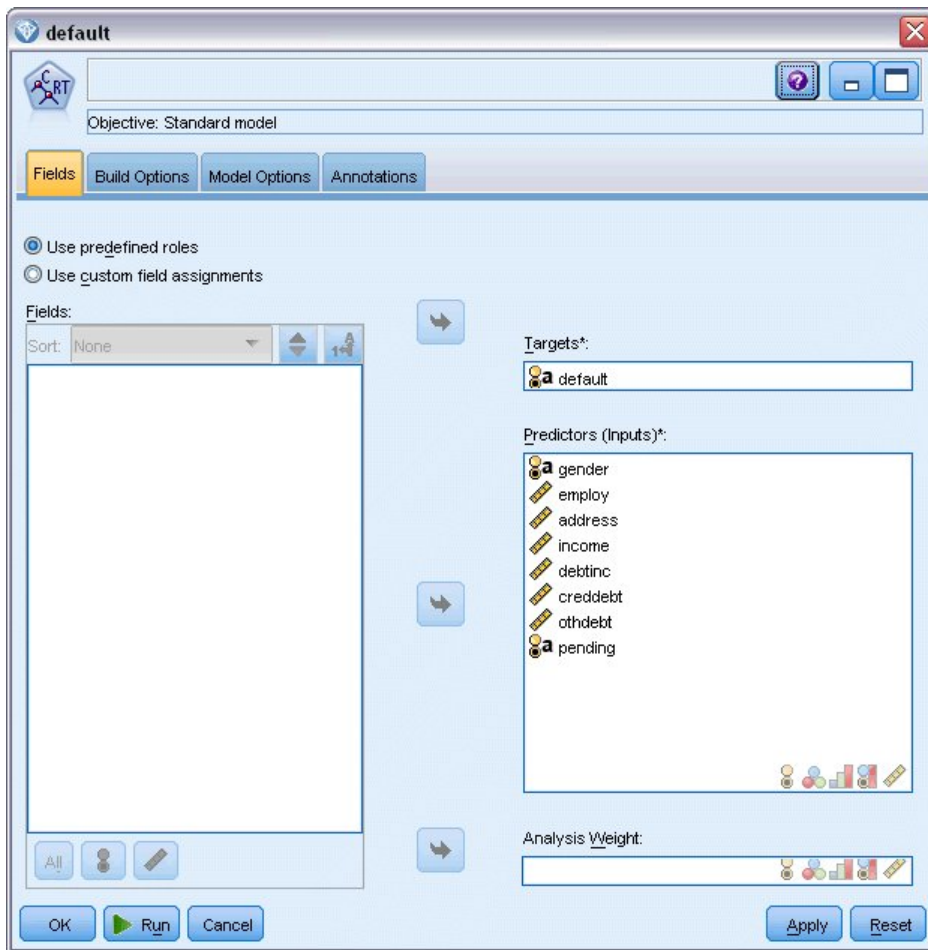


Рисунок 5. Назначение полей предиктора и назначения

Поля, используемые для вычисления прогноза, обозначаются как поля предиктора, а поле, значение которого модель пытается предсказать (в данном случае поле **default**) - как поле назначения, что было определено ранее в узле Тип.

Выполнение этого потока создаёт слепок модели, содержащий модель, построенную из полей предиктора.

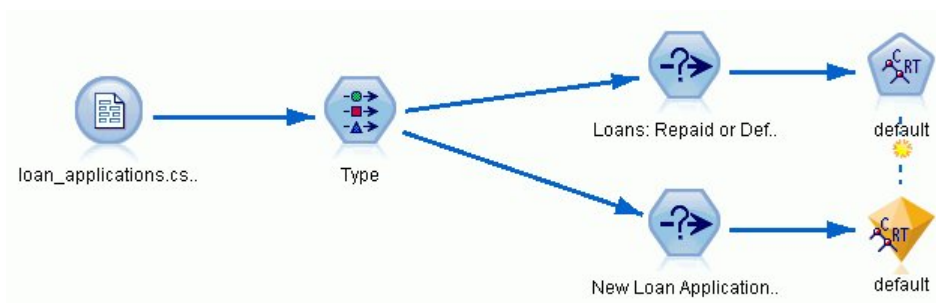


Рисунок 6. Поток с добавленным слепком модели

Теперь аналитик банка может использовать эту модель, чтобы спрогнозировать, какие заёмщики, вероятно, не будут выплачивать взносы по кредиту. Используя исходный набор данных, аналитик вставляет узел Выбор, который на этот раз включает в себя только 150 кредитных записей, помеченных как Ожидающие решения, вместо того, чтобы отбросить их. Аналитик передаёт эти записи прямо в модель, добавив узел Распределение для наглядного представления предсказаний модели.

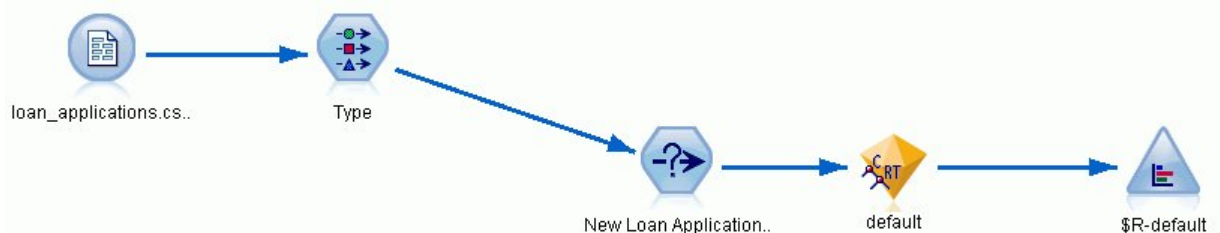


Рисунок 7. Поток выбора новых заявок на кредит с добавленным узлом Распределение

Узел Распределение показывает распределение значений поля $\$R\text{-default}$ в модели. Это поле добавляется к модели данных при выполнении узла Дерево классификации и регрессии. Это поле содержит прогноз, будет каждый новый заёмщик выплачивать взносы по кредиту или нет; позже мы будем использовать это поле, чтобы сравнить эффект от добавления Entity Analytics.

Запустив эту часть потока, аналитик получит из вывода узла Распределение прогноз, что 137 из 150 новых заёмщиков будут выплачивать взносы по кредиту. Оставшиеся 13 предсказаны как неплательщики, так что аналитик, скорее всего, будет рекомендовать банку отклонить их заявки.

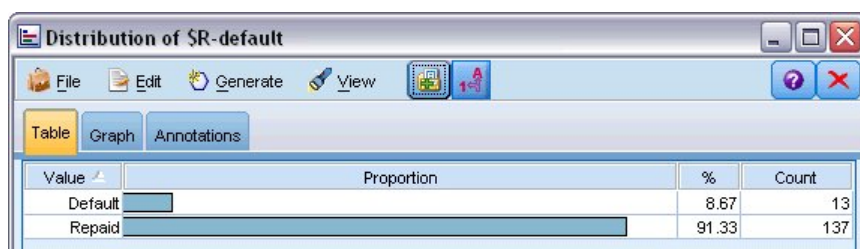


Рисунок 8. Вывод узла Распределение без Entity Analytics

Добавление Entity Analytics

Давайте посмотрим, можно ли улучшить ситуацию добавлением Entity Analytics в уравнение. Представьте себе, что вы - специалист по Entity Analytics, и вас вызывает банк, чтобы расследовать возможные записи о мошеннических операциях в записях о клиентах в данных источника. Там могут быть повторные записи, вызванные ошибками ввода данных, но также не исключено, что заёмщик мог попытаться скрыть свою личность. В любом случае банк должен знать истинную картину.

Чтобы разобрать приведённый пример, мы предположим, что репозиторий объектов уже создан. Дополнительную информацию смотрите в разделе “Создание репозитория” на стр. 14.

Получение данных источника в репозиторий

Сначала следует добавить узел Экспорт EA к узлу источника данных, чтобы можно было экспортировать данные источника в репозиторий объекта.

До начала экспорта данных следует отобразить поля в источнике данных на характеристики в репозитории объектов. Это необходимо сделать, потому что в различных источниках данных могут использоваться различные имена полей для информации одного и того же типа. В репозитории объектов есть стандартный набор типов информации (называемых также “характеристиками”), чтобы избежать дублирования.

В узле Экспорт EA задайте подробности для репозитория: подробности соединения, тег источника (чтобы идентифицировать источник данных, в данном случае - **TEST**), тип объектов (набор используемых характеристик — одна с именем **PERSON**), и поле ключа уникальности (чтобы уникальным образом идентифицировать каждую запись). В данном случае используйте в качестве ключа уникальности поле **key**.

Теперь можно задать отображения. В наборе характеристик, который вы используете, есть характеристики, соответствующие полям *fname*, *mname*, *lname*, *generation*, *dob*, *gender*, *addr1*, *city*, *country*, *postcode*, *phone*, *email*, *ssn*, *drlic* и *passport*.

Начнем с задания отображения для поля *fname*. Дважды щёлкните по столбцу **Отображено на характеристику** в таблице на строке *fname*, прокрутите ниже до записи **NAME.GIVEN_NAME** и щёлкните по ней, чтобы создать отображение.

Теперь отобразите остальные поля, у которых есть соответствующие им характеристики, чтобы полный набор отображений выглядел бы следующим образом:

Таблица 12. Поля, отображённые на характеристики репозитория.

Поле	Отображено на характеристику
<i>fname</i>	NAME.GIVEN_NAME
<i>mname</i>	NAME.MIDDLE_NAME
<i>lname</i>	NAME.SUR_NAME
<i>generation</i>	NAME.NAME_GEN
<i>dob</i>	DOB.DOB
<i>gender</i>	GENDER.GENDER
<i>addr1</i>	ADDRESS.ADDR1
<i>city</i>	ADDRESS.CITY
<i>country</i>	ADDRESS.COUNTRY
<i>postcode</i>	ADDRESS.POSTAL_CODE
<i>phone</i>	PHONE.PHONE_NUM
<i>email</i>	EMAIL_ADDR.ADDR
<i>ssn</i>	SSN.ID_NUM
<i>drlic</i>	DRLIC.ID_NUM
<i>passport</i>	PASSPORT.ID_NUM

Нажмите кнопку **Выполнить**, чтобы экспортировать данные в репозиторий. Это займет некоторое время; экспорт данных будет завершён, когда диалоговое окно Обратная связь выполнения закроется.

Чтение установленных соответствий

Когда вы экспортируете данные в репозиторий, система Entity Analytics начинает разрешать возможные конфликты соответствия, назначая уникальный идентификатор объекта, который вы позже увидите в поле *\$EA-ID*. (Примечание: Не путайте этот идентификатор с полем Ключ уникальности в узле Экспорт EA, которое используется для уникальной идентификации записей источника данных.)

Первый шаг к чтению установленных соответствий состоит в добавлении узла источника Entity Analytics (EA) к потоку. Этот узел источника не должен быть соединён ни с чем на этом этапе.

Откройте узел источника Entity Analytics (EA) и задайте подробности для репозитория объектов. При этом выводится список источников данных, которые уже экспортированы в репозиторий, - в данном случае в этом списке только один источник.

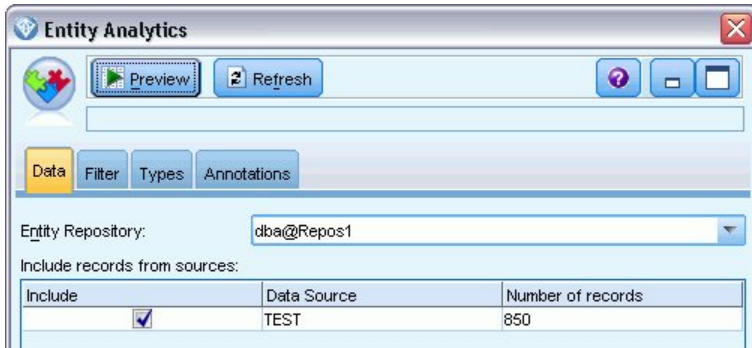


Рисунок 9. Выбор источника данных в репозитории

Включите переключатель для источника данных **TEST** и нажмите кнопку ОК.

Посмотрим, что система Entity Analytics сделала с данными. Присоедините узел Таблица к узлу источника Entity Analytics (EA), откройте узел Таблица и нажмите кнопку **Выполнить**, чтобы вывести окно вывода узла Таблица.

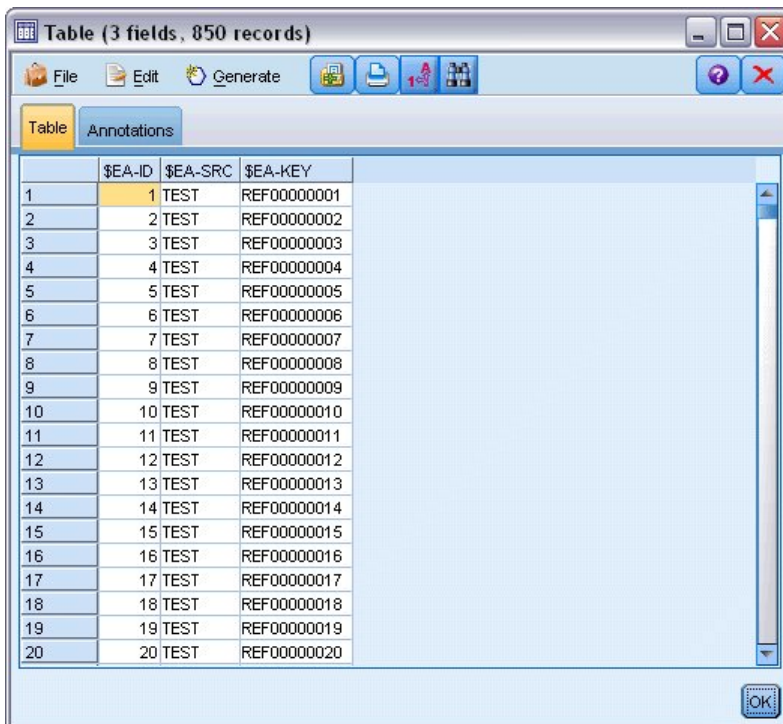


Рисунок 10. Вывод узла Таблица

Только одно поле выглядит знакомым - поле, помеченное *SEA-KEY*. На самом деле это поле *key* из источника данных, и оно находится здесь, потому что вы выбрали его в качестве поля ключа уникальности в узле Экспорт Entity Analytics (EA).

Система добавила еще два поля. Поле *SEA-ID* - это уникальный идентификатор, но не для исходных записей, а для установленных соответствий. Разницу мы сейчас увидим. Поле *SEA-SRC* определяет, откуда поступили данные - здесь в нем будет значение **TEST**, потому что такое значение вы задали для тега источника в узле Экспорт EA.

Что случилось со всеми другими полями в источнике данных? Не беспокойтесь, они всё ещё находятся в репозитории – это происходит только из соображений производительности; узел источника Entity Analytics (EA) передает для дальнейшей обработки только минимальный набор полей.

Теперь прокрутите вывод узла Таблица вниз к строке 385.

	\$EA-ID	\$EA-SRC	\$EA-KEY
376	376	TEST	REF00000376
377	377	TEST	REF00000377
378	378	TEST	REF00000378
379	379	TEST	REF00000379
380	380	TEST	REF00000380
381	381	TEST	REF00000381
382	382	TEST	REF00000382
383	383	TEST	REF00000383
384	384	TEST	REF00000384
385	45	TEST	REF00000385
386	386	TEST	REF00000386
387	387	TEST	REF00000387
388	388	TEST	REF00000388
389	389	TEST	REF00000389
390	390	TEST	REF00000390
391	391	TEST	REF00000861
392	392	TEST	REF00000392
393	393	TEST	REF00000393
394	394	TEST	REF00000394
395	395	TEST	REF00000395
396	396	TEST	REF00000396

Рисунок 11. Различия между строками вывода узла Таблица и номерами \$EA-ID

Заметьте, как номер \$EA-ID, кажется, выпадает здесь из последовательности. Система Entity Analytics определила, что в записи REF00000385 упоминается человек, идентифицируемый как объект 45, который упоминается также в записи REF0000045. Прокрутив вывод дальше вниз, находим, что есть ещё номера вне последовательности, например, в строках 485, 517, 520 и так далее. Нам следует рассмотреть это более внимательно.

Во-первых, давайте подчеркнём тот факт, что поле \$EA-KEY содержит данные из поля *key* в данных источника, переименовав его в поле *key*. Присоедините узел Фильтр к узлу источника Entity Analytics (EA) и откройте этот узел Фильтр. Дважды щёлкните по строке \$EA-KEY во втором столбце Поле и введите *key*.

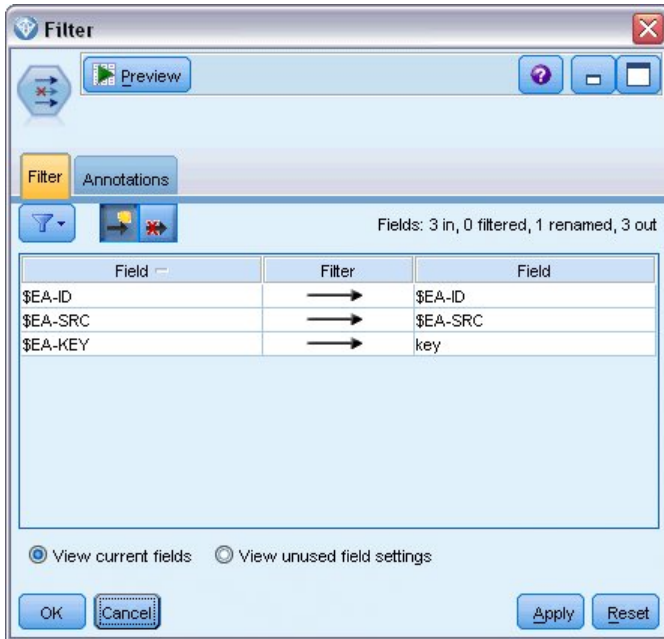


Рисунок 12. Переименование поля \$EA-KEY

Нажмите кнопку **OK**, чтобы закрыть узел Фильтр.

Теперь нам нужно отсортировать идентификаторы объекта *\$EA-ID* в возрастающем порядке. Присоедините узел Сортировка к узлу Фильтр. Откройте узел Сортировка, нажмите верхнюю кнопку рядом с таблицей **Сортировать по**, выберите **\$EA-ID** и нажмите кнопку **OK**.



Рисунок 13. Сортировка идентификаторов объекта в возрастающем порядке.

Оставьте порядок сортировки **По возрастанию** и нажмите кнопку **OK**.

Теперь нужно создать дополнительное поле, указывающее, является ли запись уникальной или дубликатом. Присоедините узел Производные данные к узлу Сортировка. Откройте узел Производные данные и задайте для имени **Производное поле** значение `IsDuplicate`. Из списка **Произвести как** выберите **Флаг**, что задаст также для **Типа поля** значение **Флаг**. Задайте для поля **Значение true** значение **Дубликат**, а для поле **Значение false** - **Уникальный**.

Чтобы найти повторные записи, следует использовать специальную функцию последовательности с именем @OFFSET, поставляемую вместе сSPSS Modeler.

Введите в поле **Если** следующее:

```
'$EA-ID' = @OFFSET('$EA-ID',1) or '$EA-ID' = @OFFSET('$EA-ID',-1))
```



Рисунок 14. Задание условия в узел Производные данные

Используя ID объектов, отсортированные по возрастанию, функция @OFFSET проверяет, совпадают ли соседние ID объектов; в случае совпадения соответствующие записи - это дубликаты. В таком случае для *IsDuplicate* задается значение *Дубликат*, иначе задается значение *Уникальный*.

Нажмите кнопку **ОК**, чтобы закрыть узел.

Чтобы увидеть эффект узла Производные данные, присоедините узел Таблица к узлу Производные данные, откройте узел Таблица и нажмите кнопку **Выполнить**. Прокрутите окно вывода узла Таблица вниз до строки 45.

	\$EA-ID	\$EA-SRC	key	IsDuplicate
39	39	TEST	REF00000039	Unique
40	40	TEST	REF00000040	Unique
41	41	TEST	REF00000041	Unique
42	42	TEST	REF00000042	Unique
43	43	TEST	REF00000043	Unique
44	44	TEST	REF00000044	Unique
45	45	TEST	REF00000045	Duplicate
46	45	TEST	REF00000385	Duplicate
47	46	TEST	REF00000046	Unique
48	47	TEST	REF00000047	Unique
49	48	TEST	REF00000048	Unique
50	49	TEST	REF00000049	Unique
51	50	TEST	REF00000050	Unique
52	51	TEST	REF00000051	Unique
53	52	TEST	REF00000052	Unique
54	53	TEST	REF00000053	Unique
55	54	TEST	REF00000054	Unique
56	55	TEST	REF00000055	Unique
57	56	TEST	REF00000056	Unique
58	57	TEST	REF00000057	Unique

Рисунок 15. Вывод узла Производные данные

Помните, мы смотрели на вывод непосредственно узла источника Entity Analytics (EA)? Система уже определила, что в записи REF00000385 упоминается тот же самый человек, что и в объекте 45. Теперь сделаем еще шаг вперед и отметим тот факт, что записи REF00000045 и REF00000385 - это дубликаты, поскольку в обеих упоминается объект 45.

Прокрутите окно вывода дальше вниз, и вы увидите другие записи, которые помечены, как дубликаты.

Чтобы получить отчет по внесённым в список дубликатам записей, присоедините узел Отчёт (из вкладки Вывод палитры узлов) к узлу производных данных *IsDuplicate*. Откройте узел Отчёт, скопируйте следующий текст в поле ввода вкладки Шаблон и нажмите кнопку **Выполнить**.

```
<html>
<h1>Список дубликатов записей клиентов.

<h2>Сгенерирован следующий отчет: [@TODAY]

<h2>Дубликаты записей
<table>
  <tr>
    <td>ID объекта</td>
    <td>Ключ</td>
  </tr>

#WHERE IsDuplicate = "Дубликат"

  <tr>
    <td>['$EA-ID']</td>
    <td>[key]</td>
  </tr>
```

```
#  
</table>  
</html>
```

Это даёт следующий вывод:

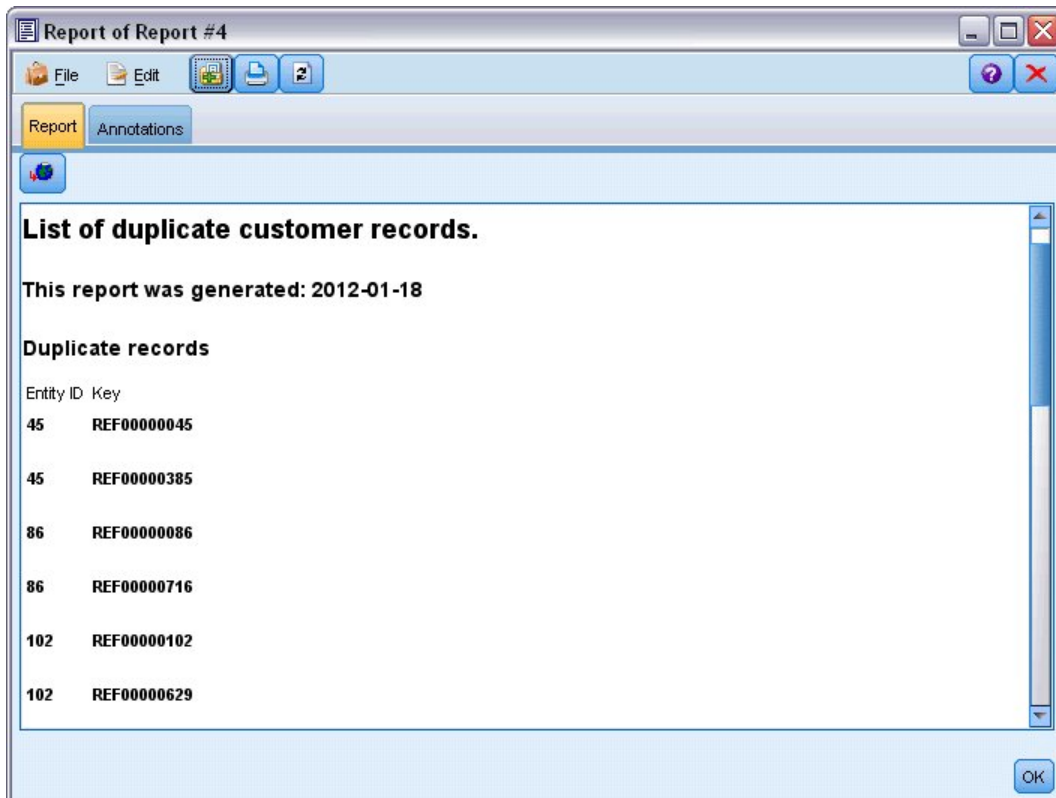


Рисунок 16. Вывод узла Отчёт

В данном случае отчёт использует формат HTML, хотя доступны также форматы XML или ASCII.

Сравнение вывода Entity Analytics с исходной моделью

Заключительный этап этого примера состоит в том, чтобы увидеть, вносит ли добавление Entity Analytics какое-нибудь отличие от первоначальных прогнозов банка. Возможно, вы помните, что исходная модель предсказывала 13 неплательщиков из 150 рассматриваемых заявок. Вы собираетесь использовать узел Слияние, чтобы совместить вывод этой модели с информацией о дубликатах записей из Entity Analytics и посмотреть, изменит ли это предсказания.

Сначала следует убедиться, что у новых полей, добавляемых Entity Analytics, правильные типы данных или уровни измерения, как они называются в SPSS Modeler. Присоедините узел Тип к узлу производных данных **IsDuplicate**, откройте узел Тип и нажмите кнопку **Читать значения**.

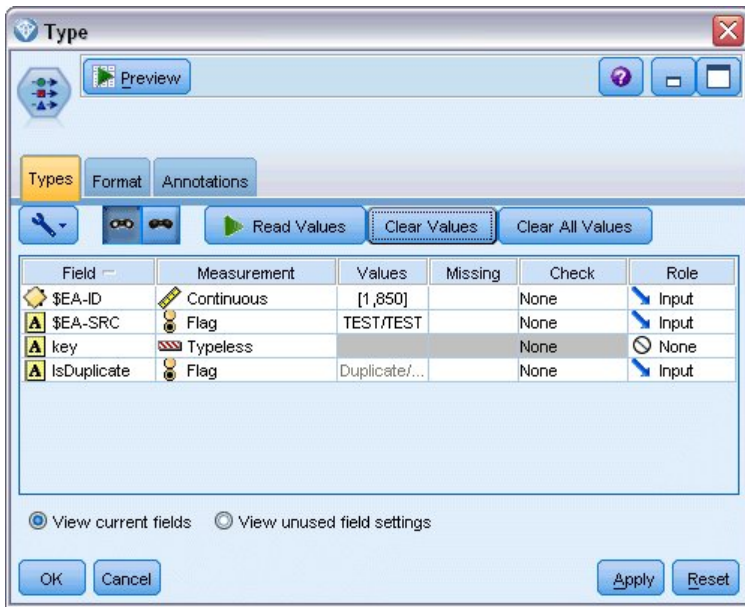


Рисунок 17. Параметры для узла Тип

Теперь можно добавить узел Слияние. Присоедините его к узлу Тип и также соедините его с желтым слепком, содержащим исходную модель. Для этого щелкните правой кнопкой мыши по желтому слепку, выберите **Соединить** и затем щелкните по узлу Слияние - теперь у него будет две входящие стрелки.

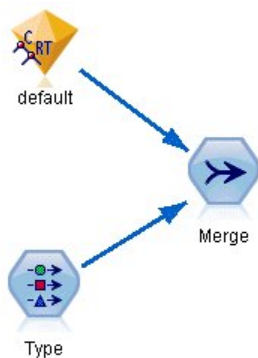


Рисунок 18. Вводы для узла Слияние

Откройте узел Слияние, задайте для **Метода слияния** значение **Ключи** и нажмите кнопку со стрелкой вправо, чтобы переместить поле **key** из списка **Возможные ключи** в список **Ключи для слияния**, затем нажмите кнопку **ОК**.

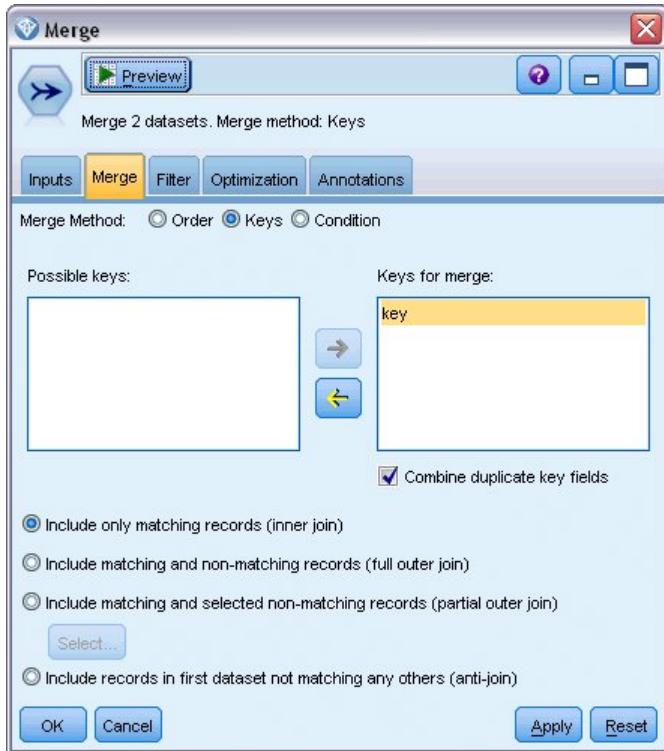


Рисунок 19. Задание поля ключа для операции слияния

Теперь вы почти готовы выполнить сравнение. Но если вы присоединили узел Распределение и запустите его в этот момент, вы не увидите никаких отличий от первоначального прогноза. Хотя теперь поток соединяется с выходом из исходного слепка модели с новыми полями, созданными Entity Analytics, в само поле прогноза (*\$R-default*) в модели данных ещё занесена новая информация.

Чтобы обновить данные, надо использовать узел Заполнение, который может заменить значения полей. Присоедините узел Заполнение к узлу Слияние и откройте узел Заполнение.

Нажмите верхнюю кнопку справа от **Заполнить поля**, прокрутите список до конца, выберите **\$R-default** и нажмите кнопку **ОК**. Это то поле, значения в котором следует изменить, если условие, указанное в остальной части диалогового окна, выполнено.

Чтобы задать это условие, убедитесь, что в поле **Замена** задано значение **На основе условия**, затем в поле **Условие** введите:

```
default != "default" and IsDuplicate = "Дубликат"
```

В поле **Заменить на** введите:

```
"default"
```

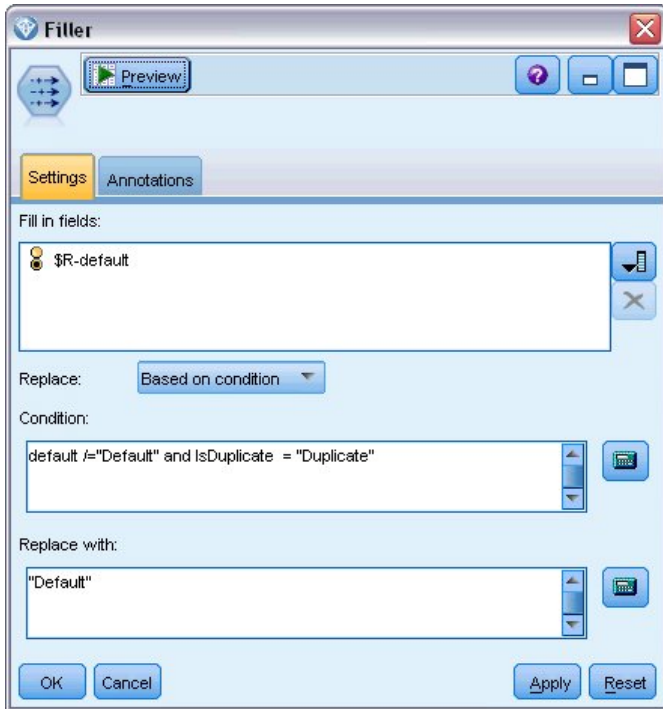


Рисунок 20. Указание условия для замены значений поля

Эти параметры требуют некоторого пояснения. Условие указывает, что для каждой записи, где значение поля *default* в исходном наборе данных не равно **Default**, а запись была помечена как дубликат, для поля *\$R-default* в модели задаётся значение **Default**.

Поле *\$R-default* - это поле в модели, которое содержит прогноз, будет ли некий клиент неплатёжеспособным по кредиту. Таким образом, клиенты с повторяющимися записями добавляются в модель в качестве потенциальных неплательщиков.

Нажмите кнопку **OK**, чтобы закрыть узел Заполнение.

Теперь вы, наконец, готовы увидеть те самые различия, которые появляются в результате работы Entity Analytics. Присоедините узел Распределение из палитры Графики к узлу Заполнение и откройте узел Распределение. Щёлкните по списку **Поле** и выберите поле **\$R-default**.

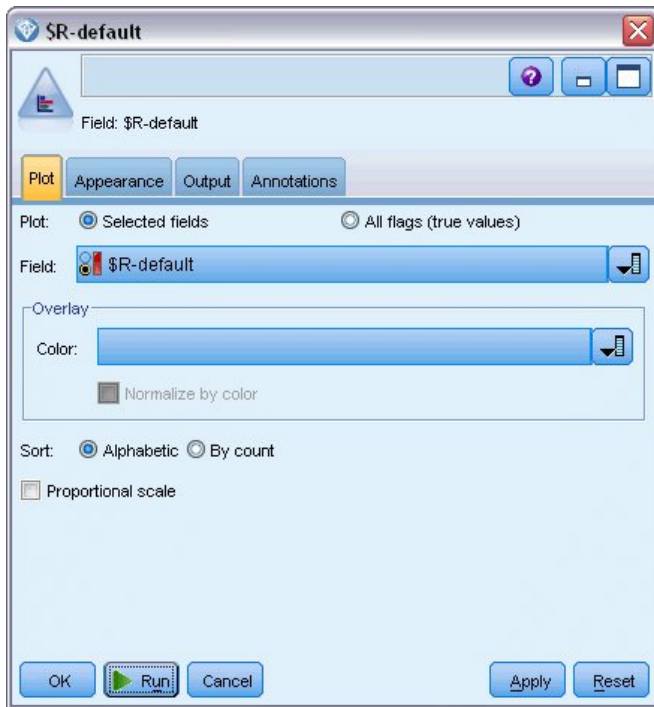


Рисунок 21. Параметры для узла Распределение

Нажмите кнопку **Выполнить**, чтобы построить диаграмму нового прогноза.

Value	Proportion	%	Count
Default		10.67	16
Repaid		89.33	134

Рисунок 22. Вывод узла Распределение после Entity Analytics

Теперь есть 16 связанных с риском заявок, а не 13. Эти дополнительные заявки могли бы обойтись очень дорого, если бы клиенты отказались выплачивать кредит, так что можно графически показать выгоды банка от включения Entity Analytics в систему оценки рисков для банка.

Итог

Данный пример показывает, как, используя Entity Analytics, можно устранить дублирование записей в данных о людях или организациях, что приводит к улучшению качества предсказаний.

Примечание: В идеале вам следовало бы устранить дублирующие записи, прежде чем делать любую другую обработку. Это можно было бы сделать, используя узел автоматизированной подготовки данных (Automated Data Preparation, ADP) для анализа данных и выявления исправлений, для отсеивания сомнительных и бесполезных полей, для генерирования новых атрибутов, когда это уместно, и повышения производительности с помощью интеллектуальных методов экранирования.

Сочетание Entity Analytics и автоматизированной подготовки данных поможет обеспечить чистоту ваших данных, насколько это возможно.

Приложение. Свойства сценариев для IBM SPSS Modeler Entity Analytics

Сценарии в IBM SPSS Modeler Entity Analytics

Сценарии в SPSS Entity Analytics - это мощное средство для процессов автоматизации в пользовательском интерфейсе. Сценарии могут выполнять действия того же типа, которые выполняются с помощью мыши или клавиатуры, и их можно использовать для автоматизации задач, которые при ручном выполнении могли бы быть многократно повторяемыми или требующими большого времени. Объяснение использования сценариев смотрите в руководстве *ModelerScriptingAutomation.pdf*, доступном с SPSS Modeler.

Общие свойства

В следующей таблице приводятся свойства, общие для узлов IBM SPSS Modeler Entity Analytics. Информация о конкретных узлах дана в следующих разделах.

Таблица 13. Общие свойства

Имя свойства	Тип переменной	Описание свойства
entity_repository	['поле', 'поле', ... , 'поле']	Строка соединения для репозитория. Формат: ['имя_репозитория', 'имя_пользователя', 'пароль'] Пример: entity_repository= ['repos1', 'dba', 'psw1']
entity_type	string	Используемый тип объектов (набор характеристик). Пример: entity_type = 'PERSON'

Свойства entityanalytics_exportnode



Узел экспорта EA - это конечный узел, который считывает данные объектов из источника данных и экспортирует данные в репозиторий для разрешения объекта.

Таблица 14. Свойства entityanalytics_exportnode

Свойства entityanalytics_exportnode	Тип переменной	Описание свойства
mode	Add PurgeFirst	Режим экспорта. Add добавляет записи файла источника к существующему содержимому репозитория; PurgeFirst удаляет существующее содержимое репозитория перед экспортом.
source_tag	string	Идентификатор источника данных. Пример: source_tag = 'CUST'

Таблица 14. Свойства *entityanalytics_exportnode* (продолжение)

Свойства <i>entityanalytics_exportnode</i>	Тип переменной	Описание свойства
unique_key_field	<i>string</i>	Входное поле используемых идентификаторов уникальности для записей данных. Пример: unique_key_field = 'ID'
field_mapping	[['field_name' 'характеристика.элемент' 'тип_использования']...]	Отображает входные Поля на соответствующие характеристики в репозитории. Пример: field_mapping = [['fname' 'NAME.GIVEN_NAME' ''] ['addr1' 'ADDRESS.ADDR1' 'PRIMARY']] <i>Примечание:</i> Чтобы задать для <i>usage_type</i> эквивалент "(Auto)", используйте '', как в первом примере выше.

Свойства *entityanalytics_sourcenode*



Узел источника аналитики объектов (Entity Analytics, EA) теперь считывает разрешенные объекты из репозитория и передает данные в поток для дальнейшей обработки, такой как форматирование и включение в отчеты.

Таблица 15. Свойства *entityanalytics_sourcenode*

Свойства <i>entityanalytics_sourcenode</i>	Тип переменной	Описание свойства
source_tags	<i>список</i>	Список тегов для источников данных, помещаемых в репозиторий. Пример: source_tags=['LOANS', 'CUSTOMERS']
relationships	None Закреть All	Критерий соответствия для получения подробностей взаимосвязей из репозитория. None не возвращает никаких взаимосвязей. Close возвращает тесные соответствия в зависимости от подробностей, таких как степень разделения. All возвращает все возможные взаимосвязи.
max_degree_separation	<i>целое</i>	Минимум 0, максимум 3.
output_entity_type	<i>строка</i>	Список типов объектов, используемых в репозитории.

Свойства *entityanalytics_processnode*



Узел потоковой EA сравнивает новые наблюдения с данными объекта в репозитории.

Таблица 16. Свойства *entityanalytics_processnode*

Свойства <i>entityanalytics_processnode</i>	Тип переменной	Описание свойства
match	Exact ByIdentifier All	Критерий соответствия для получения объектов из репозитория. Exact возвращает только точные совпадения. ByIdentifier возвращает точные совпадения и объекты с одинаковыми идентификаторами. All возвращает все возможные соответствия.
save_search_records	<i>логический</i>	
relationships	None Close All	Критерий соответствия для получения подробностей взаимосвязей из репозитория. None не возвращает никаких взаимосвязей. Close возвращает тесные соответствия в зависимости от подробностей, таких как степень разделения. All возвращает все возможные взаимосвязи.
max_degree_separation	<i>целое</i>	Минимум 0, максимум 3.
output_entity_type	<i>строка</i>	Список типов объектов, используемых в репозитории.

Уведомления

Эта информация относится к продуктам и сервису, предлагаемым в США. Этот материал может быть доступен от IBM на других языках. Однако для его получения может понадобиться приобрести продукт или версию продукта на нужном языке.

IBM может не предоставлять в других странах продукты, услуги и аппаратные средства, описанные в данном документе. За информацией о продуктах и услугах, предоставляемых в вашей стране, обращайтесь к местному представителю IBM. Ссылки на продукты, программы или услуги IBM не означают и не предполагают, что можно использовать только указанные продукты, программы или услуги IBM. Разрешается использовать любые функционально эквивалентные продукты, программы или услуги, если при этом не нарушаются права IBM на интеллектуальную собственность. Однако ответственность за оценку и проверку работы любого продукта, программы или сервиса, не произведенного корпорацией IBM, лежит на пользователе.

IBM может располагать патентами или рассматриваемыми заявками на патенты, относящимися к предмету данного документа. Предъявление данного документа не предоставляет какую-либо лицензию на эти патенты. Вы можете послать письменный запрос о лицензии по адресу:

*IBM Director of Licensing
IBM Corporation
North Castle Drive, MD-NC119
Armonk, NY 10504-1785
US*

По поводу лицензий, связанных с использованием наборов двухбайтных символов (DBCS), обращайтесь в отдел интеллектуальной собственности IBM в вашей стране или направьте запрос в письменной форме по адресу:

*Intellectual Property Licensing
Legal and Intellectual Property Law
IBM Japan Ltd.
19-21, Nihonbashi-Hakozakicho, Chuo-ku
Tokyo 103-8510, Japan*

INTERNATIONAL BUSINESS MACHINES CORPORATION ПРЕДСТАВЛЯЕТ ДАННУЮ ПУБЛИКАЦИЮ "КАК ЕСТЬ", БЕЗ КАКИХ-ЛИБО ГАРАНТИЙ, КАК ЯВНЫХ, ТАК И ПОДРАЗУМЕВАЕМЫХ, ВКЛЮЧАЯ, НО НЕ ОГРАНИЧИВАЯСЬ ТАКОВЫМИ, ПРЕДПОЛАГАЕМЫЕ ГАРАНТИИ СОБЛЮДЕНИЯ ЧЬИХ-ЛИБО АВТОРСКИХ ПРАВ, ВОЗМОЖНОСТИ КОММЕРЧЕСКОГО ИСПОЛЬЗОВАНИЯ ИЛИ ПРИГОДНОСТИ ДЛЯ КАКИХ-ЛИБО ЦЕЛЕЙ И СООТВЕТСТВИЯ ОПРЕДЕЛЕННОЙ ЦЕЛИ. В некоторых странах для ряда сделок не допускается отказ от явных или предполагаемых гарантий; в таком случае данное положение к вам не относится.

Эта информация может содержать технические неточности и типографские ошибки. В представленную здесь информацию периодически вносятся изменения; эти изменения будут включаться в новые издания данной публикации. Фирма IBM может в любое время без уведомления вносить изменения и усовершенствования в продукты и программы, описанные в этой публикации.

Любые ссылки в этой публикации на сайты, не принадлежащие IBM, приведены только для удобства и никоим образом не означают их поддержки. Материалы на этих сайтах не входят в число материалов по данному продукту IBM, и весь риск пользования этими сайтами несете вы сами.

Любую предоставленную вами информацию IBM может использовать или распространять любым способом, какой сочтет нужным, не беря на себя никаких обязательств по отношению к вам.

Если обладателю лицензии на данную программу понадобятся сведения о возможности: (i) обмена данными между независимо разработанными программами и другими программами (включая данную) и (ii) совместного использования таких данных, он может обратиться по адресу:

*IBM Director of Licensing
IBM Corporation
North Castle Drive, MD-NC119
Armonk, NY 10504-1785
US*

Такая информация может быть доступна при соответствующих условиях и соглашениях, включая в некоторых случаях взимание платы.

Описанную в данном документе лицензионную программу и все прилагаемые к ней лицензированные материалы IBM предоставляет на основе положений Соглашения между IBM и Заказчиком, Международного Соглашения о Лицензиях на Программы IBM или любого эквивалентного соглашения между IBM и заказчиком.

Данные производительности и примеры клиентов представлены только для иллюстрации. Фактическая производительность зависит от конкретной конфигурации и условий работы.

Информация о продуктах других компаний (не IBM) получена от поставщиков этих продуктов, из их опубликованных объявлений или из иных общедоступных источников. IBM не производила тестирование этих продуктов и никак не может подтвердить информацию о их точности работы и совместимости, а также прочие заявления относительно продуктов других компаний (не IBM). Вопросы о возможностях продуктов других компаний (не IBM) следует направлять поставщикам этих продуктов.

Все утверждения о будущих планах и намерениях IBM могут быть изменены или отменены без уведомлений, и описывают исключительно цели фирмы.

Эти сведения содержат примеры данных и отчетов, используемых в повседневных деловых операциях. Чтобы проиллюстрировать их настолько полно, насколько это возможно, данные примеры включают имена индивидуумов, названия компаний, брендов и продуктов. Все эти имена являются вымышленными и любое их сходство с реальными именами и адресами предприятий является случайным.

Товарные знаки

IBM, логотип IBM, и ibm.com являются товарными знаками или зарегистрированными товарными знаками компании International Business Machines Corp., зарегистрированными во многих странах мира. Прочие наименования продуктов и услуг могут быть товарными знаками, принадлежащими IBM или другим компаниям. Текущий список товарных знаков IBM смотрите на веб-сайте "Copyright and trademark information" (Информация об авторских правах и товарных знаках) по адресу www.ibm.com/legal/copytrade.shtml.

Adobe, логотип Adobe, PostScript и логотип PostScript являются либо зарегистрированными товарными знаками, либо товарными знаками корпорации Adobe Systems в Соединенных Штатах и/или других странах.

Intel, логотип Intel, Intel Inside, логотип Intel Inside, Intel Centrino, логотип Intel Centrino, Celeron, Intel Xeon, Intel SpeedStep, Itanium и Pentium являются товарными знаками или зарегистрированными товарными знаками компании Intel или ее дочерних компаний в Соединенных Штатах и других странах.

Linux является зарегистрированным товарным знаком Linus Torvalds в Соединенных Штатах и других странах.

Microsoft, Windows, Windows NT и логотип Windows являются товарными знаками корпорации Microsoft в Соединенных Штатах и других странах.

UNIX является зарегистрированным товарным знаком The Open Group в Соединенных Штатах и других странах.

Java и все основанные на Java товарные знаки и логотипы - товарные знаки или зарегистрированные товарные знаки Oracle и/или его филиалов.

Правила и условия для документации продукта

Разрешения для использования этих публикаций предоставляются на следующих условиях.

Применимость

Данные правила и условия являются дополнением к правилам использования для сайта IBM.

Персональное использование

Вы можете воспроизводить эти публикации для персонального некоммерческого использования при условии сохранения всех замечаний о правах собственности. Вам запрещается распространять эти публикации, полностью или по частям, демонстрировать их или создавать из них производные продукты без явного на то согласия от IBM.

Коммерческое использование

Вам предоставляется право воспроизводить эти публикации исключительно в пределах своего предприятия при условии, что будут воспроизведены все замечания об авторских правах. За пределами вашего предприятия вам запрещается распространять эти публикации, полностью или по частям, демонстрировать их или создавать из них производные продукты без явного на то согласия от IBM.

Права

За исключением прав, явным образом предоставляемых настоящим разрешением, никаких иных разрешений, лицензий и прав, ни явных, ни подразумеваемых, в отношении публикаций и любой содержащейся в них информации, данных, программ или иной интеллектуальной собственности, не предоставляется.

IBM оставляет за собой право отозвать разрешения, предоставленные этим документом, если, по мнению IBM, использование публикаций наносит ущерб IBM или, как это установлено IBM, вышеприведенные инструкции не соблюдаются должным образом.

Запрещается загружать, экспортировать или реэкспортировать эту информацию, если при этом не будут полностью соблюдаться все применимые законы и постановления, включая все законы и постановления США, касающиеся экспорта.

IBM НЕ ДАЕТ НИКАКИХ ГАРАНТИЙ ОТНОСИТЕЛЬНО СОДЕРЖАНИЯ ЭТИХ ПУБЛИКАЦИЙ. ПУБЛИКАЦИИ ПРЕДСТАВЛЯЮТСЯ "КАК ЕСТЬ", БЕЗ КАКИХ-ЛИБО ГАРАНТИЙ, ЯВНЫХ ИЛИ ПОДРАЗУМЕВАЕМЫХ, ВКЛЮЧАЯ (НО НЕ ОГРАНИЧИВАЯСЬ ТАКОВЫМИ) ПРЕДПОЛАГАЕМЫЕ ГАРАНТИИ ОТСУТСТВИЯ НАРУШЕНИЙ, КОММЕРЧЕСКОЙ ПРИГОДНОСТИ ИЛИ СООТВЕТСТВИЯ ОПРЕДЕЛЕННОЙ ЦЕЛИ.

Индекс

Е

- Entity Analytics
 - использование с IBM SPSS Modeler 5
 - использование с другими продуктами IBM SPSS 31
 - определение 1
 - сравнение с прогнозной аналитикой 2

А

- анонимизация характеристик репозиторий объектов 22

В

- вывод
 - из Entity Analytics 31

И

- информация о типе, задание для Entity Analytics 27
- источник данных, выбор для Entity Analytics 26
- источники данных
 - просмотр для Entity Analytics 15, 30
 - соединение с Entity Analytics 6, 14
- исходные узлы
 - Entity Analytics 9, 26

К

- ключи уникальности
 - Entity Analytics 7
 - репозиторий объектов 17
- конфигурация
 - репозиторий объектов 18, 25

Н

- назначения портов
 - конфигурирование для Entity Analytics 32
- новые наблюдения, сравнение с репозиторием Entity Analytics 28

О

- отображение полей
 - на характеристики репозитория объектов 7, 16, 17, 18, 19, 29, 30
- очистка
 - репозиторий объектов 35

П

- переименование
 - поля данных для Entity Analytics 27
- порог для соответствия объектов, Entity Analytics 25
- правила распознавания, Entity Analytics 25

Р

- распознавание объектов, Entity Analytics 7
- распознавание, Entity Analytics 7
- регистрационные данные администратора
 - управление для Entity Analytics 33
- репозиторий
 - Entity Analytics 6, 7, 13, 14, 15, 16, 17, 18, 19, 21, 22, 25, 28, 29, 30, 35, 36
 - каталог хранения для Entity Analytics, изменение 33
 - управление Entity Analytics 32
- репозиторий объектов 13
 - административные задачи 32
 - анонимизация 22
 - задание свойств потока 34
 - конфигурирование 13, 18, 25
 - конфигурирование назначений портов 32
 - опции 15
 - очистка 35
 - перемещение в другой каталог хранения 33
 - соединение с IBM SPSS Modeler 7
 - создание 6, 14, 15
 - сравнение новых наблюдений с 28
 - удаление 35, 36
 - удаление ненужных данных 35
 - управление 19
 - управление регистрационными данными администратора 33
 - элементы 21

С

- свойства
 - сценарий 53
- свойства потока
 - задание для Entity Analytics 34
- соответствие объектов, задание порога 25
- сценарий
 - свойства 53

Т

- теги источника
 - репозиторий объектов 17
- типы использования, Entity Analytics 22

типы объектов

- Entity Analytics 22
- репозиторий объектов 17

У

- удаление
 - репозиторий объектов 35, 36
- удаление ненужных данных
 - репозиторий объектов 35
- узел источника Entity Analytics (EA) 26
- узел потоков EA, Entity Analytics 28
- узел Экспорт EA, Entity Analytics 13
- узлы
 - добавление к потоку Entity Analytics 27
- узлы процесса
 - Entity Analytics 10, 28
- узлы экспорта
 - Entity Analytics 7, 13
- установленные соответствия, анализ с помощью Entity Analytics 26

Х

- характеристики
 - репозиторий объектов 19

Э

- экспорт
 - данных в репозиторий объектов 7
- элементы
 - репозиторий объектов 7, 16, 17, 18, 19, 21, 29, 30



Напечатано в Дании