

IBM® Analytical Decision Management
버전 17 릴리스 0

IBM® SPSS® Modeler
Advantage

IBM

참고

이 정보와 이 정보가 지원하는 제품을 사용하기 전에, 반드시 103 페이지의 『주의사항』의 일반 정보를 읽으십시오.

제품 정보

이 개정판은 새 개정판에서 별도로 표시되지 않는 한, IBM Analytical Decision Management의 버전 17, 릴리스 0, 수정 0과 모든 후속 릴리스 및 수정에 적용됩니다.

© Copyright IBM Corporation 2010, 2015.

목차

서문	v	데이터 미리보기	44
제 1 장 IBM Analytical Decision Management		데이터 개요	44
정보	1	측정 수준	45
제품 개요	1	데이터 필드 값 편집	46
버전 17의 새로운 기능	1	필드 맵핑	46
이전 버전	2	필드 파생	47
제 2 장 고객 반응 모델링 예제	11	표현식 빌더	48
IBM SPSS Modeler Advantage	11	함수 선택	49
직접 메일 캠페인의 수익 최대화	11	필드 및 값 선택	50
샘플 파일	12	표현식 확인	50
애플리케이션 시작	13	제 5 장 규칙 작업	53
데이터 소스 설정	14	규칙 작성	53
모델 빌드	15	세그먼트 규칙 정의	53
모델 평가	16	선택 규칙 정의	55
예측 규칙 모델링	17	내보내기 및 재사용 규칙	56
모델 테스트	18	주석 추가	58
모델 스코어링	18	IBM Operational Decision Management의 규칙 사 용	58
군집 모델링 예제	19	제 6 장 예측 모델 작성	61
연관 모델링 예제	20	예측 모델 정보	61
제 3 장 작업공간 탐색	25	모델 작성	61
표시된 언어 변경	25	예측 모델 작성	61
애플리케이션 시작	25	예측 모델링 결과	62
홈 페이지	27	예측 규칙 모델 작성	63
갤러리	27	군집 모형 작성	65
환경 설정	27	연관 모델 작성	66
프로젝트, 모델, 규칙 열기 및 저장	29	선택적 모델 설정	68
리포지토리에서 열기 및 저장	30	모델 평가	72
레이블 작업	31	최대 수익 시뮬레이션	73
프로젝트 잠금	31	모델 테스트	74
특성	32	애플리케이션에서 모델 스코어 사용	75
제 4 장 데이터 소스 관리	33	제 7 장 스코어링 모형	77
새 데이터 소스 작성	35	데이터베이스 테이블, 파일, Analytic Server 또는 IBM Cognos BI 서버에 모델 스코어링	77
파일 소스	36	데이터 및 서브셋 선택	78
데이터베이스 소스	39	스코어링될 레코드 선택	78
엔터프라이즈 뷰 소스	40	스코어링 대상 선택	78
데이터 보기 소스	40	결과 필드 선택	82
IBM Cognos BI 소스	41	모델 스코어링	83
Analytic Server 소스	43	스코어링 요약 보고서	84
입력 필드 선택	44		

제 8 장 애플리케이션 관리	85	부록 A. 스트림 공유	99
일반 옵션	86	부록 B. 내게 필요한 옵션	101
애플리케이션 단축키 제어.	87	주의사항	103
홈 페이지 숨기기	87	상표	105
데이터 옵션 잠금	87	색인	107
스코어링 대상 옵션 잠금.	88		
외부 규칙 작성에 사용하기 위한 프로젝트 메타데이 터 다운로드	88		
XML 스키마 정의 편집	90		
제 9 장 도움말 가져오기.	97		
사용 가능한 문서	97		

서문

IBM Business Analytics 정보

IBM Business Analytics 소프트웨어는 의사결정자가 비즈니스 성과를 개선하기 위해 신뢰하는 완전하고 일관되며 정확한 정보를 제공합니다. 비즈니스 지능, 예측 분석, 금융 성과와 전략 관리 및 분석 애플리케이션의 종합 포트폴리오는 현재의 성과에 대한 분명하고 즉각적이며 실행 가능한 통찰을, 그리고 미래의 결과를 예측하는 능력을 제공합니다. 풍부한 업계 솔루션, 검증된 절차 및 전문 서비스와 결합되어, 모든 크기의 조직이 최대의 생산성을 발휘하고, 자신있게 의사결정을 자동화하며 더 나은 결과를 낼 수 있습니다.

이 포트폴리오의 일부인 IBM SPSS Predictive Analytics 소프트웨어는 조직이 미래의 사건을 예측하고 이 통찰을 기반으로 더 나은 비즈니스 결과를 얻기 위해 미리 대처하는 데 도움을 줍니다. 전 세계의 상업, 정부 및 학술 분야 고객들은 IBM® SPSS® 기술을 사기를 줄이고 위험을 낮추는 한편 고객을 끌어들이고, 유지하고, 성장시키는 데 대한 경쟁력으로 삼고 있습니다. 조직은 IBM SPSS 소프트웨어를 일일 작업과 결합하여 비즈니스 목적을 만족시킬 수 있도록 의사결정을 이끌고 자동화하며 적절한 경쟁력을 갖추으로써, 예측력을 갖춘 기업으로 성장할 수 있습니다. 담당자에게 문의하기 위한 자세한 정보를 얻으려면 <http://www.ibm.com/spss> 를 방문하십시오.

기술 지원

유지보수 고객은 기술 지원을 사용할 수 있습니다. IBM Corp. 제품 사용이나 지원되는 하드웨어 환경 중 하나에 대한 설치 도움말이 필요한 경우에는 기술 지원에 문의하십시오. 기술 지원에 문의하려면 <http://www.ibm.com/support>의 IBM Corp. 웹 사이트를 참조하십시오. 지원을 요청하려면 본인의 신상, 소속 조직 및 지원 계약을 제시해야 합니다.

제 1 장 IBM Analytical Decision Management 정보

제품 개요

IBM Analytical Decision Management는 예측 분석의 이점을 실제 비즈니스 문제점에 도입하여 사용자의 고객 또는 업계에 특화된 사용자 정의 애플리케이션을 빌드할 수 있게 해 줍니다. 애플리케이션은 보통 특정한 문제를 해결하기 위해 구성되지만 모든 애플리케이션은 다음과 같은 공통적인 기능 세트를 기반으로 하고 있습니다.

- 비즈니스 규칙을 사용하여 의사결정을 자동화합니다.
- 예측 모델을 사용하여 통찰력을 얻습니다.
- 우선순위, 최적화 또는 시뮬레이션을 사용하여 이에 기반한 최선의 의사결정을 내립니다.

특정 비즈니스 문제점을 해결하는 데 특화된 몇 가지 패키징된 애플리케이션을 사용할 수 있습니다. 자세한 정보는 영업 담당자에게 문의하십시오.

버전 17의 새로운 기능

IBM Analytical Decision Management 17(17.0.0.0)에서는 다음과 같은 새 기능이 추가되었습니다.

다음은 새 주요 기능에 대한 설명입니다.

표현식 빌더. 더 복잡하고 강력한 표현식을 허용하도록 표현식 빌더가 개선되었습니다. 이제 IBM Analytical Decision Management의 표현식 빌더는 IBM SPSS Modeler 클라이언트의 표현식 빌더와 더 잘 일치합니다. http://www-01.ibm.com/support/knowledgecenter/SS6A3P_17.0.0/configurableapps/dm_expressions.dita의 내용을 참조하십시오.

파생 탭. 대부분의 IBM Analytical Decision Management 애플리케이션에서 표현식, 모델 결과, 세그먼트 규칙에서 작성된 필드 또는 보조 테이블의 필드와 같은 추가 필드를 프로젝트 데이터 모델에 추가할 수 있습니다. 데이터 탭의 새 파생 하위 탭은 데이터의 일부로 포함할 필드를 표시합니다. http://www-01.ibm.com/support/knowledgecenter/SS6A3P_17.0.0/configurableapps/dm_app_derived.dita의 내용을 참조하십시오.

보조 테이블 섹션. 새 보조 테이블 섹션이 대부분의 IBM Analytical Decision Management 애플리케이션에 추가되었습니다. 이는 이전 릴리스에 있던 다양한 테이블 추가/편집 링크를 대체합니다. http://www-01.ibm.com/support/knowledgecenter/SS6A3P_17.0.0/configurableapps/dm_app_additional_fields.dita의 내용을 참조하십시오.

MFA(Multi-Factor Authentication). MFA(Multi-Factor Authentication)를 지원하는 기능이 IBM SPSS Collaboration and Deployment Services 7, IBM SPSS Modeler 17 및 IBM Analytical Decision Management 17에 추가되었습니다. 이러한 개선사항은 대형 엔터프라이즈 환경을 위한 보안을 향상시킬 수 있습니다. 예를 들어, SPSS 제품은 회사의 기존 Kerberos 싱글 사인온 인프라 및 SmartCard 인프라와 통합될 수 있습니다

(예를 들어, 회사 직원이 시스템 및 애플리케이션 인증을 위해 보안 배지 및 카드 리더기를 사용할 경우).
http://www-01.ibm.com/support/knowledgecenter/SS6A3P_17.0.0/configurableapps/dms_install_security.dita의 내용을 참조하십시오.

데이터 소스 맵핑. 가능한 경우 이제 IBM Analytical Decision Management에서는 사용자가 데이터 소스를 추가할 때 호환 가능한 스토리지 유형을 자동으로 맵핑합니다. 호환 가능하지 않은 유형인 경우 사용자에게 유형 변환이 허용됩니다.

파일 스토리지 유형. 사용자는 이제 값을 정확하게 읽도록 고정 및 변수 형식 데이터 파일에서 읽을 경우 스토리지 유형 및 형식을 설정할 수 있습니다. 예를 들어, IBM Analytical Decision Management에서 새 데이터 소스를 추가할 때 데이터 소스 유형이 파일이고 파일 형식이 일반 텍스트의 변수 길이 필드인 경우 데이터 소스 대화 상자에는 이제 사용자가 각 필드의 형식 및 입력 유형을 지정할 수 있는 새 데이터 형식 지정 섹션이 있습니다. 이는 IBM SPSS Modeler 클라이언트에서 사용 가능한 기능과 유사합니다.

프로젝트 요약. 배포 탭이 있는 애플리케이션의 경우, 새 프로젝트 요약 섹션이 추가되었습니다. 이 섹션을 사용하여 프로젝트의 유효성을 검증하고 배포에 영향을 줄 수 있는 발견된 오류를 포함하여 중요한 프로젝트 요약 정보를 표시하십시오. 모든 속성 참조 및 모든 리포지토리 참조 같은 프로젝트 요약 표시 정보는 *.csv 파일로 내보낼 수 있습니다. 또한 애플리케이션의 다른 탭으로 이동하여 오류를 수정한 후 배포 탭의 프로젝트 요약으로 돌아갈 수도 있습니다.

모델 참조. IBM Analytical Decision Management 애플리케이션 내에서 모델을 참조할 경우, 모델 링크의 동작이 변경되었습니다. 모델 참조 링크를 클릭하면 모델에 대한 세부사항을 볼 수 있는 모델 뷰어 대화 상자가 열립니다. 또는 이제 모델 참조를 마우스 오른쪽 단추로 클릭하여 참조 또는 빌드 옵션을 선택할 수 있습니다. 규칙 참조를 클릭하면 규칙 뷰어가 열립니다(이전 릴리스에서와 같이). 또는 규칙 참조를 마우스 오른쪽 단추로 클릭하고 참조를 선택할 수 있습니다.

Analytic Server 신임 정보 접미문자. 관리자는 브라우저 기반 IBM SPSS Collaboration and Deployment Services Deployment Manager를 사용하여 Analytic Server 신임 정보에 붙일 접미문자를 구성할 수 있습니다. 이는 IBM Analytical Decision Management에서 Analytic Server 연결에 사용할 신임 정보의 이름을 파생합니다.

IBM Cognos 신임 정보. IBM Analytical Decision Management에서 사용하기 위해 IBM Cognos BI 데이터베이스 데이터 또는 단일 목록 보고서를 가져올 경우, 이제 사용자 이름 및 비밀번호 정보를 입력하거나 익명 연결을 사용하는 대신 IBM SPSS Collaboration and Deployment Services Repository에 저장된 IBM Cognos 신임 정보를 선택할 수 있습니다.

이전 버전

버전 8

IBM SPSS Analytical Decision Management 8은 사용자에게 이전보다 나은 이해력과 제어를 통한 폭넓은 결정 능력을 지원하여 일상의 비즈니스 프로세스에 분석을 적용할 수 있도록 합니다.

- 개선된 시뮬레이션 기능을 통해 비즈니스 결과에 대한 규칙, 모델 및 최적화의 결합된 영향을 조정하고 시각화합니다.
- 가격 최적화 및 분류 최적화 등과 같이 사전 빌드된 새 애플리케이션 및 새로운 최적화 기법으로 다양한 문제점을 해결합니다.
- IBM Operational Decision Management와의 통합이 개선되고 IBM SPSS Analytical Server 통합을 통해 대량 데이터 지원이 가능합니다.

다음은 버전 8에 추가된 새 주요 기능에 대한 설명입니다.

새 설치 기술. 제품 설치가 간소화되고 개선되었습니다. 제품은 이제 Installation Manager 기술을 사용합니다. 자세한 정보는 *IBM SPSS Analytical Decision Management 설치 안내서*를 참조하십시오.

사용자 인터페이스 개선. 사용성을 개선하기 위해 사용자 인터페이스가 많이 변경되었습니다.

IBM Analytical Decision Management for Predictive Maintenance. IBM SPSS Analytical Decision Management for Predictive Maintenance는 장비 장애가 발생하기 전에 최적화를 사용하여 이를 예측하고 방지하는 새로운 사전 빌드된 애플리케이션입니다. 이를 통해 장애를 예측하고 올바른 시간에 올바른 정정 조치를 스케줄링함으로써 비용 소모적인 가동 중지를 최소화하고 귀중한 인프라와 자원의 사용을 최적화할 수 있습니다. 또한 맨 먼저 장애가 발생할 소지가 있는 부분을 예측하고 적절한 시기에 이를 교체할 수 있습니다. 시스템의 장애를 예측하게 되면 수익을 최대화하고 시스템 장애 및 작동 중지로 인한 손실을 막는 데 도움이 됩니다.

IBM Analytical Decision Management for Demand Optimization. IBM SPSS Analytical Decision Management for Demand Optimization은 적절한 양의 각 제품을 각 상점에 할당함으로써 최적화를 사용하여 비용이 많이 드는 재고 과잉과 부족을 방지하는 새로운 사전 빌드된 애플리케이션입니다. 이를 통해 각 아울렛에 올바른 양의 재고 명세를 할당하여 수익을 극대화하고 재고 명세 부족분과 환불수수료로 인한 손실을 막을 수 있습니다.

시뮬레이션 개선. 새 시뮬레이션 기능은 이전 시뮬레이션 단추(이전에는 정의 탭에 있었음)의 기능을 이전의 가정 기능(이전에는 최적화, 우선순위 또는 결합 탭에 있었음)과 단일 기능으로 결합한 기능입니다. 새 시뮬레이션 기능 및 개선된 시뮬레이션 기능은 사용자 인터페이스의 오른쪽 상단에 자체 소유의 도구 모음을 가지며 시나리오를 프로젝트와 함께 저장하고 세션 간에도 유지되도록 해줍니다.

배포 및 스코어링. 배포/스코어 탭이 여러모로 개선되었습니다.

애플리케이션 URL. IBM SPSS Collaboration and Deployment Services에서는 IBM SPSS Analytical Decision Management를 실행할 수 있는 URL을 포함하여 웹 기반 애플리케이션에 대한 URL을 사용자 정의하는 기능이 추가되었습니다. 자세한 정보는 관리자에게 문의하십시오. *IBM SPSS Collaboration and Deployment Services 리포지토리 설치 및 구성 안내서*의 애플리케이션 컨텍스트 루트 장에 추가 정보가 있습니다.

통계 파일 개선. IBM SPSS Analytical Decision Management에서는 이제 비밀번호로 보호된 IBM SPSS Statistics 압축 데이터 파일(.sav 및 .zsav)이 지원됩니다.

트리 뷰어 개선. IBM SPSS Modeler Advantage에서 개별 모델 결과를 볼 때 각 모델의 이름은 대상 필드 및 입력 필드와 같이 모델에 대한 정보가 포함된 새 대화 상자가 열리는 링크입니다. 선택된 모델이 트리 모델 인 경우에는 트리 뷰어도 표시됩니다. 트리 뷰어가 개선되었습니다. 이제 트리 확대 및 축소, 트리 방향 변경, 큰 트리 탐색, 트리 내에서 개별 노드 펼치기 또는 접기 등의 조치를 수행할 수 있습니다.

최적화 탭의 고급 옵션. IBM SPSS Analytical Decision Management for Campaign Optimization, IBM SPSS Analytical Decision Management for Predictive Maintenance 및 IBM SPSS Analytical Decision Management for Demand Optimization 등과 같이 CPLEX(Complex Mathematical Optimization)를 사용하는 애플리케이션의 경우 이제 CPLEX 최적화 알고리즘과 관련된 고급 옵션을 사용할 수 있습니다. 이제 OPL(Optimization Programming Language) 모델 파일 및 사용자 정의 설정을 내보낼 수 있습니다.

데이터 보기 소스. IBM SPSS Collaboration and Deployment Services 분석 데이터 보기에 정의된 스트림의 데이터를 포함하도록 새 데이터 보기 노드가 IBM SPSS Modeler 16에 추가되었습니다. 새 데이터 보기 데이터 소스는 IBM SPSS Analytical Decision Management의 데이터 탭에서도 사용할 수 있습니다. IBM Operational Decision Management 규칙과 함께 사용할 연결을 작성할 수 있습니다. 계층 구조 데이터 모델을 사용하는 복합 규칙의 경우, 관리자는 성능이 저하되지 않도록 해당 규칙의 사용을 제한할 수 있습니다.

버전 8 이전에는 외부 규칙에 IBM SPSS Analytical Decision Management 프로젝트 데이터 모델과 일치하는 플랫폼 입력 매개변수가 필요했습니다. 그러나 이제 외부 규칙은 이 새 분석 데이터 보기 지원을 통해 계층 구조 입력 매개변수를 가질 수 있습니다. 자세한 정보는 40 페이지의 『데이터 보기 소스』 및 87 페이지의 『데이터 옵션 잠금』 주제와 애플리케이션 디자이너 안내서의 5장을 참조하십시오.

IBM SPSS Analytic Server 통합 IBM SPSS Modeler 16에서, IBM SPSS Analytic Server이 소개되어 데이터 소스로 HDFS(Hadoop Distributed File System)의 사용을 허용합니다. IBM SPSS Modeler 사용자는 Hadoop 대형 데이터 시스템에 기반하여 HDFS Analytic Server 소스 노드 및 빌드 모델을 사용하여 스트림을 작성할 수 있습니다. IBM SPSS Analytical Decision Management 사용자는 이러한 모델 스트림을 보기 위해 IBM SPSS Modeler Advantage에서 열 수 있습니다(Analytic Server 데이터 소스를 통해 새 프로젝트를 작성할 수 없음).

애플리케이션 디자이너 안내서의 새 장. *IBM SPSS Analytical Decision Management 애플리케이션 디자이너 안내서*에 새 장이 추가되었습니다. 여기에서는 사용자 정의 최적화 애플리케이션의 구성 예를 제공합니다. 표준 IBM SPSS Analytical Decision Management for Campaign Optimization 애플리케이션은 사전 구성된 OPL(Optimization Programming Language) 모델 파일(*CampaignOptimization.mod*)을 사용합니다. 어떤 경우에는 표준 Campaign Optimization 애플리케이션 및 해당 OPL 모델이 비즈니스 문제점의 목적에 부합하지 않을 수 있습니다. 새 장에서는 가상 회사에 대해 IBM ILOG CPLEX Optimization Studio 및 사용자 정의 IBM SPSS Analytical Decision Management 최적화 애플리케이션을 사용하여 빌드된 사용자 정의 OPL 모델의 예를 설명합니다.

표 형식 평가 결과. 새 표 형식 결과 포함 옵션은 평가 실행 시에 사용할 수 있습니다. 표 형식 결과를 포함하는 평가를 실행하는 데 많은 시간이 소요되기 때문에 이 옵션을 선택 해제하면 성능이 개선됩니다.

명목 필드의 최대 수에 대한 새 설정. 명목 데이터 필드에 대해 허용된 최대 멤버 수를 설정할 수 있는 새 설정이 특성 대화 상자에 추가되었습니다. 이는 데이터 탭에서 사용됩니다. 자세한 정보는 32 페이지의 『특성』의 내용을 참조하십시오.

싱글 사인온. ODBC 연결은 이제 IBM SPSS Modeler Server에 대한 싱글 사인온(SSO) 및 IBM SPSS Analytical Decision Management와 ODBC 데이터베이스 연결을 사용할 수 있습니다. 일부 구성 단계가 필요합니다. 설치 문서의 보안 고려사항 주제를 참조하십시오.

SQL 푸시백. 데이터베이스에서 소스 데이터가 시작될 경우 가능한 한 언제든지 실행이 데이터베이스로 푸시백됩니다("SQL 푸시백"이라는 프로세스에서). 일반적으로 IBM SPSS Analytical Decision Management는 IBM SPSS Modeler와 동일한 레벨의 SQL 푸시백을 지원하지만 다음 제한사항이 있습니다.

- 프로젝트에 모델 또는 외부 규칙에 대한 참조를 가지는 선택 규칙 또는 추가 필드가 포함되어 있는 경우 이러한 참조는 데이터베이스로 푸시백되지 않으며, 이는 전체적으로 프로젝트 스트림에 대해 달성할 수 있는 푸시백의 차수를 제한할 수 있습니다.
- IBM SPSS Modeler Advantage 이외의 애플리케이션에서는 선택 규칙에 대해서만 푸시백을 사용할 수 있습니다(첫 번째 차원에서의 멤버 선택 및 글로벌 선택). 프로젝트 데이터 모델에서 엔티티 ID로 정수 필드를 선택할 경우 푸시백의 차수가 늘어날 수 있습니다.

버전 7.0.0.1

IBM SPSS Decision Management 7 수정팩 1(7.0.0.1)에서는 다음과 같은 기능이 추가되었습니다.

데이터 개요. 모든 애플리케이션의 데이터 탭, IBM SPSS Modeler Advantage의 모델링 탭과 스코어 탭, 그리고 지금 스코어링 기능을 포함하는 애플리케이션의 배포 탭에서 새 데이터 개요 기능을 사용할 수 있습니다. 자세한 정보는 44 페이지의 『데이터 개요』 주제를 참조하십시오. 이 기능은 IBM SPSS Modeler Server 15 수정팩 1(15.0.0.1)에서만 사용 가능합니다. 버전 15에서는 사용할 수 없습니다.

스코어링 요약 보고서. 프로젝트를 스코어링하면 새 스코어링 요약 보고서 섹션에서 데이터가 스코어링된 위치, 파일 크기(적용 가능한 경우), 데이터베이스 및 테이블 이름(적용 가능한 경우), 모델 이름 또는 프로젝트 이름 스코어링된 레코드의 수, 그리고 어떤 경우에는 모델을 요약하는 차트 및/또는 테이블 등의 세부사항을 제공합니다. 스코어링 요약 보고서 섹션은 IBM SPSS Modeler Advantage와 지금 스코어링 기능을 포함하는 애플리케이션의 배포 탭에서 사용 가능합니다. 파일 크기는 IBM SPSS Modeler Server 15 수정팩 1(15.0.0.1)에서 실행될 경우에만 표시된다는 점을 참고하십시오. 버전 15에서는 사용할 수 없습니다.

스코어링된 파일의 로컬 사본 다운로드. 스코어링 결과를 플랫폼 파일에 전송하도록 선택하는 경우에는 필요한 경우 이 파일의 로컬 사본을 다운로드할 수도 있습니다. 관리자는 브라우저 기반 IBM SPSS Collaboration and Deployment Services Deployment Manager를 사용하여 이 기능을 사용으로, 또는 사용 안함으로 설정하거나 허용된 최대 파일 크기 한계를 늘리고 줄일 수 있습니다. 파일 다운로드 단추가 비활성 상태인 경우에는 파일 크기가 관리자가 설정한 최대 파일 한계를 초과하거나 관리자가 이 기능을 사용 안함으로 설정한 것입니다. 이 기능은 기본적으로 사용 안함으로 설정되어 있습니다. 이 기능은 IBM SPSS Modeler Server 15 수정팩 1(15.0.0.1)에서만 사용 가능합니다. 버전 15에서는 사용할 수 없습니다.

모델 작성에 대한 새 선택적 설정. IBM SPSS Modeler Advantage에서 모델을 작성할 때, 모델 작성 시 포함시킬 모델링 기법(알고리즘)을 선택하고, 보유하고 결합할 수 있는 모델의 최대 수를 지정하고, 모델을 작성하기 전에 수익 기준을 설정하는 데이터 파티셔닝 옵션과 같은 새 선택적 설정을 사용할 수 있습니다. 사용 가능한 옵션은 모델의 유형과 선택된 대상에 따라 달라집니다. 자세한 정보는 68 페이지의 『선택적 모델 설정』 주제를 참조하십시오. 다음 새 기능은 IBM SPSS Modeler Server 15 수정팩 1(15.0.0.1)에서만 사용 가능하다는 점을 참고하십시오. 버전 15에서는 사용할 수 없습니다.

- 모델 작성 시 포함시킬 모델링 기법 선택
- 수익 기준 설정
- 개별 모델에 대한 높은 수준의 세부사항 보기
- 선택된 모델에 대한 추가 세부사항 보기

트리 뷰어. IBM SPSS Modeler Advantage에서 모델을 작성한 후 개별 모델 결과를 볼 때, 각 모델의 이름은 대상 필드나 입력 필드와 같은 모델에 대한 정보를 포함하고 있는 새 대화 상자를 여는 링크입니다. 선택된 모델이 판별 모형인 경우에는 트리 뷰어도 표시됩니다. 트리 뷰어는 IBM SPSS Modeler Server 15 수정팩 1(15.0.0.1)에서만 사용 가능하다는 점을 참고하십시오. 버전 15에서는 사용할 수 없습니다.

평가 개선. IBM SPSS Modeler Advantage의 평가 대화 상자가 여러모로 개선되었습니다. 예를 들어, 평가에 포함시킬 수 있도록 모델을 수정할 수 있게 해 주는 새로운 모델 선택 수정 단추를 사용할 수 있습니다. 선택한 후 원하는 경우에는 현재 모델 선택을 개별 모델 결과 테이블로 다시 복사할 수도 있습니다. 새 옵션 중 일부는 특정 모델 유형(대부분의 새 옵션은 예측 모델 유형에만 사용 가능하며 일부는 예측 규칙 모델 유형에 사용 가능)에만 사용할 수 있습니다. 평가 대화 상자의 다음 새 기능은 IBM SPSS Modeler Server 15 수정팩 1(15.0.0.1)에서만 사용할 수 있다는 점을 참고하십시오. 버전 15에서는 사용할 수 없습니다.

- 숫자 대상에 대한 성능 차트
- 숫자 대상에 대한 수익 시뮬레이션
- 일치 교차표 차트(모든 대상)
- 모든 차트에 대한 테이블 보기
- 모델 선택 수정 단추
- 성능 차트를 위한 개별 모델 평가 포함 옵션

Microsoft Windows 8 및 Internet Explorer 10. IBM SPSS Collaboration and Deployment Services 5 수정팩 1(5.0.0.1)에는 Windows 8 및 Internet Explorer 10에 대한 지원이 포함되어 있습니다. 이들은 IBM SPSS Collaboration and Deployment Services 4.2.1이나 5에서는 지원되지 않습니다.

사용자 인터페이스 개선. 사용성을 개선하기 위해 사용자 인터페이스가 약간 변경되었습니다.

버전 7

IBM SPSS Decision Management 7(7.0.0.0)에서는 다음과 같은 기능이 추가되었습니다.

최적화. 최적화의 목표는 마케팅 캠페인의 수입 최대화, 또는 사기나 이탈의 위험 최소화와 같은 특정 목표에 가장 부합하는 솔루션을 식별하는 것입니다. 최적화의 결과는 모집단 내의 모든 가능한 조합 중에서 최적화하고자 하는 값(예: 수익)을 최대화함으로써 이러한 질문의 해답을 주는 솔루션입니다.

군집 모형. IBM SPSS Modeler Advantage의 새 기능인 군집 모델링 방법은 유사한 값 또는 패턴이 있는 데이터 레코드의 그룹을 찾는 데 도움을 줍니다. 이러한 기술은 마케팅(고객 세분화) 및 기타 비즈니스 애플리케이션(단일 레코드 군집에 속하는 레코드는 오류를 포함하고 있거나 사기 인스턴스일 수 있음)에 사용됩니다. 군집 모형은 대상 필드를 사용하지 않으며 true 또는 false로 평가할 수 있는 측정 예측을 리턴하지 않으므로 종종 비감독 학습 모델이라고 하기도 합니다. IBM SPSS Decision Management를 통해 자동으로 클러스터 모델을 빌드할 수 있습니다. 또한 군집을 결합하거나, 이름을 바꾸거나, 수정하거나, 삭제할 수 있으며 사전 비즈니스 지식을 기반으로 사용자 정의 군집을 추가할 수도 있습니다.

연관 모델. IBM SPSS Modeler Advantage의 새 기능인 연관 모델링은 함께 나타날 가능성이 가장 높은 항목을 예측할 수 있게 해 주며 이들 사이에 있는 관계의 강도를 예측합니다. 연관 규칙은 특정 결과(예: 특정 제품의 구매)를 조건 세트(예: 다른 여러 제품의 구매)와 연관시킵니다.

데이터. 데이터 탭은 이제 여러 데이터 입력을 갖고, 애플리케이션 보기로 프로젝트 데이터 모델을 정의하고, 모델 결과를 데이터인 것처럼 참조하고, 로컬 데이터 파일을 업로드하고, 개수 데이터를 미리 보는 기능을 포함하고 있습니다.

IBM Cognos 데이터 통합 개선. IBM SPSS Modeler 15와 같이, IBM SPSS Decision Management는 이제 Cognos 사전 빌드 보고서 가져오기, Cognos 필터 및 계산, 자동 분해의 제어를 지원합니다.

여러 요소. 애플리케이션은 이제 두 개 이상의 차원 계층 구조를 포함할 수 있습니다. 애플리케이션을 시작할 때 선택된 애플리케이션이 둘 이상의 요소를 포함하도록 구성된 경우에는 애플리케이션에서 사용할 가능한 한 요소(계층 구조)를 하나 이상 선택하도록 프롬프트됩니다. 정의 탭의 차원 트리는 리턴될 수 있는 가능한 의사결정 또는 권장사항의 범위를 지정합니다. 애플리케이션이 둘 이상의 차원 계층 구조를 포함할 경우 이들은 정의 탭에 있는 차원 트리 영역의 별도의 탭에 포함됩니다.

동적 할당. 이전 릴리스에서는 무엇인가를 할당하려면 이 항목이 차원 트리에 있어야 했습니다. 이제는 하위가 없는 차원 항목의 경우 프로젝트 데이터 모델에 있는 키로 링크된 테이블의 항목을 읽을 수 있습니다. 예를 들어, 하위가 없으며 이름이 *Cars*인 차원이 있습니다. *BMW*, *Audi* 및 *Ford*와 같은 테이블의 필드를 참조하도록 *Cars*에 동적 할당을 설정할 수 있습니다.

다중 할당. 새 다중 할당 옵션은 각 세그먼트에 대해 레코드를 다중 차원 항목(예: 다중 오퍼)에 할당할 수 있게 해 줍니다. 다중 할당은 규칙당 단일 할당 대신 각 규칙에 대해 작성될 수 있습니다.

사용자 인터페이스 개선. 모든 애플리케이션의 사용성 및 확장성을 개선하기 위해 IBM SPSS Decision Management 사용자 인터페이스에 많은 변경과 추가가 이루어졌습니다.

갤러리 개선. 한 번에 두 개 이상의 갤러리 항목을 제거하는 갤러리 지우기 옵션과 같이, 갤러리가 여러모로 개선되었습니다.

규칙 편집기 개선. 규칙 편집기가 실행된 위치에 따라, 사용자는 이제 선택 규칙(포함 또는 제외)의 규칙 유형을 설정하고, 통합 규칙에 위험 점을 지정하고, 모든 할당 규칙에 대한 차원을 선택하고, *OR* 또는 *Split OR* 제어를 사용하여 규칙 내의 여러 표현식을 결합하고 *OR* 표현식을 작성할 수 있습니다. 이제 규칙을 작성할 때는 계층 구조 트리에서 차원을 선택하거나(둘 이상의 트리가 있는 애플리케이션의 경우) 표현식 편집기를 실행하여 표현식을 작성할 수도 있습니다. 이전에는 규칙 편집기 대화 상자에서 이 기능을 사용할 수 없었습니다.

확장된 IBM Operational Decision Management 통합 IBM SPSS Collaboration and Deployment Services Deployment Manager에서는 새 외부 규칙 추가 대화 상자를 사용하여 IBM SPSS Decision Management에서 사용될 수 있는 IBM Operational Decision Management 규칙 세트에 대한 참조를 정의할 수 있습니다. 이들 규칙은 이전에는 웹 서비스를 통해서만 액세스할 수 있었습니다. 이제는 이 릴리스를 사용하여 IBM Operational Decision Management 데이터베이스에서 로컬 규칙을 참조할 수 있습니다. 새 로컬 옵션의 경우 규칙은 IBM SPSS Modeler Server(IBM SPSS Decision Management에서 사용하도록 Deployment Manager에 정의된 것과 같은 서버)를 통해 실행됩니다.

WhatIf? 개선. WhatIf? 대화 상자의 기능이 여러모로 개선되었습니다. 예를 들어, 시뮬레이션 날짜의 경우에는 이제 WhatIf? 평가를 수행할 날짜 범위를 입력할 수 있습니다.

업로드 스트림 파일. 이제 로컬 파일 시스템에서 모든 IBM SPSS Modeler 스트림 파일을 열 수 있으며 원하는 경우에는 이를 리포지토리에 저장할 수 있습니다.

여러 IBM SPSS Modeler Server 신임 정보. IBM SPSS Collaboration and Deployment Services Deployment Manager에 여러 IBM SPSS Modeler Server 신임 정보가 정의되어 있는 경우에는 IBM SPSS Decision Management와 함께 사용할 신임 정보를 두 개 이상 지정할 수도 있습니다. 그런 다음, IBM SPSS Decision Management 사용자가 로그인하면 서버는 사용자에게 권한이 있는지 확인하기 위해 목록의 첫 번째 신임 정보를 시도하며, 적절한 권한이 있는 신임 정보를 찾을 때까지 순서대로 목록을 계속 시도합니다. 사용자에게 정렬된 목록에 있는 신임 정보에 대한 권한이 없는 경우 이 사용자는 IBM SPSS Decision Management를 사용할 수 없습니다. 이는 어느 사용자가 IBM SPSS Modeler Server 파일 시스템의 어느 부분에 대해 액세스 권한을 갖는지에 대해 더 나은 제어를 제공합니다.

버전 6.2

IBM SPSS Decision Management 6.2에서는 다음과 같은 기능이 추가되었습니다.

Cognos 데이터 지원. IBM SPSS Modeler Server 14.2 이상과 함께 사용하는 경우 IBM SPSS Decision Management 6.2에는 데이터 소스를 정의할 때 IBM Cognos Business Intelligence 서버에서 데이터를 읽고 스코어링 결과를 Cognos BI 서버에 다시 기록하는 기능이 추가됩니다. Cognos BI 서버와의 연결을 정의한 후에는 원하는 필드를 선택하기 위해 패키지, 차원, 쿼리 및 보고서를 찾아볼 수 있습니다. 자세한 정보는 41 페이지의 『IBM Cognos BI 소스』 주제를 참조하십시오.

ILOG 비즈니스 규칙 관리 시스템 통합. ILOG와 같은 비즈니스 규칙 관리 시스템에서 작성된 규칙은 현재 IBM SPSS Decision Management 프로젝트에서 사용된 것과 같은 데이터 모델을 지원하도록 개발된 경우 IBM SPSS Decision Management 애플리케이션에서 참조하고 사용할 수 있습니다. 이러한 외부 규칙은 IBM SPSS Decision Management 애플리케이션 내의 다른 모델 및 규칙과 같은 방식으로 선택하고 사용할 수 있

습니다. 자세한 정보는 58 페이지의 『IBM Operational Decision Management의 규칙 사용』 주제를 참조하십시오. ILOG는 이제 IBM Operational Decision Management로 명칭이 변경되었다는 점을 참고하십시오.

System Z용 Linux 지원. SUSE Enterprise Server 10이 64비트 IBM System z10에서 지원됩니다.

제 2 장 고객 반응 모델링 예제

IBM SPSS Modeler Advantage

비즈니스 사용자는 사용하기 쉬운 애플리케이션인 IBM SPSS Modeler Advantage를 사용하여 예측 모델링을 수행할 수 있습니다. 예측 모델을 사용하면 이전에 있었던 일로부터 패턴을 식별할 수 있으며 이를 사용하여 미래에 발생할 일을 예측할 수 있습니다.

예를 들어, 모델을 사용하여 수입, 나이, 가입한 조직이나 멤버십과 같은 특성을 기반으로 이탈할 가능성이 가장 낮은 고객, 또는 특정 오퍼에 반응할 가능성이 가장 높은 고객을 예측하려 할 수 있습니다. 예측 결과는 전략 계획에 대한 입력이 될 대상 고객의 목록이나 흥미로운 케이스를 생성하는 데 사용되거나 예측 애플리케이션의 컨텍스트에 있는 규칙과 통합될 수 있습니다.

직접 메일 캠페인의 수익 최대화

이 예측 모델 예제에서는 가상의 은행에서 고객에게 금융 상품 및 서비스를 마케팅합니다. 최근의 테스트 캠페인에서 다수의 고객에서 몇 가지 신규 서비스 중 하나를 제안하고 그들의 반응을 기록했습니다. 그러나 반응률이 기대했던 것보다 낮았으며 캠페인이 수익을 얻는 데 실패했습니다.

IBM SPSS Modeler Advantage를 통해, 예상 비용, 수입 및 기타 요인을 고려할 때 가장 적합한 고객을 대상으로 하여 그러한 손실을 수익으로 전환하려고 합니다.

다음은 테스트 캠페인 결과의 요약입니다.

표 1. 테스트 캠페인 결과

테스트 캠페인 결과	
메일 개수	3000
반응 개수	111
반응률	3.7%
비용	6750
수입	6660
총 수익(손실)	(90)

이 예제에서는 IBM Analytical Decision Management와 함께 분배 및 설치되는 *bank_customer_data.txt* 및 *bank_response_data.txt* 데이터 파일을 사용합니다. 이 예제의 완전한 버전도 *bank_response_model.str* 파일에서 제공됩니다. 필요한 경우 관리자에게 표본 파일 설치에 대한 세부사항을 문의하십시오. 자세한 정보는 12 페이지의 『샘플 파일』 주제를 참조하십시오.

샘플 파일

사용자 안내서에 나와 있는 예제를 복제하는 데 사용할 수 있는 데이터 및 기타 샘플 파일은 IBM Analytical Decision Management와 함께 포함되어 있습니다.

데이터

설치 중에는 몇 가지 샘플 데이터 파일이 IBM SPSS Modeler Server \Demos 디렉토리에 설치됩니다 (예: C:\Program Files\IBM\SPSS\ModelerServer\17\Demos). 이 데이터는 사전 빌드된 샘플 애플리케이션과 함께 사용할 수 있으며, 다음 항목을 포함합니다.

- *bank_customer_data.txt*
- *bank_response_data.txt*
- *customer_contact_history.txt*
- *customer_summary_data.txt*
- *insurance_claim_data.txt*
- *insurance_fraud_data.txt*
- *inventory_orders.txt*
- *inventory_stock.txt*
- *machine_log_data.txt*
- *machine_summary_data.txt*
- *retail_purchase_data.txt*

데이터 파일을 사용하는 샘플 애플리케이션 파일, 모델 및 규칙에서는 데이터 파일이 위에 표시된 기본 위치에 설치되었다고 가정합니다. 데이터 파일이 다른 드라이브 또는 위치에 설치된 경우에는 샘플 애플리케이션에서 적절히 데이터 소스 경로를 업데이트해야 합니다.

샘플

사전 빌드된 IBM Analytical Decision Management 애플리케이션에서 사용할 몇 가지 샘플 애플리케이션 파일, 모델 및 규칙이 제공되어 있습니다. 샘플은 설치 디스크의 \demo 디렉토리에 패키징되어 있거나, 전자 다운로드 가능한 파일에 제공되어 있습니다. 하나 이상의 사전 빌드된 애플리케이션용 샘플 파일을 IBM SPSS Collaboration and Deployment Services Repository로 가져오려면 다음 단계를 완료하십시오.

1. **DVD.** DVD의 \demo 폴더에 있는 *DecisionManagementExamples.pes* 파일을 임시 위치로 복사하십시오.

또는

다운로드한 파일, *ADM_17_Demo.zip* 파일을 임시 위치로 다운로드하고 압축을 푸십시오.
2. IBM SPSS Collaboration and Deployment Services Deployment Manager를 열고 콘텐츠 서버 연결에 로그인하십시오.

3. 콘텐츠 리포지토리 폴더를 마우스 오른쪽 단추로 클릭하고 가져오기를 선택하십시오.
4. 가져오기 대화 상자에서 찾아보기를 클릭하고 1단계에서 임시 위치에 복사 또는 추출한 *DecisionManagementExamples pes* 파일을 선택하십시오. 확인을 클릭하십시오.
5. 가져오기 충돌 대화 상자에서 확인을 클릭하십시오. 가져오기가 완료되면 IBM Analytical Decision Management에서 열 수 있는 예제 프로젝트가 포함된 다음과 같은 새 폴더를 볼 수 있습니다.

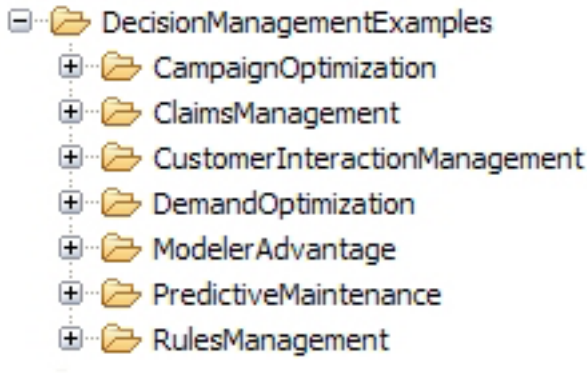


그림 1. 샘플 아티팩트의 폴더

모든 샘플 애플리케이션 파일, 모델 및 규칙은 데이터 파일이 IBM SPSS Modeler Server의 *Demos* 디렉토리(예: *C:\Program Files\IBM\SPSSModelerServer\17\Demos*)에 설치되어 있다고 가정합니다. 데이터 파일이 다른 위치에 설치된 경우에는 로컬 구성을 위해 샘플 애플리케이션에서 데이터 소스 경로를 업데이트해야 합니다.

애플리케이션 시작

IBM Analytical Decision Management 시작 페이지에는 사용자 사이트에서 사용 가능한 애플리케이션에 대한 링크가 있습니다. 이러한 링크는 관리자가 구성할 수도 있지만, 사용자가 오른쪽 상단에 있는 애플리케이션 추가 링크를 클릭하여 추가 애플리케이션에서 선택할 수도 있습니다.

1. 예제를 통해 작업을 시작하려면 IBM Analytical Decision Management 시작 페이지를 여십시오. 나중에 쉽게 액세스할 수 있도록 브라우저에서 이 페이지에 책갈피를 추가하거나 데스크탑에 단축키를 작성할 수 있습니다.
2. 새 모델을 작성하려면 IBM SPSS Modeler Advantage 그룹의 애플리케이션 드롭 다운 목록에서 새로 작성을 선택한 후 이동을 클릭하십시오. IBM SPSS Modeler Advantage 홈 페이지가 표시됩니다.



그림 2. IBM SPSS Modeler Advantage 홈 페이지

홈 페이지의 각 아이콘은 데이터 정의에서 모델 빌드 및 스코어링에 이르는 모델링 프로세스에서 서로 다른 각 단계를 나타냅니다. 시작하려면 왼쪽의 데이터 아이콘을 클릭하십시오.

데이터 소스 설정

데이터 탭은 예제에 사용된 표본 데이터 파일의 데이터 소스를 구성하는 데 사용됩니다. 데이터 세트는 관리자나 사용자가 직접 추가할 수 있습니다.

프로젝트 데이터 모델은 애플리케이션 전반에서 사용 가능한 필드를 정의하며, 규칙 및 모델을 평가하는 데 필요할 수 있습니다. 기타 모든 데이터 소스는 이 데이터 모델로 맵핑해야만 애플리케이션에서 사용할 수 있습니다.

이 예제에서는 애플리케이션과 함께 제공되는 두 개의 데이터 파일을 사용합니다.

- **bank_customer_data.txt**에는 나이 및 수입과 같은 인구 통계학적 정보와 함께 고객 ID, 지불 히스토리, 고객이 소유한 제품 수, 고객으로 등록된 기간(월) 등을 포함하여 현재 고객에 대한 정보가 포함됩니다. 프로젝트 데이터 모델은 이 파일을 기반으로 합니다.
- **bank_response_data.txt**에는 이전 캠페인에서 하나 이상의 서비스를 제공한 고객에 대한 정보가 포함되며 고객의 반응이 기록됩니다. 이 데이터를 분석함으로써 애플리케이션은 향후 캠페인에서 유사한 제안에 반응할 가능성이 가장 높은 고객을 식별할 수 있습니다.

기본적으로 표본 데이터 파일은 사용하는 애플리케이션에서 액세스 가능한 IBM SPSS Modeler Server Demos 폴더(예: *C:\Program Files\IBM\SPSS\ModelerServer\17\Demos*)에 설치됩니다. 필요한 경우 관리자에게 이러한 파일의 설치 지원을 요청하십시오.

프로젝트 데이터 모델을 정의하려면 다음을 수행하십시오.

1. 프로젝트 데이터 소스 패널의 데이터 소스 드롭 다운에서 데이터 소스 추가를 선택하십시오.
2. 은행 고객 데이터를 데이터 소스 이름으로 지정하십시오.
3. 파일을 선택하고 *bank_customer_data.txt*.
4. 입력 필드 지정에서 불량 지불 기록이 있음 필드의 측정 유형이 플래그로 설정되어 있는지 확인하십시오. 플래그 설정은 필드가 해당 속성의 존재 여부를 표시하는 *true* 및 *false*(일반적으로 *T* 및 *F*로 코딩되거나 숫자 필드에서 0 및 1로 코딩됨) 두 값을 가질 수 있음을 의미합니다.
5. 입력 필드 지정에서 성별 필드의 측정 유형이 명목으로 설정되어 있는지 확인하십시오. (이 필드의 값은 플래그와 같이 *true/false*로 표시하지 않고 단순히 남성 또는 여성으로 표시합니다.)

6. 링크를 클릭하십시오. 링크 대화 상자가 표시됩니다.
7. 저장을 클릭하십시오.

이 소스는 프로젝트에 대해 정의하는 첫 번째 데이터 소스이므로 프로젝트 데이터 모델에 자동으로 선택되며 애플리케이션에 필요한 필드를 정의하는 데 사용됩니다. 모든 다른 데이터 세트는 이 소스를 기준으로 맵핑됩니다.

반응 데이터 소스를 정의하려면 다음을 수행하십시오.

반응 데이터 세트에는 프로젝트 데이터 모델에서 사용되는 것과 동일한 다수의 필드와 함께 캠페인 ID 및 반응이라는 두 개의 추가 필드가 포함됩니다. 이러한 필드는 각 고객에게 제안된 내용과 고객의 반응 여부를 표시하며, 향후 반응할 가능성이 가장 높은 고객을 예측하는 모델을 빌드하는 데 사용될 수 있습니다.

1. 프로젝트 데이터 소스 패널의 데이터 소스 드롭 다운에서 데이터 소스 추가를 선택하십시오.
2. 은행 반응 데이터를 데이터 소스 이름으로 지정하십시오.
3. 파일을 선택하고 `bank_response_data.txt`를 선택하십시오.
4. 입력 필드 지정에서 반응 필드의 측정 유형이 플래그로 설정되어 있는지 확인하십시오.
5. 입력 필드 지정에서 캠페인 ID 필드 옆에 있는 선택란을 선택 취소하여 모델링 작업 시 이 필드를 입력으로 선택해서는 안 됨을 표시하십시오. (이 필드는 과거에 고객이 선택되었던 캠페인을 표시하는 데는 사용할 수 있지만 향후 반응을 예측하는 데는 유용하지 않습니다.)
6. 저장을 클릭하십시오.

데이터 모델에 있는 필드 중 일부가 반응 데이터 세트에는 없으므로, 이 데이터 소스의 경우 데이터 탭의 호환 가능 열에서 일부 필드는 일치하지만 일부 필드가 일치하지 않음을 나타내는 주황색 아이콘이 표시됩니다. 이 데이터 세트는 모델링에는 적합하지만 시뮬레이션, 스코어링 또는 배포에는 적합하지 않습니다. (이러한 목적을 위해서는 데이터 모델의 모든 필드가 존재해야 합니다.)

모델 빌드

모델링 탭을 사용하면 데이터 소스, 대상 필드 및 기타 옵션을 지정하여 모델을 빌드하고 평가할 수 있습니다.

1. 데이터 소스에 은행 반응 데이터를 선택하고 반응을 대상 필드로 선택하십시오.
2. 데이터 스캔이 필요함 메시지가 표시되면 아이콘을 클릭하여 데이터 스캔을 수행하십시오. 그런 다음 모델 빌드를 클릭하십시오.

모델이 완료되면 분포도가 표시됩니다. 분포도는 관측된 값과 긍정적 반응을 가리키는 1 값을 가로축에 표시합니다.

- 막대 높이는 전체 반응률을 반영합니다. 전체 적중률이 약 3.8%인 경우, 반응 데이터 세트의 3000명의 고객 중 111명이 반응했습니다.
- 각 막대의 색상 영역은 모델이 예측한 반응을 나타냅니다. 오른쪽 상단에 표시된 바와 같이, 빨간색은 긍정적 예측을 표시하고 파란색은 부정적 반응을 표시합니다. 전체 반응률이 낮으므로 예측 반응 수(빨간색 영

역)가 비교적 적습니다. 그러나 false 막대(반응 = 0)는 파란색이 압도적이며 이는 반응하지 않을 것으로 예측된 고객의 거의 대부분이 실제로 반응하지 않았음을 표시합니다.

모델이 수행되는 방법을 더 잘 이해하려면 다음 절(『모델 평가』)로 이동하십시오.

모델 평가

고객 데이터에 모델을 적용할 때 모델이 수행되는 방법을 알고 모델 사용이 최종 결과에 어떤 영향을 미치는지 이해하려면 모델링 탭에서 평가를 클릭하십시오.

1. 데이터 소스에 은행 반응 데이터를 선택하십시오.
2. 빌드할 그래프 유형에 **Gains** 차트 및 **Profit** 차트를 선택하십시오.
3. 최대 수익 시뮬레이션을 선택하고 인원 3000을 지정하십시오. 이는 테스트 캠페인의 인원 수와 일치합니다.
4. 비용에 2.25를 지정하고 수입에 60을 지정하십시오. 이 값은 각 제안에 드는 비용(예: 생산, 처리 및 메일 비용)과 긍정적인 각 반응에서 예상되는 수입을 표시합니다.
5. 찾는 반응에 1을 선택하십시오.
6. 실행을 클릭하십시오.

Gains 차트는 모델을 사용하여 얻을 수 있는 이익 또는 "리프트"를 표시합니다. 모델을 빌드할 때 데이터 자동 파티션 옵션을 선택했으므로(선택적 설정에서) 훈련 및 테스트 파티션 각각에 대한 개별 차트가 표시됩니다. 각 차트는 표본의 적중(긍정적 반응) 총계에 대한 각 백분위수에서의 적중 비율을 표시합니다. 곧은 대각선은 모델을 사용하지 않는 경우의 예상 반응을 표시합니다--적중 수의 1/2을 얻으려면 전체 인원의 1/2에 해당하는 사람들에게 접촉해야 합니다. 곡선은 모델을 사용할 경우의 이익을 표시합니다--반응할 가능성이 가장 높은 고객의 60%를 대상으로 함으로써 훨씬 적은 메일 전송으로 거의 모든 적중을 얻을 수 있습니다.

Profit 차트는 지정된 비용 및 수입 값을 기반으로 전체 수익에 있어서의 이익을 표시합니다. 모델 훈련에 사용되는 표본의 경우, 수익은 30번째 백분위수 바로 아래에서 최고에 이르는 것으로 표시됩니다.

이 차트의 작동 방법을 이해하려면 시뮬레이션 탭으로 이동하여 지정된 인원 수에 대해 예상되는 최대 수익을 확인하십시오. 예상 수익은 다음 공식으로 계산됩니다.

$$\text{expected profit} = (\text{probability to respond} * \text{revenue}) - \text{cost}$$

여기서 Probability to Respond는 모델을 기반으로 고객마다 판별됩니다.

모델에 의해 반응할 가능성이 가장 높은 고객으로 식별된 고객을 대상으로 함으로써 고객당 수익을 최대화할 수 있습니다. 반응률이 낮은 고객이 포함되면 그만큼 고객당 수익은 점점 줄어들지만, 더 많은 고객을 포함할수록 특정 지점까지는 총 수익이 증가합니다. 시뮬레이션은 최대 수익을 얻기 위해 어디서 선을 그을지 판별하는 데 도움이 됩니다.

수익 비교 테이블에서는 이 결과를 자세히 요약합니다.

- **대상 지정 44%.** 이 열은 모델에 의해 반응률이 가장 높다고 예측된 44%의 고객을 대상으로 지정했을 때의 결과를 표시합니다. 반응할 가능성이 가장 높은 1320명의 고객을 선택적으로 포함시켜 수익을 최대화함으로써 결과적으로 \$3270의 예상 수익을 얻을 수 있습니다.
- **무작위 44%.** 크기가 동일한 무작위 표본을 작성할 경우의 결과를 표시합니다. 이 경우에는 동일한 수의 제안(1320개)을 보내지만 모델에서 제공하는 추가 정보의 혜택을 받지 못하므로 예상 반응률은 더 낮습니다(테스트 캠페인의 반응률과 일치함). 전체 반응률이 각 제안에 든 메일 비용을 회수하기에 부족하므로 손실로 이어집니다.
- **모든 고객.** 표본 고객 3000명을 모두 포함할 경우의 결과를 표시하며, 테스트 캠페인의 결과와 정확히 일치합니다. 모델의 혜택을 받지 못하므로 이 경우에도 손실을 입게 됩니다.

결과를 보면, 손해를 보던 캠페인을 이익을 남기는 캠페인으로 바꾸기 위해 모델을 사용하여 반응률을 얼마나 향상시킬 수 있는지 알 수 있습니다. \$90.00 손실과 \$3277.50 수익의 차이는 예측 모델링을 통해 얻는 혜택을 보여줍니다. 이러한 결과는 테스트 캠페인의 3000명의 고객을 기반으로 하지만, 대상 인원 수를 늘리면 그만큼 이익도 증가합니다.

7. 시뮬레이션 결과를 검토한 후 예측 결과 차트에서 **스코어링에 대해 사용**을 선택하여 지정된 백분율(이 경우 44%를 기반으로 스코어 탭 설정을 업데이트하십시오. 자세한 정보는 18 페이지의 『모델 스코어링』 주제를 참조하십시오.
8. 프로젝트를 `bank_response_model.str`로 저장하십시오. 다음 절에서는 모델 유형을 변경하고 새 프로젝트를 작성합니다.

예측 규칙 모델링

예측 모델이 사용자가 필요로 하는 모든 예측력을 제공할 수도 있지만, 모델의 내부적 작업에 대한 제어력과 통찰력이 더 필요한 경우가 있습니다. 이러한 경우 예측 규칙(대화형) 모델링을 사용할 수 있습니다. 예측 규칙 모델을 사용하면 데이터를 세그먼트화하는 규칙을 자동으로 작성하거나 또는 선택한 필드의 값 집중이 높거나 낮은 세그먼트를 자동으로 찾을 수 있습니다. 예를 들어, 캠페인에 대해 긍정적으로 반응한 고객을 찾은 후 반응 확률이 더 높은 세그먼트를 식별합니다.

1. 모델링 탭으로 돌아가 **모델 변경**을 클릭하고 **예측 규칙 모델**을 선택하십시오.
2. 데이터 소스에 **은행 반응 데이터**를 선택하고 **반응**을 대상 필드로 선택하십시오.
3. 찾는 반응으로 **1**을 지정하여 긍정적인 반응을 보인 고객을 찾고자 함을 표시하십시오.
4. **빌드 시작**을 선택하십시오.
5. **모델 성장**을 선택하여 반응 확률이 높은 세그먼트를 식별하십시오. 반응 확률이 가장 높은 세그먼트를 식별하는 가장 우수한 세 개의 규칙이 작성되어 표시됩니다.
6. **포함/제외** 아이콘을 클릭하여 각 규칙에서 선택한 고객을 포함시킬지 또는 제외시킬지 여부를 선택하십시오. 예를 들어, 두 번째 및 세 번째 규칙 세그먼트는 8% 바로 위의 반응률을 표시하는데, 이 값은 전체 테스트 캠페인 반응률 3.7%보다 훨씬 높지만 첫 번째 세그먼트의 반응률 15.65%만큼 높지 않습니다. 이 두 규칙의 포함 또는 제외가 전체 수익에 어떠한 영향을 미치는지 확인하기 위해, 다음과 같이 두 방식으로 수익 시뮬레이션을 실행하고 결과를 비교할 수 있습니다.
7. 세 규칙 세그먼트를 모두 포함으로 설정하여 평가를 클릭하십시오.

8. 빌드할 그래프 유형에 **Gains** 차트 및 **Profit** 차트를 선택하십시오.
9. 최대 수익 시뮬레이션을 선택하고 인원에 **3000**, 비용에 **2.25**, 수입에 **60**을 지정하십시오. 이들은 예측 모델을 평가할 때 사용한 스펙과 동일합니다. 찾는 반응이 자동으로 완료됩니다.
10. 실행을 클릭하고 그래프가 빌드되면 시뮬레이션 탭으로 이동하십시오.

이 모델의 시뮬레이션된 최대 수익은 이 케이스의 자동화된 모델의 수익보다 조금 낮은 \$3178.90입니다. 선택적으로, 평가 대화 상자를 닫고 두 번째 및 세 번째 규칙 세그먼트에 대한 설정을 제외로 변경한 후 수익 시뮬레이션을 반복하여 결과가 어떻게 변경되는지 확인할 수 있습니다. 예측 규칙(대화형) 모델링의 실질적인 혜택은 제어력과 통찰력이 향상된다는 점입니다.

11. 프로젝트를 `bank_predictive_rule_model.str`로 저장하십시오.

모델 테스트

모델을 빌드했으면 선택한 고객 레코드를 테스트하여 고객이 캠페인에 어떻게 반응할지 알 수 있습니다.

1. 모델 스코어링에 의해 특정 레코드가 어떻게 처리되는지 미리보려면 모델링 탭에서 테스트를 선택하십시오. 테스트 대화 상자가 표시됩니다.
2. 레코드 영역에서 리턴 수치를 100으로 늘리십시오.
3. 표시할 필드 선택 영역에서 고객 ID를 선택하십시오. 또한 *\$XF-Response* 및 *\$XFRP-Response* 필드가 포함됩니다. 이 필드는 각 고객의 예측 반응(이 케이스에서는 아니오인 경우 0, 예인 경우 1) 및 성향 스코어를 표시합니다.

선택적으로, 지정된 고객이 모델에 의해 특정 방식으로 스코어링되는 이유나 방법을 이해하는 데 도움이 될 수 있는 나이, 마지막 제안 후 경과 시간(주) 및 불량 지불 기록이 있음 등의 추가 필드를 포함시킬 수 있습니다.

모델 스코어링

모델을 스코어링하는 것은 관심이 있는 데이터 소스에 모델을 적용하는 것을 의미합니다. 예를 들어, 반응할 가능성이 가장 높은 고객을 예측하는 모델을 생성했으면 이제 예상 고객을 스코어링하여 캠페인에 포함시킬 고객을 식별할 수 있습니다. 평가 대화 상자에서 수익을 최대화하기 위해 스코어링해야 하는 레코드의 백분율을 판별했으며 이 결과를 기반으로 스코어링할 레코드의 백분율을 표시했습니다.

1. 스코어 탭으로 이동하십시오.
2. 스코어링 데이터 소스의 경우, 은행 반응 데이터가 선택되어 있는지 확인하십시오.
3. 스코어링 결과의 대상을 지정하십시오. (단지 스코어를 미리보려는 경우에는 이 단계를 건너뛰어도 됩니다.)
4. 결과 필드에서 고객 ID 필드가 선택되었는지 확인하십시오.
5. 스코어 설정은 모델 평가 중에 실행한 수익 시뮬레이션과 동일한 퍼센트의 레코드를 스코어링하도록 이미 구성되어 있어야 합니다. (시뮬레이션 결과를 볼 때 스코어링에 대해 사용 단추를 선택하지 않았으면 여기서 설정을 업데이트할 수 있습니다.)
6. 대화 상자 맨 아래에서 미리보기를 클릭하여 스코어링에서 얻게 될 결과의 표본을 표시하십시오.

7. 스코어링에서 사용할 데이터의 개요를 보려면 **데이터 개요**를 클릭하십시오. 자세한 정보는 44 페이지의 『데이터 개요』 주제를 참조하십시오.

성향을 기반으로 스코어링하도록 선택했으므로 반응할 성향이 가장 높은 고객만 포함됩니다. 고객 ID를 다시 고객 데이터베이스에 링크함으로써 이 결과를 기반으로 메일 목록을 생성할 수 있습니다.

8. 스코어를 클릭하십시오. 모델이 고객 데이터에 적용되고 대화 상자 맨 아래에 메시지가 표시되어 이를 확인합니다.

군집 모델링 예제

군집 모형을 사용하여 레코드를 유사한 특성을 갖는 그룹으로 분류할 수 있습니다. 이는 데이터에서 자연 그룹을 찾는 데 도움이 될 수 있습니다. 예를 들어 인구 통계학적 특성이나 구매 행동에 따라 고객을 세그먼트화할 수 있습니다.

이 예제에서는 제품과 함께 분배되고 설치되는 *bank_customer_data.txt* 데이터 파일을 사용합니다. 이 예제의 완전한 버전도 *bank_cluster_model.str* 파일에서 제공됩니다. 필요한 경우 관리자에게 표본 파일 설치에 대한 세부사항을 문의하십시오. 자세한 정보는 12 페이지의 『샘플 파일』 주제를 참조하십시오.

1. 애플리케이션 시작 페이지로 돌아가 IBM SPSS Modeler Advantage 프로젝트를 새로 작성하십시오.
2. 데이터 탭의 프로젝트 데이터 소스 패널에서 14 페이지의 『데이터 소스 설정』에 설명된 대로 **은행 고객 데이터**라는 새 데이터 소스를 추가하십시오.
3. 모델링 탭에서 **모델 변경**을 선택하고 **군집 모형**을 선택한 후 **저장**을 클릭하십시오.
4. 데이터 소스에 **은행 고객 데이터**를 선택하십시오.
5. **선택적 설정** 섹션을 펼치십시오. 군집 모형의 경우, **평가 필드**를 선택하거나 **찾을 군집의 범위**를 설정하는 추가 **자동 군집 옵션** 선택적 설정을 사용할 수 있습니다.

군집 모형을 작성하면 이면에서 다수의 다양한 군집 모형이 실제로 작성됩니다. 그런 다음 이 모형들은 특정 품질 척도로 비교되고 순위가 정해지며, 가장 좋은 모형이 자동 군집 모형에서 사용되도록 선택됩니다. 따라서 이 두 자동 군집 옵션 설정은 실제로 선택되는 모델에 대한 제어를 제공합니다. 여기서 평가 필드를 지정할 경우, 평가 필드의 값을 가장 잘 구별하는 모델이 선택됩니다. 그리고 원하는 범위가 설정되어 있으면, 지정된 범위 밖에 있는 다수의 군집을 찾는 모델은 버려집니다.

6. 이 예제의 경우, **검색할 군집 수에 대한 원하는 범위 설정** 옵션을 선택하고 기본 최소값 **3** 및 최대값 **15**를 사용하십시오. 그러면 군집 모형에 적절한 수의 군집이 포함됩니다.
7. 나머지 선택적 설정은 다른 모델링 유형과 동일합니다. 자동 군집 예제를 더 이해하기 쉽게 만드는 경우에만 신뢰할 수 있는 **모델 빌드**를 위해 **자동으로 데이터 정리 및 준비** 옵션을 선택 취소하십시오.
8. **선택적 설정** 패널을 접고 **군집** 패널을 펼치십시오.

군집 패널에는 수동 군집을 위한 섹션과 자동 군집을 위한 섹션이 있습니다. 수동 군집을 통해 데이터에 대한 지식을 기반으로 군집을 정의할 수 있습니다. 이 예제에서는 **고가치 고객**에 대한 수동 군집 및 **젊은 미혼 고객**에 대한 수동 군집을 다음과 같이 정의합니다.

9. 새 규칙 작성 아이콘을 클릭하십시오. 규칙의 이름을 *고가치 고객*으로 지정하고 다음 표현식을 추가한 후 확인을 클릭하십시오.

Months as a Customer > 12

Income > 65000

10. 다음 표현식을 사용하여 *젊은 미혼 고객*이라는 또하나의 새 규칙을 작성하십시오.

Age < 35

Marital Status = U

11. 수동 군집 섹션 아래에 두 개의 새 규칙이 나열됩니다. 이 수동 군집에 의해 캐치되는 고객 수를 확인하려면 *군집 레코드 개수*를 클릭하십시오. 그런 다음 수동 군집을 다시 펼쳐 개수 열에서 각 군집에 해당하는 고객 수를 확인하십시오. 67명의 *고가치 고객*이 발견되고 33명의 *젊은 미혼 고객*이 발견되었음을 알 수 있습니다.
12. *자동 군집 찾기*를 클릭하여 관심을 끌 수 있는 다른 군집을 자동으로 찾으십시오. 완료한 후에는 *자동 군집* 섹션을 펼치십시오. 세 개의 자동 군집이 발견되었음을 확인할 수 있습니다.
13. 자세히 보려면 *자동 군집 보기*를 클릭하십시오. 자동 군집 결과 뷰어가 표시됩니다.

모델 요약 페이지에 사용된 자동 군집 모형 알고리즘(이 예제에서는 *TwoStep*)에 대한 기본 정보, 검색할 자동 군집 수 및 군집 품질에 대한 기본 개념이 표시됩니다.

기타 차트도 사용 가능합니다. 각 차트의 다양한 영역에 마우스를 놓으면 그에 해당하는 자세한 내용을 볼 수 있으며 일부 차트(예: *예측변수중요도 차트*)와 상호작용할 수 있습니다.

군집 보기(마지막 차트)에서 군집을 비교하여 흥미있는 패턴을 찾을 수도 있습니다. 실제 구현에서는 자동 군집을 연구하여 보다 의미있는 이름을 지정하게 됩니다(군집 보기에서 각 군집 레이블을 클릭하여). 군집 레이블 이름을 업데이트한 후에는 기타 모든 차트와 모델링 탭에서 새 이름이 사용됩니다. 그런 다음 모델을 평가하고 테스트하여 모델이 마음에 들면 마지막으로 모델을 스코어링합니다.

14. 프로젝트를 *bank_cluster_model.str*로 저장하십시오.

군집 모델링에 대한 자세한 정보는 65 페이지의 『군집 모형 작성』의 내용을 참조하십시오.

연관 모델링 예제

연관 모델링은 예측 분석에서 중요한 기법입니다. 연관 모델을 사용하여 함께 표시될 가능성이 가장 높은 항목과 그러한 항목들 간의 관계 강도를 예측할 수 있습니다. 예를 들어, 고객의 현재 장비구비에 담긴 항목을 기반으로 고객이 다음으로 구매할 가능성이 높은 항목을 식별할 수 있습니다.

이 예제에서는 제품과 함께 분배되고 설치되는 *retail_purchase_data.txt* 데이터 파일을 사용합니다. 이 데이터는 TV, 컴퓨터 및 스마트폰 등의 전자제품 항목에 대한 구매 정보로 구성됩니다.

이 예제의 완전한 버전도 *retail_association_model.str* 파일에서 제공됩니다. 필요한 경우 관리자에게 표본 파일 설치에 대한 세부사항을 문의하십시오. 자세한 정보는 12 페이지의 『샘플 파일』 주제를 참조하십시오.

1. 애플리케이션 시작 페이지로 돌아가 IBM SPSS Modeler Advantage 프로젝트를 새로 작성하십시오.
2. 데이터 탭의 프로젝트 데이터 소스 패널에서 소매 구매 데이터라는 새 데이터 소스를 추가하십시오. 파일을 선택하고 *retail_purchase_data.txt*를 선택하십시오.
3. 모델링 탭에서 모델 변경을 선택하고 연관 모델을 선택한 후 저장을 클릭하십시오.
4. 데이터 소스로 소매 구매 데이터를 선택하십시오.
5. 데이터 형식으로 표 형식을 선택하십시오. 트랜잭션 데이터 세트는 포함된 항목만 표시하며 모든 항목이 하나의 열에 표시됩니다. 그러나 이 데이터 세트에서는, 항목 구매 여부를 나타내는 *true* 또는 *false* 플래그가 있는 개별 열에 각 항목이 표시됩니다.
6. 모델 작성을 클릭하십시오. 빌드에 실패했음을 알리고 임계값이 너무 높을 수 있음을 제시하는 오류가 수신됩니다. 오류를 닫으십시오.

연관 규칙을 찾을 때는 일반적으로 적용 가능하고 신뢰할 수 있으며 단순한 규칙을 찾는 것이 좋습니다.

7. 선택적 설정 섹션을 펼치십시오. 연관 규칙 옵션에서, 첫번째 임계값을 10에서 1로 변경하고 두 번째 임계값을 80에서 1로 낮추십시오. 일반적으로 매번 임계값을 달리하여 모델을 다시 빌드해 봄으로써 규칙 과다와 부족 사이의 균형을 맞춥니다.
8. 모델 작성을 클릭하십시오.

모델 결과에서 각 행은 개별 연관 규칙을 나타냅니다. 목록의 첫 번째 규칙에서는 스마트폰을 구매하는 고객은 대형 화면 TV를 구매할 가능성이 높음을 표시합니다. 범위(%) 열은 약 15%(또는 고객 6명 중 1명)가 스마트폰을 구매했음을 표시하고 신뢰도(%) 열은 그러한 고객 중 약 30%가 대형 화면 TV를 구매할 가능성이 있음을 표시합니다.

따라서 규칙 세트가 작더라도(이 예제에서는 239) 비즈니스 컨텍스트에서 이러한 규칙을 사용하여 고객이 다음에 구매할 제품을 예측하거나 추천할 수 있습니다.

9. 테스트를 클릭한 후 실행을 클릭하여 예측 또는 추천이 작성되는 방법을 확인하십시오.

이 케이스의 경우, 모델은 신뢰 수치가 가장 높은 규칙과 사용자가 이미 구매한 항목을 검토하여 이들을 기반으로 고객이 구매할 가능성이 높은 항목에 대한 몇 가지 추천을 작성했습니다.

10. 예를 들어, 첫 번째 레코드 옆의 미리보기 아이콘을 클릭하면 첫 번째 고객이 대형 화면 TV 및 스피커를 구매했지만 다른 제품은 구매하지 않았음을 알 수 있습니다. 미리보기를 닫으십시오.

이 고객에 대한 추천은 두 번째 열에 표시된 바와 같이 스탠다드 TV를 제안하는 것입니다. 세 번째 열의 값 0.3은 대형 화면 TV 및 스피커를 이미 구매한 고객이 스탠다드 TV도 구매할 가능성이 30%임을 의미합니다.

그러나 이것이 이상 항목인지 일반적으로 적용 가능한 예측 추세인지 어떻게 알 수 있을까요? 돌아가 규칙을 자세히 검토할 수 있도록 규칙 번호(이 예제에서는 115)를 적어 두십시오.

11. 테스트 대화 상자를 닫으십시오.

12. 모델 결과에서 규칙 번호 115를 찾을 때까지 페이지 이동하십시오. 결과 페이지가 많은 대형 모델인 경우에는 **ID**로 규칙 찾기를 클릭하여 규칙 번호를 입력할 수도 있습니다.
13. 결과 열 헤더 중 하나의 옆에 있는 화살표를 클릭하십시오.

열 > 범위(N)

범위(N) 열이 표시됩니다. 인스턴스 수를 기반으로 할 경우, 규칙 115는 세 개의 제품을 함께 구매한 58 명의 서로 다른 고객을 기반으로 하므로 이것이 이상 항목이 아님을 확실히 알 수 있습니다. 이로써 이 규칙을 향후 추천의 기초로 사용할 수 있음을 충분히 신뢰할 수 있습니다.

단일 규칙이 추천에 어떻게 사용될 수 있는지 살펴보았습니다. 그러나 IBM Analytical Decision Management 에서도 특정 목적과 더 밀접한 관련을 갖도록 모델을 사용자 정의하는 강력한 옵션을 제공합니다. 예를 들어, 대형 화면 TV 구매에 관심이 있는 고객을 예측하고자 할 수 있습니다.

14. 필터 적용을 클릭하십시오. 필터와 일치하는 규칙만 모델에 표시되도록 필터를 정의할 수 있는 대화 상자가 표시됩니다.
15. 필터 사용을 선택하고 항목에 대형 화면 TV를 선택하십시오.
16. 저장을 클릭하여 모델 결과로 돌아가십시오.

규칙 수가 239개에서 34개로 줄었습니다.

적절한 수의 고객을 기반으로 하는 규칙만 사용하고자 할 수도 있습니다. 이를 위해 10명 미만의 고객의 동작을 기반으로 하는 규칙을 제외시킬 수 있습니다.

17. 범위(N) 열 헤더 옆의 화살표를 클릭하여 오름차순 정렬을 선택하십시오.
18. 제외? 열에서 범위(N) 값이 10개 미만의 인스턴스인 모든 규칙을 선택하십시오.
19. 많은 수의 규칙이 필터링되고 제외되었지만 이 시점에서 이들은 아직 모델에 존재합니다. 모델을 스코어링할 때 제외되거나 필터링된 규칙이 사용되지 않도록 이들을 영구히 제거하려면 제외 및 필터링된 규칙 삭제를 클릭하고 확인을 클릭하십시오.

리턴되는 규칙 수가 20개로 줄고, 이들 모두 10개 이상의 인스턴스를 기반으로 하며 모두 대형 화면 TV를 예측합니다.

20. 이러한 규칙 중 몇 개가 적용되는지 보려면 테스트를 클릭한 후 실행을 클릭하십시오. 이제 몇 명의 추가 고객에게 대형 화면 TV가 추천되며 일부 고객은 다른 고객보다 신뢰수준이 높습니다.
21. 테스트 대화 상자를 닫으십시오.
22. 선택적 설정 섹션에서 스코어링 옵션 섹션을 펼쳐 연관 모델에 고유한 스코어링 옵션을 표시하십시오.

기본적으로 예측 최대 수는 3입니다. 이 수는 최고 신뢰 수치를 기반으로 하며, 그렇지 않을 경우 범위, 규칙 지원, 리프트(특정 항목의 구매 가능성 증가) 최고량 또는 배포성을 기반으로 규칙을 선택하도록 규칙 기준을 변경할 수 있습니다. 예측이 존재하지 않는지 확인을 선택하여 대형 화면 TV를 이미 구매한 고객에게는 대형 화면 TV가 제안되지 않게 할 수도 있습니다.

기본 스코어링 옵션은 대부분의 상황에 적합합니다. 그러나 몇몇 경우에는 이러한 옵션에서 제공하는 추가된 유연성이 유익할 수도 있습니다.

이 프로세스의 다음 단계는 스코어링 탭을 사용하여 모델을 스코어링하는 단계입니다. 자세한 정보는 77 페이지의 『데이터베이스 테이블, 파일, Analytic Server 또는 IBM Cognos BI 서버에 모델 스코어링』 주제를 참조하십시오.

23. 프로젝트를 *retail_association_model.str*로 저장하십시오.

연관 모델링에 대한 자세한 정보는 66 페이지의 『연관 모델 작성』의 내용을 참조하십시오.

제 3 장 작업공간 탐색

표시된 언어 변경

각 사용자는 서버를 다시 시작할 필요 없이 애플리케이션의 언어를 쉽게 변경할 수 있습니다. 예를 들어, 독일어, 영어, 일본어 사용자가 동일한 애플리케이션을 각자의 언어로 사용할 수 있습니다.

사용하는 브라우저의 도구 메뉴에서 인터넷 옵션 또는 옵션으로 이동하여 원하는 언어로 전환한 후 화면을 새로 고치십시오. IBM Analytical Decision Management가 원하는 언어로 제공되지 않으면 기본적으로 영어로 표시되거나 선택한 제 2 언어가 사용 가능한 경우 해당 언어로 표시됩니다.

IBM Analytical Decision Management 인터페이스가 다음 언어를 지원합니다. 이 목록은 릴리스간에 변경되고 문서가 사용자 인터페이스와 동일한 언어로 사용 가능하지 않습니다.

- 영어
- 독일어
- 스페인어
- 프랑스어
- 이탈리아어
- 일본어
- 한국어
- 포르투갈어(브라질)
- 러시아어
- 중국어(중국)
- 중국어(대만)

애플리케이션 시작

애플리케이션 시작 페이지는 설치된 애플리케이션에 대한 구성 가능한 링크를 제공합니다. 각 패널은 설치된 애플리케이션을 나타냅니다. 최근에 저장된 프로젝트, 모델 또는 규칙을 선택하려면 드롭 다운을 클릭하십시오. 특정 레이블된 버전을 가리키는 단축키를 하나 이상 추가할 수도 있습니다.

- 애플리케이션 시작 페이지에 액세스하려면 `http://hostname:port/DM`으로 이동하십시오. 여기서 `hostname`은 리포지토리 시스템의 IP 주소 또는 이름이며 `port`는 애플리케이션 서버 포트 번호입니다. 이들 정보를 모르는 경우에는 관리자에게 문의하십시오. 브라우저에 이 페이지의 책갈피를 추가하거나 데스크탑에 단축키를 작성할 수도 있습니다.

IBM Analytical Decision Management URL 또는 임의의 IBM SPSS Collaboration and Deployment Services URL에 액세스하는 중 문제점이 발생한 경우 관리자가 이를 사용자 정의했는지 확인하십시오. 관

리자에게 문의하십시오. *IBM SPSS Collaboration and Deployment Services* 리포지토리 설치 및 구성 안내서의 애플리케이션 컨텍스트 루트 장에 추가 정보가 있습니다.

- 새 프로젝트, 모델 또는 규칙을 작성하려면 원하는 애플리케이션에 대한 드롭 다운 목록에서 새로 작성을 선택하고 이동을 클릭하십시오.
- 저장된 프로젝트, 모델 또는 규칙을 열려면 애플리케이션 패널의 드롭 다운 목록에서 최근에 사용된 버전 또는 단축키를 선택하거나, 드롭 다운 목록에서 찾아보기를 선택하여 다른 오브젝트 또는 버전을 선택하십시오. 목록에 있는 항목의 리포지토리 위치를 빠르게 보려면 드롭 다운 메뉴에서 커서를 해당 항목 위에 두십시오.
- 드롭 다운 목록에서 항상 사용 가능한 새 사용자 정의 단축키를 추가하려면 애플리케이션 패널의 오른쪽 상단에 있는 삼각형 화살표 아이콘을 클릭하십시오. 선택적으로, 표시할 단축키의 수를 지정할 수 있습니다.

애플리케이션을 여는 다른 방법

- 브라우저의 주소 표시줄에 URL을 복사 또는 입력하거나, 브라우저의 저장된 책갈피 또는 즐겨찾기 목록에 URL을 추가하십시오.
- URL을 가리키는 데스크탑 단축키를 작성하십시오.
- 회사의 인트라넷 사이트 또는 포털에서 애플리케이션 URL을 링크하십시오.

시작 페이지 사용자 정의

애플리케이션 시작 페이지의 오른쪽 상단에 있는 제어를 사용하면 애플리케이션을 추가하고 패널과 목록 보기 표시 사이를 전환할 수 있습니다. 페이지에서 패널을 끌어 재배열할 수도 있습니다.

- 새 애플리케이션 패널을 추가하려면 애플리케이션 시작 페이지의 오른쪽 상단에 있는 애플리케이션 추가를 선택하십시오. (사용 가능한 추가 애플리케이션이 없는 경우에는 이 링크를 사용할 수 없습니다.)

사용 가능한 애플리케이션의 목록이 표시됩니다. 선택적으로, 하나 이상의 레이블된 버전에 대한 단축키를 작성할 수 있습니다.

- 시작 페이지에서 애플리케이션 패널의 배열을 변경하려면 패널의 제목 표시줄을 클릭하고 마우스 단추를 누른 채 패널을 새 위치로 끌어 놓으십시오.
- 애플리케이션을 제거하려면 애플리케이션 패널의 오른쪽 상단에 있는 제거 아이콘(적색 X)을 클릭하십시오. 해당 애플리케이션을 가리키는 모든 단축키도 제거됩니다. (페이지에 있는 애플리케이션 패널이 관리자에 의해 잠긴 경우에는 제거 아이콘을 사용할 수 없습니다.)
- 저장되지 않은 프로젝트, 모델 및 규칙은 시작 페이지의 갤러리에 나열됩니다. 이 갤러리는 아직 작성 중인 모델과 작성이 완료되었지만 저장되지 않은 모델을 표시합니다. 자세한 정보는 27 페이지의 『갤러리』 주제를 참조하십시오.

홈 페이지

애플리케이션은 홈 페이지에 있는 아이콘에 나타나 있는 바와 같이 단계별 워크플로우를 제공합니다. 임의의 단계로 점프하려면 해당 아이콘을 클릭하십시오.

데이터. 애플리케이션에서 사용할 데이터 세트를 정의하십시오.

모델. 분석 데이터를 사용하여 모델을 작성하고, 데이터에 숨겨진 패턴을 발견함으로써 데이터를 꿰뚫어 볼 수 있도록 예측변수 중요도, 분포도 및 Gains 차트 등을 사용하여 모델을 평가하고, 이러한 패턴을 사용하여 더 정확한 의사결정을 내리십시오.

스코어링. 모델을 스코어링하십시오.

갤러리

갤러리는 애플리케이션 시작 페이지에 표시됩니다. 여기에는 밤 사이에 작성되도록 놓아 둔 모델과 같은 저장되지 않은 작업이 나열되며, 퇴근하기 전에 모델을 시작하고 브라우저를 닫은 후 아침에 돌아와서 결과를 확인할 수 있게 해 줍니다. 브라우저를 실수로 닫은 후 진행하던 작업을 검색할 경우에도 갤러리를 사용할 수 있습니다. 갤러리의 항목에는 프로젝트, 규칙, 모델, 데이터 스캔, 평가, 그리고 해당 애플리케이션에 적절한 기타 장기간 작업이 포함됩니다.

- 항목의 이름에 밑줄이 있는 경우에는 이름을 클릭하여 관련 항목을 여십시오. 예를 들어, 저장되지 않은 데이터 스캔이 있는 경우에는 갤러리의 데이터 스캔 제목을 클릭하여 데이터 탭을 여십시오.
- 갤러리에 표시된 항목이 필요하지 않은 경우에는 삭제 아이콘을 클릭하여 제거하십시오. 이 항목을 영구적으로 삭제할 것인지 확인하기 위한 경고 메시지가 표시됩니다. 또는 여러 갤러리 항목을 한 번에 제거하려면 갤러리 지우기를 클릭하십시오. 갤러리 지우기 대화 상자가 표시되며 포함된 모든 오브젝트를 삭제할 탭을 선택할 수 있습니다.
- 갤러리의 프로젝트 탭에서 만든 위치 드롭 다운 메뉴는 각 애플리케이션에 대한 탭에 표시된 항목을 필터링할 수 있게 해 줍니다. 예를 들어, 한 애플리케이션에서 작업 중에 다른 애플리케이션의 갤러리 항목을 보이지 않으려는 경우가 있습니다.

환경 설정

사용자의 필요에 따라 IBM Analytical Decision Management를 사용자 정의하기 위해 조정할 수 있는 환경 설정에는 몇 가지가 있습니다. 기본적으로 이 사용자 정의는 작업을 작성하고, 테스트하고 저장할 때 같은 조치를 반복적으로 선택할 필요가 없도록 자신만의 기본 옵션을 설정하는 것으로 구성되어 있습니다. 사용자가 설정한 환경 설정은 자신에게 고유하며 사용하는 모든 애플리케이션에 적용됩니다.

사용자 환경 설정 설정에 액세스하려면 임의의 애플리케이션에서 도구 모음에 있는 사용자 환경 설정 설정 아이콘을 클릭하십시오.



그림 3. 사용자 환경 설정 설정 아이콘

다음 절에 자세히 설명되어 있는 내용과 같이 각 영역에 대해 원하는 설정을 지정할 수 있습니다.

일반 설정

애플리케이션 단축키에 있는 최근에 열어 본 오브젝트의 기본 수. 애플리케이션 시작 페이지에 있는 드롭 다운 목록에서 사용 가능한 항목의 수를 선택하십시오.

날짜 형식. 날짜를 표시하거나 기록할 때 사용할 형식으로 선택하십시오.

시간 형식. 시간을 표시하거나 기록할 때 사용할 형식으로 선택하십시오.

통화. 각 새 프로젝트의 기본 통화를 선택하십시오. 이는 다른 통화를 사용하는 기존 데이터를 겹쳐쓰지 않는다는 점을 참고하십시오.

소수점 이하 자릿수. 숫자에 대해 표시할 소수점 이하 자릿수의 수를 선택하십시오.

미리보기에 표시할 행 수. 미리보기를 표시할 때(예: 가져오기 시 데이터 소스를 확인할 때) 표시할 레코드의 최대 수를 선택하십시오.

페이지당 결과 수(페이지 번호를 매긴 테이블). 테이블을 표시할 때 각 페이지에 표시할 결과의 최대 수를 선택하십시오.

그래프에 표시되는 항목의 최대 수. 표시할 항목이 너무 많은 경우에는 그래프 결과가 원활하게 작동하지 않을 수 있습니다. 그래프에 표시할 항목의 최대 수를 지정하십시오. 지정하면 항목의 수가 이 최대값을 초과할 경우 그래프 대신 테이블이 표시됩니다. 이 설정은 데이터 개요의 범주형 필드, 그리고 범주형 대상이 있는 모델에 대한 스코어링 요약 보고서 그래프에만 적용됩니다.

차트 및 테이블에 적용할 데이터 세그먼트(타일) 수. 적용 가능한 경우, 차트 및 테이블에 대한 애플리케이션에 사용할 데이터 세그먼트의 최대 수를 선택하십시오.

모델 설정

기본적으로 빌드되는 모델 유형. 특정 결과에 대해 더 높거나 낮은 가능성을 보이는 세그먼트를 식별하는 모델(예측 규칙 모델)을 사용하려 하거나, 레코드를 몇 개의 군집으로 분류하는 모델(군집 모형)을 사용하려 하거나, 같이 나타날 가능성이 높은 항목들을 예측하고 이 항목들 사이에 있는 관계의 강도를 예측하는 모델(연관 모델)을 사용하려 하는 경우에 기본적으로 애플리케이션이 사용 가능한 데이터에서 최선의 모델(예측 모델)을 자동으로 빌드하도록 할 것인지 지정하십시오. 이 드롭 다운 메뉴에서는 애플리케이션에 표시되도록 구성된 모델 유형만 사용 가능합니다.

평가 및 테스트를 위해 빌드 데이터 소스에 대한 모형 평가를 할 수 있도록 자동으로 데이터를 파티셔닝. 이 옵션을 선택하면 훈련 및 모델의 테스트를 위해 데이터를 별도의 서브세트 또는 샘플로 분할합니다. 하나의 서

브세트에 모델을 빌드하고 이를 다른 서브세트에 테스트하면 이것이 다른 데이터 세트에 어떻게 적용될지 알 수 있습니다. 임의로 생성되는 훈련 백분율 값 및 테스트 파티션 크기의 백분율 값을 지정할 수도 있습니다. 이러한 옵션은 연관 모델링을 제외한 모든 모델 유형에 대해 사용할 수 있습니다.

신뢰할 수 있는 모델 빌드를 위해 자동으로 데이터 정리 및 준비. 이 옵션을 선택하면 애플리케이션에서 데이터 준비의 많은 부분을 수행하도록 할 수 있습니다. 예를 들어, 데이터 문제를 복구하거나, 덜 유용한 데이터를 선별하거나, 새 속성을 파생하는 것 등이 있습니다.

매우 큰 데이터 세트에 대해 작업하거나 빠른 결과를 얻으려 할 때는 정리 조치가 속도와 정확성 간에 균형을 맞추도록, 또는 속도나 정확성 중 한 쪽을 우선하도록 선택할 수 있습니다.

참고: 자동 데이터 준비 설정은 새 모델을 작성할 경우에만 적용되며 구간화는 수행되지 않습니다.

하나의 모델을 작성하는 데 소요되는 시간 제한. 각 모델을 작성하는 데 소요되는 기본 시간을 지정하려면 이 옵션을 사용하십시오. 이는 큰 데이터 세트를 처리할 때 특히 유용하며 모델 작성이 완료되는 데 필요한 시간에 큰 영향을 줄 수 있습니다.

테스트 및 시뮬레이션 설정

테스트되는 레코드의 기본 수. 테스트를 실행할 때 포함시킬 레코드의 기본 수입니다.

기본 테스트/시뮬레이션 날짜. 기본 시뮬레이션 날짜를 현재 날짜로 설정할 것인지, 또는 특정 날짜로 설정할 것인지 지정합니다.

프로젝트, 모델, 규칙 열기 및 저장

IBM SPSS Collaboration and Deployment Services Repository에서 열거나, 삭제하거나, 저장하거나, 로컬 파일 시스템으로 다운로드할 수 있습니다. 또한 IBM SPSS Modeler 스트림 파일을 로컬 파일 시스템에서 업로드하여 IBM SPSS Collaboration and Deployment Services Repository에 저장할 수 있습니다. IBM SPSS Collaboration and Deployment Services Repository에 저장하면 다중 버전, 레이블 지정 및 보안에 대한 지원을 포함, 중요한 이점을 제공받을 수 있습니다. 로컬 구성에 따라 지원되지 않는 옵션이 있을 수 있습니다.



새 모델을 작성합니다.



현재 프로젝트를 IBM SPSS Collaboration and Deployment Services Repository에서 마지막으로 알려진 위치에 저장합니다. 선택적으로, 저장할 때 새 레이블을 적용할 수 있습니다.



현재 프로젝트를 저장하고 레이블 지정하거나 새 이름으로 저장합니다. 적용 가능한 경우에는 IBM SPSS Collaboration and Deployment Services Repository에 저장하거나 로컬 파일 시스템에 다운로드하는 것 중 하나를 선택할 수 있습니다.



적용 가능한 경우 IBM SPSS Collaboration and Deployment Services Repository 또는 로컬 파일에서 프로젝트를 엽니다. 또는 적용 가능한 경우 로컬 파일 시스템에서 프로젝트를 업로드할 수도 있습니다. 모든 애플리케이션에서 파일 시스템 액세스를 지원하지는 않습니다. IBM SPSS Modeler 스트림을 업로드하는 경우에는 파일을 열거나 리포지토리에 저장하도록 선택할 수 있습니다.

리포지토리에서 열기 및 저장

IBM SPSS Collaboration and Deployment Services Repository는 다중 버전, 레이블 지정 및 보안에 대한 지원을 포함한 중요한 이점을 제공합니다. 저장할 때마다 새 버전이 작성되며 언제든지 이전 버전으로 되돌릴 수 있습니다. 또한 레이블을 사용하여 버전을 관리할 수 있으며 보안 설정을 적용하여 오브젝트를 액세스하고, 보고, 제거할 수 있는 사람을 결정할 수 있습니다.

왼쪽 창의 트리 보기는 IBM SPSS Collaboration and Deployment Services Repository에서 사용자가 볼 권한이 있는 폴더를 표시합니다. 폴더 트리를 탐색하거나 검색을 수행하여 원하는 항목을 찾으십시오. 왼쪽 하단에 있는 분할창에서 사용 가능한 레이블로 파일 목록을 필터링할 수도 있습니다. 예를 들어, 생산 레이블을 클릭하여 해당 레이블로 표시된 파일만 표시할 수 있습니다.

파일에 대한 세부사항을 보려면 옆에 있는 미리보기 아이콘이나 정보 아이콘(사용 가능한 경우)을 클릭하십시오.

파일을 삭제하려면 옆에 있는 삭제 아이콘을 클릭하십시오. 이는 IBM SPSS Collaboration and Deployment Services Deployment Manager 클라이언트 소프트웨어에 대한 액세스가 없는 사용자에게 특히 유용합니다.

폴더. 사용 가능한 폴더를 표시하며 계층 구조를 탐색할 수 있게 해 줍니다. 저장할 폴더를 선택하십시오. 폴더 목록 위에 있는 아이콘을 사용하여 새 폴더를 작성하거나 기존 폴더의 이름을 바꿀 수도 있습니다.

레이블로 필터링. 선택된 레이블로 오브젝트의 표시를 제한합니다. 폴더를 찾아볼 때는 이 레이블이 지정된 오브젝트만 표시됩니다.

애플리케이션에서 사용된 규칙을 내보낼 때는, 다른 이름으로 저장 또는 다른 이름으로 내보내기 드롭 다운도 사용 가능합니다. 드롭 다운에서 단일 규칙 세트 또는 개별 규칙을 선택하여 규칙을 저장할 방식을 지정하십시오. 참조된 규칙만 사용 가능한 경우에는 드롭 다운을 사용할 수 없게 되며 규칙이 단일 규칙 세트로 저장됩니다(참조된 규칙은 개별적으로 저장할 수 없음).

개별 규칙으로 저장하도록 선택한 경우에는 저장할 각 개별 규칙에 대한 세부사항을 입력할 레이블이 표시됩니다.

이름. 저장하는 오브젝트의 고유 이름을 입력하십시오.

키워드. 저장한 오브젝트를 검색할 때 이를 빨리 식별할 수 있는 키워드를 입력하십시오. *판매 1-2009*와 같은 것을 예로 들 수 있습니다.

설명. 저장된 오브젝트에 대한 자세한 설명을 입력하십시오. *소매 판매 분석, 2009년 1분기*와 같은 것을 예로 들 수 있습니다.

레이블. 저장되는 동일한 오브젝트의 여러 버전을 보유할 것 같은 경우에는 고유 레이블로 각각을 식별할 수 있습니다. 기존 레이블을 선택하거나 새 레이블을 작성하십시오.

레이블 작업

레이블은 프로젝트, 모델 또는 규칙의 특정 버전을 식별하는 데 사용됩니다. 예를 들어, 여러 프로젝트 간에 규칙을 공유할 경우 생산으로 레이블된 버전을 사용하도록 지정할 수 있습니다. 규칙의 새 버전이 작성된 경우 사용자의 프로젝트에 사용된 버전은 생산 레이블이 명시적으로 새 버전으로 이동되지 않는 한 변경되지 않습니다.

공유 오브젝트를 참조할 때는 최신 버전의 변경사항이 의도하지 않은 결과를 발생시킬 수 있으므로, 특히 생산 환경에서는 레이블된 버전을 사용하는 것이 좋습니다. 규칙 내에서 규칙을 참조할 때와 같은 특정 경우에는 반드시 레이블된 버전을 사용해야 합니다. 자세한 정보는 56 페이지의 『내보내기 및 재사용 규칙』 주제를 참조하십시오.

버전이 지정되지 않은 경우에는 최신 버전이 사용되며 이는 규칙에 대한 후속 변경사항을 자동으로 적용한다는 것을 의미합니다. (최신은 각 새 버전이 작성되면 자동으로 해당 버전으로 이동되는 특수 레이블이라고 생각하십시오.)

새 오브젝트 또는 버전을 저장할 때마다 해당 버전에 적용할 레이블을 선택할 수 있습니다. 자세한 정보는 29 페이지의 『프로젝트, 모델, 규칙 열기 및 저장』 주제를 참조하십시오.

레이블 이름의 제한사항. IBM Analytical Decision Management에서 사용할 레이블을 정의할 때 문자 ? # & 및 :(콜론)은 레이블 이름에 포함시킬 수 없는데, 이는 이러한 문자가 이와 같은 레이블을 사용하는 오브젝트에 대한 참조를 애플리케이션이 올바르게 해석할 수 없게 만들 수 있기 때문입니다. 예를 들어, 다음과 같은 이름은 올바른 레이블 이름이 아닙니다.

테스트 & 배포

배포:PHASE1

프로젝트 잠금

프로젝트, 모델 또는 규칙을 열면 오른쪽 상단에 있는 잠금 선택란을 사용하여 다른 사용자가 편집할 수 없도록 IBM SPSS Collaboration and Deployment Services Repository에서 오브젝트를 잠글 수 있습니다.

- 프로젝트가 잠기면 다른 사용자는 이 프로젝트를 보기 위해 열 수 있으나 변경사항을 저장할 수는 없습니다.
- 잠긴 프로젝트를 닫으면(예: 브라우저 닫기) 잠금이 해제되어 다른 사용자가 오브젝트를 편집할 수 있습니다.
- 프로젝트를 닫지 않고 잠금 해제하려는 경우(잠금 선택란 선택 취소를 통해) 변경사항 저장 여부에 대해 프롬프트됩니다. 저장되지 않은 변경사항은 오브젝트가 잠금 해제되면 사라집니다.
- 프로젝트 잠금을 해제하고 나면 다른 사용자가 잠금을 수행하지 않을 때까지 잠금이 해제되어 있는 동안 계속해서 이에 대한 변경을 수행할 수 있습니다.

특성

특성 대화 상자에는 프로젝트 탭 및 리포지토리 탭이 포함됩니다. 특성 대화 상자를 표시하려면 애플리케이션에서 특성 아이콘을 클릭하십시오.



그림 4. 특성 아이콘

프로젝트 탭에서는 명목 필드에 허용된 최대 멤버 수를 지정할 수 있습니다. 이는 데이터 탭에서 사용됩니다. 명목 필드에 대한 최대 멤버 수를 선택하고 값을 지정하십시오. 기본값은 **250**입니다. 이는 다음과 같은 상황에서 유용합니다.

1. 사용자가 새 데이터 소스를 작성하고 데이터 스캔을 실행합니다.
2. 데이터에 있는 고유 값의 수가 특성 대화 상자에 설정된 기본 최대 크기 한계를 초과하므로 유용한 범주형 정보가 포함된 다수의 필드가 프로젝트 데이터 모델에 유형 없음으로 표시됩니다.
3. 특성 아이콘을 클릭하고 명목 필드에 대한 최대 멤버 수 값을 자체 데이터에 적합한 크기로 늘립니다.
4. 데이터 탭으로 돌아가 데이터 스캔 새로 고치기를 클릭하지만 사용자가 작성한 변경사항을 유지하기 위해 새로 고치기 조치가 측도를 변경하지 않기 때문에 해당 필드는 유형 없음으로 유지됩니다. 다시 말해서, 새로 고치기는 필드에 대한 기존 측도를 유지하며 사용자가 데이터 모델에서 작성한 변경사항이 유지되어야 하기 때문에 값을 재계산하기만 합니다.
5. 값을 프로젝트 데이터 모델로 가져오려면 새로 고치기를 실행하기 전에 유형 없음에서 범주형으로 각 필드의 측도를 수동으로 변경해야 합니다. 영향받는 필드가 많을 경우 더 빠르게 데이터 소스를 재작성할 수 있습니다.

명목 필드에 대한 최대 멤버 수 설정은 파일 유형 *.sav*의 데이터 소스에 적용되지 않습니다. 대다수의 *.sav* 파일에는 사용자 정의 파일 및 값이 있으므로 이 설정에 의해 영향을 받지 않습니다. 이 필드는 사용자가 수동으로 변경해야 합니다.

리포지토리 탭에서는 나중에 프로젝트, 모델 또는 규칙 세트를 찾아볼 때 이를 식별하는 데 도움을 줄 수 있도록 키워드 및 설명을 추가할 수 있습니다.

- **키워드.** 저장한 오브젝트를 검색할 때 이를 빨리 식별할 수 있는 키워드를 입력하십시오. 판매나 *Q1-2013* 과 같은 것을 예로 들 수 있습니다.
- **설명.** 저장된 오브젝트에 대한 자세한 설명을 입력하십시오. 소매 판매 분석, 2013년 1분기와 같은 것을 예로 들 수 있습니다.

제 4 장 데이터 소스 관리

데이터 탭에서는 분석, 시뮬레이션 및 테스트, 스코어링 또는 기타 용도로 데이터 소스를 정의할 수 있습니다. 데이터 세트는 관리자가 구성하거나 사용자가 직접 추가할 수 있습니다.

탭에는 소스 탭과 파생 탭이 있습니다. 소스 탭은 다음의 기본 영역으로 구분되어 있습니다.

- **프로젝트 데이터 모델.** 애플리케이션에서 필요로 하는 필드를 정의합니다. 모든 다른 데이터 세트는 이 소스를 기준으로 맵핑됩니다. 데이터 소스 필드 섹션에는 프로젝트가 사용하는 입력 필드와 유형이 표시되어 있습니다. 원하는 경우 파생 탭을 사용하면 서로 다른 데이터 소스의 표현식 또는 모델 결과를 추가하여 프로젝트 데이터 모델을 확장할 수 있습니다.
- **데이터 소스.** 프로젝트 데이터 소스 섹션에는 현재 프로젝트, 모델 또는 규칙에 사용하기 위해 저장된 데이터 소스가 표시됩니다. 내 데이터 소스 섹션에는 정의하거나 저장한 데이터 소스가 표시됩니다. 이 목록의 데이터 세트는 다른 프로젝트로 복사하거나 다른 프로젝트로부터 복사한 후 프로젝트 데이터 모델로 사용하도록 승격시킬 수 있으며 이를 통해 애플리케이션 간에 데이터를 공유할 수 있습니다.

프로젝트 데이터 모델

프로젝트 데이터 모델은 프로젝트에서 사용할 입력 필드 및 유형을 나열하는 템플릿과 같은 역할을 합니다.

데이터 소스. 애플리케이션에 대해 선택된 데이터 소스를 표시합니다.

레코드 ID. 원하는 경우 엔티티 ID로 사용할 필드를 선택하여 SQL 푸시백을 사용할 수 있습니다. 따라서 데이터베이스에서 데이터가 시작되고 선택한 필드가 정수인 경우 성능이 향상될 수 있습니다. IBM SPSS Modeler Advantage에서는 여기에서 선택된 필드가 일괄처리 스코어의 결과로써 기본적으로 선택되고 모델 빌드의 입력으로써 기본적으로 선택 취소됩니다. 이 필드는 선택사항입니다.

운영. 이 선택란의 체크 표시는 필드의 데이터가 규칙, 표현식 및 예측에 사용 가능함을 나타냅니다. 필드를 분석의 대상 필드로 사용할 경우 선택란을 제거하십시오. 예를 들어, 고객이 DM 캠페인에 대해 보이는 반응 정도를 확인하는 조회를 실행하는 경우 반응 값을 미리 알지 못하므로 결과 필드가 분석이 됩니다.

참고: IBM SPSS Collaboration and Deployment Services Enterprise View 데이터 소스를 프로젝트 데이터 모델로 선택하는 경우에는 이 열의 상자가 사용 안함으로 설정됩니다.

필드 이름. 데이터 소스에 나열된 각 필드의 이름을, 저장 공간 유형을 식별하는 아이콘과 함께 나열합니다.

수치. 데이터 소스에 있는 각 필드의 수치 유형을 표시합니다. 수치 유형을 변경할 경우에는 데이터 호환성을 확인하기 위해 데이터 스캔을 새로 고쳐야 합니다. 자세한 정보는 45 페이지의 『측정 수준』 주제를 참조하십시오.

값. 데이터 소스에 있는 각 필드의 값을 나열합니다. 예를 들어, 범위 내의 최대 및 최소값이 있습니다. 플래그 필드의 경우에는 필드에 대해 정의된 "true" 및 "false" 값을 표시하는 도구 톱을 표시하려면 값 위에 마우스

포인터를 두십시오. 값을 클릭하여 편집할 수도 있습니다.

프로젝트 데이터 소스

프로젝트 데이터 소스는 관리자가 사전 정의한 항목이나 사용자가 추가한 항목을 포함할 수 있습니다. 선택적으로, 관리자는 사용자가 수정하거나 제거할 수 없도록 하나 이상의 데이터 소스를 잠갔거나, 사용자가 새 데이터 소스를 작성할 수 없도록 모든 데이터 옵션을 잠갔을 수 있습니다.

데이터 소스의 입력이 프로젝트 데이터 모델의 입력과 직접 일치하지 않는 경우 전자를 후자에 매핑하고 차이를 수정할 수 있습니다. 예를 들어, 프로젝트 데이터 모델에서 이름이 구매이며 값이 예 및 아니오(측정 수준 플래그)인 필드를 필요로 하는 경우 사용된 데이터 소스에는 적절히 매핑될 수 있는 비교 가능한 필드가 있어야 합니다.

이름. 데이터 소스 이름을 표시하며 해당 파일 유형을 식별하는 아이콘을 표시합니다.

레코드의 수. 테이블에 있는 데이터 소스의 레코드 수를 표시하려면 이 열의 아이콘을 클릭하십시오. 개수는 아이콘 옆에 표시됩니다.

미리보기. 소스에 포함된 데이터의 샘플을 미리 보려면 이 열의 아이콘을 클릭하십시오. 자세한 정보는 44 페이지의 『데이터 미리보기』의 내용을 참조하십시오.

개요. 데이터 소스의 개요를 보려면 이 열의 아이콘을 클릭하십시오. 자세한 정보는 44 페이지의 『데이터 개요』의 내용을 참조하십시오.

호환 가능. 데이터 소스가 프로젝트 데이터 모델로 사용되었다는 노트를 표시하거나 데이터 소스가 프로젝트 데이터 모델과 호환되는 정도를 나타내는 녹색, 주황색 또는 빨간색 원을 표시합니다.

- 녹색 원은 데이터 소스가 프로젝트 데이터 모델 데이터 소스와 사용 가능하게 호환됨을 나타냅니다. 사용 가능하게 호환되는 데이터 소스는 프로젝트 데이터 모델의 모든 운영 필드를 포함하며 그 외에 추가 필드가 있는 데이터 소스입니다. 이 데이터 소스는 규칙, 스코어링, 시뮬레이션 및 테스트 조작에 적합합니다.
- 황색 원은 데이터 소스에 프로젝트 데이터 모델과 호환되는, 이름과 유형이 같은 필드가 하나 이상 있음을 나타냅니다. 이 데이터 소스에는 추가 필드가 있을 수 있으며 이 데이터 소스는 모델을 작성하고 평가하는데 적합합니다.
- 적색 원은 데이터 소스가 프로젝트 데이터 모델과 호환이 불가능하며 애플리케이션에서 사용하려면 먼저 필드를 매핑해야 함을 나타냅니다. 호환되지 않는 데이터 소스는 동격의 프로젝트 데이터 모델 유형과 호환되지 않는 유형의 필드가 하나 이상 있는 데이터 소스입니다.

필드 매핑. 이 옵션을 사용하면 데이터 소스 필드를 프로젝트 데이터 모델에 있는 것과 비교할 수 있으며 프로젝트 데이터 모델에서 필요로 하는 필드와 일치시키기 위해 호환 가능 필드를 매핑하거나 매핑 취소할 수 있습니다. 자세한 정보는 46 페이지의 『필드 매핑』의 내용을 참조하십시오.

복사. 데이터 소스를 내 데이터 소스 영역으로 복사합니다.

내 데이터 소스

이 목록의 데이터 소스는 사용자 계정과 함께 저장되므로 로그인할 때마다 사용할 수 있으며 사용자가 여는 프로젝트, 모델 또는 규칙에 복사할 수 있습니다(관리자가 이러한 권한을 사용자에게 부여한 경우). 탭의 이 부분에 있는 필드는 프로젝트 데이터 소스 영역에 있는 필드와 같은 방식으로 작동하지만, 호환성 열은 없습니다.

복사 열은 데이터 소스를 프로젝트 데이터 소스 영역으로 복사할 수 있게 해 줍니다.

데이터 소스에 대한 작업

- 프로젝트 데이터 소스 또는 내 데이터 소스 목록에 새 데이터 소스를 추가하려면 데이터 소스 추가를 선택하십시오. 자세한 정보는 『새 데이터 소스 작성』의 내용을 참조하십시오.
- 프로젝트 데이터 모델의 데이터 소스에 있는 필드의 측정 수준 또는 유형을 변경하려면 관련 수준을 선택하십시오. 자세한 정보는 45 페이지의 『측정 수준』의 내용을 참조하십시오.
- 프로젝트 데이터 소스 목록에서, 또는 이 목록으로 데이터 소스를 복사하려면 복사 열에 있는 적절한 회살 표를 클릭하십시오.
- 프로젝트 데이터 모델에 데이터 소스의 필드 이름을 맵핑하려면 호환 가능 열에 있는 적절한 링크를 클릭하십시오. (필드 이름이 맵핑된 후에는 링크가 더 이상 표시되지 않습니다.) 자세한 정보는 46 페이지의 『필드 맵핑』의 내용을 참조하십시오.
- 데이터 소스를 미리 보려면 미리보기 아이콘을 클릭하십시오. 자세한 정보는 44 페이지의 『데이터 미리보기』의 내용을 참조하십시오.
- 필드(표현식, 세그먼트 규칙을 사용하는 필드 또는 다른 데이터 소스의 모델 결과)를 추가하여 프로젝트 데이터 모델을 확장하려면 파생 탭을 클릭하십시오. 자세한 정보는 47 페이지의 『필드 파생』의 내용을 참조하십시오.

새 데이터 소스 작성

1. 데이터 소스 편집기에 액세스하려면 프로젝트 데이터 소스 또는 내 데이터 소스 영역에 있는 데이터 소스 추가를 선택하십시오.
2. 데이터 소스의 이름을 입력하십시오.
3. 소스 유형을 선택하십시오. 사용 가능한 데이터베이스 목록은 IBM SPSS Modeler 시스템에서 사용할 수 있는 목록입니다.
 - 파일. 지원되는 파일 유형에는 텍스트(*.txt, *.csv 등), Excel, IBM SPSS Statistics에서 사용되는 *.sav 및 *.zsav 형식, SAS 파일 등이 있습니다. 자세한 정보는 36 페이지의 『파일 소스』 주제를 참조하십시오.
 - 데이터베이스. Microsoft SQL Server, DB2, Oracle 등과 같이 ODBC를 지원하는 모든 데이터베이스가 해당됩니다. 자세한 정보는 39 페이지의 『데이터베이스 소스』 주제를 참조하십시오.

- 엔터프라이즈 뷰. IBM SPSS Collaboration and Deployment Services에 정의된 바와 같이 IBM SPSS Collaboration and Deployment Services Enterprise View와의 연결을 작성합니다. 자세한 정보는 40 페이지의 『엔터프라이즈 뷰 소스』 주제를 참조하십시오.
- 데이터 보기. IBM Operational Decision Management 규칙과 함께 사용할 연결을 작성합니다. 계층 구조 데이터 모델을 사용하는 복합 규칙의 경우, 관리자는 성능이 저하되지 않도록 해당 규칙의 사용을 제한할 수 있습니다. 자세한 정보는 40 페이지의 『데이터 보기 소스』 주제를 참조하십시오.
- **IBM Cognos BI 서버.** IBM Cognos BI 서버를 지원하는 모든 데이터베이스가 해당됩니다. 자세한 정보는 41 페이지의 『IBM Cognos BI 소스』 주제를 참조하십시오.
- **Analytic Server.** 새 데이터 소스 작성 시 이 소스 유형을 선택할 수 없습니다. IBM SPSS Modeler 16에서, IBM SPSS Analytic Server이 소개되어 데이터 소스로 HDFS(Hadoop Distributed File System)의 사용을 허용합니다. IBM SPSS Modeler 사용자는 Hadoop 대형 데이터 시스템에 기반하여 HDFS Analytic Server 소스 노드 및 빌드 모델을 사용하여 스트림을 작성할 수 있습니다. IBM Analytical Decision Management 사용자는 보려는 IBM SPSS Modeler Advantage의 해당 모델 스트림을 열 수 있습니다(Analytic Server 데이터 소스를 통해 새 프로젝트를 작성할 수 없음).

4. 소스 유형에 적절한 옵션을 지정하십시오.

레코드의 수 아이콘을 클릭하여 현재 데이터 소스에 있는 레코드의 수를 표시할 수 있으며, 또는 미리보기를 클릭하여 데이터 소스의 레코드 중 처음 열 개의 레코드를 표시할 수 있습니다.

파일 소스

애플리케이션은 몇 가지 데이터 파일 유형을 지원합니다. 파일 이름을 입력하거나 파일을 찾아볼 때 애플리케이션은 자동으로 유형을 발견하고 데이터 소스 편집기를 펼쳐 해당 유형과 관련된 추가 필드를 표시합니다. 파일 유형이 올바르게 표시되지 않거나 특정 버전을 필요로 하는 경우(예: Microsoft Excel 1997-2003)에는 필요한 유형으로 변경할 수 있습니다.

찾아보기를 클릭하여 리포지토리에서 파일을 선택하거나 로컬 파일 업로드를 클릭하여 로컬 파일 시스템에서 파일을 선택할 수 있습니다(사용으로 설정된 경우).

로컬 파일을 업로드하는 경우 이는 IBM SPSS Modeler Server에 업로드됩니다. 파일 업로드 대화 상자에서 업로드할 로컬 파일을 찾아 선택하고 IBM SPSS Modeler Server 업로드 대상을 찾아 선택하십시오. 대상을 선택할 때 원하는 경우에는 파일 이름을 변경할 수 있습니다.

로컬 파일 업로드 기능은 기본적으로 사용 안함으로 설정되어 있습니다. 관리자는 브라우저 기반 IBM SPSS Collaboration and Deployment Services Deployment Manager를 사용하여 이를 사용으로, 또는 사용 안함으로 설정하거나 허용된 최대 파일 크기 한계를 늘리고 줄일 수 있습니다.

텍스트 기반 데이터 파일

텍스트 기반 데이터 소스를 선택하면 세부사항을 입력하도록 프롬프트됩니다.

첫 번째 행에 열 이름 있음. 각 열의 이름이 데이터 소스에 표제 행으로 포함될 경우에는 이를 선택하십시오.

인코딩. 사용되는 텍스트 인코딩 메소드를 지정합니다. 시스템 기본값이나 UTF-8을 선택할 수 있습니다.

소수점이하자리 기호. 데이터에서 소수점이하자리가 어떻게 표현될지 지정합니다.

- 마침표(.). 마침표 문자가 소수점 구분 기호로 사용됩니다.
- 쉼표(.). 쉼표 문자가 소수점 구분 기호로 사용됩니다.

구분자. 이 제어에 나열된 선택란을 사용하여 쉼표(.)와 같은, 파일의 필드 경계를 정의하는 문자를 지정할 수 있습니다. 여러 구분자를 사용하는 레코드의 경우에는 ", |"와 같이 둘 이상의 구분자를 지정할 수도 있습니다. 기본 구분자는 쉼표입니다.

참고: 쉼표가 소수점 구분 기호로도 정의되어 있는 경우에는 기본 설정이 작동하지 않습니다. 쉼표가 필드 구분자인 동시에 소수점 구분 기호인 경우에는 구분자 목록에서 기타를 선택하십시오. 그런 후 수동으로 입력 필드에 쉼표를 지정하십시오.

여러 인접 공백 구분 문자를 하나의 구분자로 취급하려면 다수 공백 구분자 허용을 선택하십시오. 예를 들어, 한 데이터 값 뒤에 네 개의 공백이 있고 그 뒤에 다른 데이터 값이 있는 경우 이 그룹은 다섯 개가 아니라 두 개의 필드로 취급됩니다.

고급 옵션

EOL 주석 문자. 데이터에서 주석을 나타낼 # 또는 !와 같은 문자를 지정하십시오. 데이터 파일에서는 이 문자가 있는 위치부터 다음 줄 바꾸기 문자가 나오기 전까지 모든 내용이 무시됩니다.

입력 필드 지정. 각 레코드에서 사용될 입력 필드의 수를 지정하십시오.

데이터 형식 지정 - 일반 텍스트의 변수 길이 필드 유형인 파일 데이터 소스 유형의 경우, 이 섹션을 사용하여 값을 정확하게 읽도록 각 필드의 형식 및 입력 스토리지 유형을 설정할 수 있습니다. 이는 IBM SPSS Modeler 클라이언트에서 사용 가능한 기능과 유사합니다. 대체 옵션은 기본값이 대체되는지 여부를 표시합니다. 대체를 선택하면 스토리지 및 입력 형식 제어를 사용할 수 있습니다. 대체를 선택 취소하면 값이 원래 기본값으로 다시 변경됩니다. 입력 형식은 실제, 날짜, 시간, 시간소인 스토리지 유형에만 적용됩니다.

헤더 문자 건너뛰기. 첫 번째 레코드의 시작에서 몇 개의 문자를 무시할지 지정하십시오.

유형에 대해 스캔할 행. 지정된 데이터 유형에 대해 스캔할 행 수를 지정하십시오.

선행 및 후미 공백 제거. 가져올 때 문자열에서 선행 및 후미 공백을 버리려면 선택하십시오. 공백은 왼쪽이나 오른쪽, 양쪽 모두에서 제거하도록 선택할 수 있으며 제거하지 않도록 선택할 수도 있습니다.

올바르지 않은 문자. 데이터 소스에서 올바르지 않은 문자를 제거하려면 버리기를 선택하십시오. 지정된 기호(한 문자)로 올바르지 않은 문자를 대체하려면 다른 문자로 대체를 선택하십시오. 올바르지 않은 문자는 널 문자 또는 지정된 인코딩 메소드에 없는 문자입니다.

따옴표. 드롭 다운 목록을 사용하여, 가져올 때 작은따옴표와 큰따옴표가 어떻게 처리되는지 지정할 수 있습니다. 버리기를 선택하여 모든 따옴표를 버리거나, 텍스트로 포함을 선택하여 따옴표를 필드 값으로 포함하거나,

짜맞춤 및 버리기를 선택하여 따옴표 쌍을 만들고 이를 제거할 수 있습니다. 따옴표의 짝이 맞지 않는 경우에는 오류 메시지를 수신합니다. 버리기와 짜맞춤 및 버리기는 모두 필드 값을 따옴표 없이 문자열로 저장합니다.

입력 필드 지정. 자세한 정보는 44 페이지의 『입력 필드 선택』 주제를 참조하십시오.

Excel 데이터 파일

Excel 데이터 소스를 선택하면 세부사항을 입력하도록 프롬프트됩니다.

첫 번째 행에 열 이름 있음. 각 열의 이름이 데이터 소스에 표제 행으로 포함될 경우에는 이를 선택하십시오.

이름 지정된 범위. Excel 워크시트에 정의된, 이름 지정된 범위의 셀을 선택할 수 있게 해 줍니다. 이름 지정된 범위를 사용할 경우 다른 워크시트 및 데이터 범위 설정은 더 이상 적용할 수 없으며 사용 안함으로 설정됩니다.

워크시트 선택. 가져올 워크시트를 이름 또는 인덱스를 사용하여 지정합니다.

- 이름 사용. 가져올 워크시트의 이름을 선택하십시오.
- 인덱스 사용. 첫 번째 워크시트부터 0으로 시작하여 두 번째 워크시트는 1, 이런 방식으로 가져올 워크시트의 인덱스 값을 지정하십시오.

워크시트의 범위. 공백이 아닌 첫 번째 행이나 명시적 셀 범위부터 데이터를 가져올 수 있습니다.

- 공백이 아닌 첫 번째 행부터 범위 시작. 공백이 아닌 첫 번째 셀을 찾고 이를 데이터 범위의 왼쪽 상단으로 사용합니다.
- 명시적 셀 범위. 행과 열의 명시적 범위를 지정할 수 있게 해 줍니다. 예를 들어, Excel 범위 A1:D5를 지정하려면 첫 번째 필드에 A1을 입력하고 두 번째 필드에 D5를 입력할 수 있습니다(또는 R1C1 및 R5C4). 지정된 범위의 행은 공백 행을 포함하여 모두 리턴됩니다.

공백 행의 경우. 공백 행을 발견하면 이를 건너뛰고 무시할지, 또는 공백 행 리턴을 선택하여 공백 행을 포함, 워크시트의 끝까지 모든 데이터를 읽을 것인지 선택할 수 있습니다.

입력 필드 지정. 자세한 정보는 44 페이지의 『입력 필드 선택』 주제를 참조하십시오.

IBM SPSS Statistics 데이터 파일

IBM SPSS Statistics 데이터 소스 (.sav 또는 .zsav 파일)를 선택하면 세부사항을 입력하도록 프롬프트됩니다. 파일의 비밀번호가 보호되는 경우 비밀번호를 입력하라는 프롬프트도 표시됩니다.

변수 이름. IBM SPSS Statistics .sav 또는 .zsav 파일에서 가져오기를 수행한 후 변수 이름 및 레이블을 처리할 방법을 선택하십시오.

- 이름 및 레이블 읽기. 변수 이름과 레이블을 모두 읽으려면 선택하십시오. 이는 기본 옵션입니다. 레이블은 차트, 모델 브라우저 및 기타 출력 유형에 표시될 수 있습니다.

- 레이블을 이름으로 읽기. IBM SPSS Statistics .sav 파일에서 짧은 필드 이름이 아니라 변수 설명 레이블을 읽고 이 레이블을 변수 이름으로 사용하려면 선택하십시오.

값. IBM SPSS Statistics .sav 또는 .zsav 파일에서 가져오기를 수행한 후 값 및 레이블을 처리할 방법을 선택하십시오.

- 데이터 및 레이블 읽기. 실제 값과 값 레이블을 모두 읽으려면 선택하십시오. 이는 기본 옵션입니다.
- 레이블을 데이터로 읽기. 값을 나타내기 위해 사용된 숫자 또는 기호 코드 대신 .sav 또는 .zsav 파일의 값 레이블을 사용하려면 선택하십시오. 예를 들어, 값 1과 2가 각각 실제로는 남성과 여성을 가리키는, 성별 필드가 있는 데이터에 대해 이 옵션을 선택하면 필드가 문자열로 변환되며 남성과 여성을 실제 값으로 가져옵니다.

이 옵션을 선택하기 전에 IBM SPSS Statistics 데이터의 결측값을 고려하는 것이 중요합니다. 예를 들어, 숫자 필드가 결측값에 대해서만 레이블을 사용하는 경우(0 = 응답 없음, -99 = 알 수 없음)에는 위의 옵션을 선택하면 값 레이블 응답 없음과 알 수 없음만 가져오며 필드를 문자열로 변환합니다. 이러한 경우에는 값 자체를 가져와야 합니다.

입력 필드 지정. 자세한 정보는 44 페이지의 『입력 필드 선택』 주제를 참조하십시오.

데이터베이스 소스

데이터베이스 소스를 선택하면 세부사항을 입력하도록 프롬프트됩니다.

데이터베이스, 사용하려는 데이터베이스를 선택하십시오.

연결. IBM SPSS Collaboration and Deployment Services Repository에 저장된 신임 정보 또는 데이터베이스 신임 정보를 사용하여 데이터베이스에 연결할지 선택하십시오. 그런 다음 데이터베이스 로그인 이름 및 비밀번호를 입력하거나 리포지토리에서 신임 정보를 선택하고 연결을 클릭하여 데이터베이스에 대한 액세스 권한이 있는지 확인하십시오.

데이터베이스 로그인 이름 및 비밀번호는 저장되지 않습니다. 따라서 관리자가 IBM SPSS Modeler Server 및 ODBC 데이터베이스 연결을 위해 싱글 사인온(SSO)을 구성하지 않은 경우 다음 번에 프로젝트를 열어 실행할 때 로그인할지 묻는 프롬프트가 표시됩니다. 자세한 정보는 IBM Analytical Decision Management 설치 안내서의 **보안 고려사항** 섹션을 참조하십시오.

테이블 이름. 사용될 데이터가 있는 테이블을 선택하십시오.

선행 및 후미 공백 제거. 가져올 때 문자열에서 선행 및 후미 공백을 버리려면 선택하십시오. 공백은 왼쪽이나 오른쪽, 양쪽 모두에서 제거하도록 선택할 수 있으며 제거하지 않도록 선택할 수도 있습니다.

테이블 및 열 이름 따옴표로 묶기. 쿼리가 데이터베이스로 전송될 때(예: 이름에 공백이나 구두점이 있는 경우) 테이블 및 열 이름을 따옴표로 묶을지 지정합니다.

- 필요에 따라 옵션은 오직 테이블 및 필드 이름에 표준이 아닌 문자가 포함되어 있을 경우에만 이 이름을 따옴표로 묶습니다. 표준이 아닌 문자에는 비ASCII 문자, 공백 문자, 그리고 마침표(.)를 제외한 비영숫자 문자가 포함됩니다.

- 모든 테이블 및 필드 이름을 따옴표로 묶으려는 경우에는 **항상**을 선택하십시오.
- **절대로** 테이블 및 필드 이름을 따옴표로 묶지 않으려는 경우에는 **묶지 않음**을 선택하십시오.

입력 필드 지정. 자세한 정보는 44 페이지의 『입력 필드 선택』 주제를 참조하십시오.

IBM Analytical Decision Management는 데이터를 읽고 쓰고 푸시하도록 IBM SPSS Modeler와 동일한 데이터베이스 데이터 소스를 지원합니다. 자세한 정보는 IBM SPSS Modeler 17의 제품 호환성 보고서를 참조하십시오. 리포지토리를 호스팅하는 데이터베이스인 IBM SPSS Collaboration and Deployment Services 7에서 지원하는 데이터베이스와 다르다는 점을 참고하십시오.

엔터프라이즈 뷰 소스

엔터프라이즈 뷰 소스를 선택하면 세부사항을 입력하도록 프롬프트됩니다.

애플리케이션 보기. 애플리케이션 보기가 있을 경우 드롭 다운 목록이 선택된 애플리케이션 보기를 표시합니다. 현재 세션에서 다른 애플리케이션 보기와 연결이 설정된 경우에는 이러한 보기도 드롭 다운 목록에 표시됩니다. 리포지토리 내의 애플리케이션 보기를 검색하려면 **찾아보기**를 클릭하십시오.

테이블. 사용될 데이터가 있는 애플리케이션 보기 테이블을 선택하십시오.

데이터 제공자. 드롭 다운 목록이 선택된 애플리케이션 보기를 참조하는 데이터 제공자 정의의 이름을 표시합니다. 데이터 제공자 정의를 선택하거나 **-데이터 소스 없음-** 옵션을 선택하십시오. **-데이터 소스 없음-** 옵션을 선택하는 것은 데이터 소스가 애플리케이션에서 사실상 쓸모없으며 테스트, 시뮬레이션 등에서 사용할 수 없음을 의미합니다.

환경. IBM SPSS Collaboration and Deployment Services에서 설정된 환경 값은 어느 데이터 제공자 정의가 사용 가능한지 결정합니다.

입력 필드 지정. 자세한 정보는 44 페이지의 『입력 필드 선택』 주제를 참조하십시오.

IBM SPSS Collaboration and Deployment Services Enterprise View에 대한 자세한 정보와 이 데이터가 구성되는 방식에 대한 전체 세부사항은 IBM SPSS Collaboration and Deployment Services 문서를 참조하십시오. IBM SPSS Collaboration and Deployment Services Enterprise View는 이후 릴리스에서 더 이상 사용되지 않습니다.

데이터 보기 소스

데이터 보기 소스에서는 IBM Operational Decision Management 규칙과 함께 사용할 연결을 작성합니다. 계층 구조 데이터 모델을 사용하는 복합 규칙의 경우, 관리자는 성능이 저하되지 않도록 해당 규칙의 사용을 제한할 수 있습니다. 세부사항은 87 페이지의 『데이터 옵션 잠금』의 내용을 참조하십시오.

IBM SPSS Collaboration and Deployment Services 분석 데이터 보기에 정의된 스트림의 데이터를 포함하도록 **데이터 보기 노드**를 IBM SPSS Modeler에서 구성할 수 있습니다. 분석 데이터 보기는 예측 모델 및 비즈니스 규칙에서 사용되는 엔티티를 설명하는 데이터 액세스를 위한 구조를 정의합니다. 보기는 데이터 구조를 분석용 실제 데이터 소스와 연관시킵니다.

자세한 정보는 IBM SPSS Modeler 도움말의 소스, 프로세스 및 결과 노드 섹션 및 IBM SPSS Modeler 문서를 참조하십시오.

데이터 보기 소스를 선택하면 세부사항을 입력하도록 프롬프트됩니다.

분석 데이터 소스 리포지토리에서 분석 데이터 소스를 선택하십시오.

테이블 이름 사용될 데이터가 있는 테이블을 선택하십시오. 이는 분석 데이터 보기의 데이터 모델에 있는 테이블입니다. 데이터 모델의 각 테이블은 예측 분석 프로세스에 포함된 개념 또는 엔티티를 나타냅니다. 테이블에 대한 필드는 테이블에 의해 표시되는 엔티티의 속성에 해당합니다. 예를 들어, 고객 주문을 분석 중인 경우 데이터 모델에 고객에 대한 테이블 및 주문에 대한 테이블이 포함될 수 있습니다. 고객 테이블에는 고객 ID, 연령, 성별, 재무 상태 및 거주 국가에 대한 필드가 있을 수 있습니다. 주문 테이블에는 주문 ID, 주문 항목 수, 총 비용, 해당 주문을 넣은 고객의 ID에 대한 필드가 있을 수 있습니다. 고객 ID 필드는 고객 테이블의 고객을 주문 테이블의 주문과 연관시키는 데 사용할 수 있습니다.

데이터 액세스 계획 사용될 데이터가 있는 데이터 액세스 계획을 선택하십시오. 데이터 액세스 계획은 분석 데이터 보기의 데이터 모델 테이블을 실제 데이터 소스와 연관시킵니다. 분석 데이터 보기에는 일반적으로 다중 데이터 액세스 계획이 포함됩니다. 사용 중인 데이터 액세스 계획을 변경할 경우, 프로젝트별로 사용되는 데이터를 변경합니다. 예를 들어, 분석 데이터 보기에 모델 교육을 위한 데이터 액세스 계획과 모델 테스트를 위한 데이터 액세스 계획이 포함된 경우 사용 중인 데이터 액세스 계획을 변경하여 교육 데이터에서 테스트 데이터로 전환할 수 있습니다.

선택적 속성 사용 가능하고 원하는 경우 데이터 소스에 추가할 선택적 속성을 선택하십시오.

입력 필드 지정. 자세한 정보는 44 페이지의 『입력 필드 선택』 주제를 참조하십시오.

IBM Cognos BI 소스

IBM Analytical Decision Management에서 사용하기 위해 IBM Cognos BI 데이터베이스 데이터 또는 단일 목록 보고서를 가져올 수 있습니다.

참고: IBM SPSS Modeler Server가 IBM Cognos BI 서버를 지원하지 않는 IBM Analytical Decision Management와 함께 사용 중인 경우 이 옵션은 사용할 수 없습니다.

IBM Cognos BI 서버를 선택하고 연결을 클릭하면 서버 연결 세부사항을 입력하도록 프롬프트됩니다.

1. 데이터를 내보내거나 가져올 IBM Cognos 서버의 서버 URL을 입력하십시오. 어느 URL을 사용할지 모르겠는 경우에는 Cognos 시스템 관리자에게 문의하십시오.
2. 연결에 사용할 모드를 선택하십시오. 특정 사용자(예: 관리자)로 로그인하려면 신임 정보 설정을 선택하고 IBM Cognos 네임스페이스, 사용자 이름 및 비밀번호를 입력하십시오.
 - 서버에 로그인하는 데 사용되는 IBM Cognos 보안 인증 제공자 네임스페이스를 입력하십시오. 인증 제공자는 사용자, 그룹 및 역할을 정의하고 유지하며 인증 프로세스를 제어하는 데 사용됩니다.
 - 서버에 로그인하는 데 사용할 IBM Cognos 사용자 이름을 입력하십시오.
 - 지정된 사용자 이름과 연관된 비밀번호를 입력하십시오.

또는

IBM SPSS Collaboration and Deployment Services Repository에 저장된 IBM Cognos 신임 정보를 찾아보고 선택하십시오.

3. 또는, 사용자 신임 정보 없이 로그인하려면(이 경우 네임스페이스, 사용자 이름 및 비밀번호 필드를 지정할 수 없음) 익명 연결 사용을 선택하십시오.

참고: 일부 서버 연결에서는 익명 연결을 허용하지 않습니다.

Cognos 오브젝트 세부사항 선택

IBM Cognos 서버 연결을 설정한 후에는 모드를 선택하십시오. 데이터 모드는 데이터를 가져올 패키지를 찾아보고 선택할 수 있게 해 줍니다. 목록 보고서 모드는 데이터를 가져올 사전 정의된 목록 보고서를 찾아보고 선택할 수 있게 해 줍니다.

패키지는 IBM Cognos 모델, 그리고 해당 모델과 연관된 모든 폴더, 쿼리, 보고서, 보기, 단축키, URL 및 작업 정의를 포함합니다. IBM Cognos 모델은 비즈니스 규칙, 데이터 설명, 데이터 관계, 비즈니스 차원 및 계층 구조, 그리고 기타 관리 작업을 정의합니다.

1. 패키지를 선택하려면 데이터 모드를 선택하고 찾아보기를 클릭하십시오. 콘텐츠를 가져올 수 있는 모든 사용 가능한 패키지를 나열하는 대화 상자가 표시됩니다.
2. 또는 하나의 목록 보고서를 선택하려면 목록 보고서 모드를 선택하고 찾아보기를 클릭하십시오. 보고서를 선택할 수 있는 패키지와 폴더를 나열하는 대화 상자가 표시됩니다. 참고: 하나의 목록 보고서만 가져올 수 있습니다. 복수의 목록은 지원되지 않습니다.
3. 필요한 경우에는 목록을 펼쳐 필요한 패키지 또는 보고서를 찾으십시오. 항목을 선택하고 확인을 클릭하십시오. 패키지 또는 보고서 세부사항이 데이터 소스 편집기에 표시됩니다.
4. 패키지를 선택한 후에는 가져올 오브젝트를 선택하십시오. 예를 들어, 패키지에서 사용할 패키지 오브젝트 쿼리 개체(데이터베이스 테이블을 나타냄) 또는 개별 필드 쿼리 항목(테이블 열을 나타냄)을 선택하십시오. 참고: 데이터는 UTF-8 형식이어야 합니다.
5. 원시 데이터가 아니라 통합 데이터를 가져오려면 가져오기 전에 데이터 통합 옵션을 선택하십시오. 통합 데이터의 레코드 수는 일반적으로 원시 데이터보다 적습니다.

오브젝트 선택. 패키지 또는 목록 보고서에서 선택할 수 있는 오브젝트(네임스페이스라고도 함)를 표시합니다. 가져올 수 있는 입력 필드를 표시하려면 하나 이상의 오브젝트를 펼치십시오. 오브젝트를 선택하면 이에 속한 입력 필드가 자동으로 모두 선택됩니다.













매개변수 검토. 데이터 오브젝트나 보고서에 대해 매개변수가 정의된 경우에는 오브젝트나 보고서를 가져오기 전에 이러한 매개변수에 대해 값을 지정할 수 있습니다. 이 테이블은 필터와 계산에 필요한 매개변수의 목록을 포함합니다. 이들 항목에는 기본값이 있을 수 있습니다. 하나 이상의 매개변수에 기본값이 없는 경우에는 경고가 표시됩니다. 값 열에 적절히 매개변수를 설정하십시오. 데이터를 가져오기 전에 매개변수를 사용하여 조정(예: 매개변수화된 계산을 수행하기 위해)을 수행할 수 있습니다.

필터 검토. 패키지에 필터가 정의된 경우에는 하나 이상의 필터를 가져올 수 있습니다. 가져오는 필터가 가져온 데이터와 연관된 경우에는 해당 데이터를 가져오기 전에 이미 이 필터가 적용되어 있습니다. 필터는 매개변수 화되거나 되지 않을 수 있습니다.

입력 필드 검토. 처리를 위해 IBM Analytical Decision Management로 가져올 입력 필드(데이터베이스 오브젝트라고도 함)를 나열합니다. 특정 필드를 필요로 하지 않는 경우에는 필드 이름 옆에 있는 선택란을 선택 취소하십시오.

Cognos 오브젝트 아이콘: Cognos BI 데이터베이스에서 가져올 수 있는 다양한 유형의 오브젝트는 다음 표에 나타나 있는 바와 같이 서로 다른 아이콘으로 표현됩니다.

표 2. Cognos 오브젝트 아이콘.

아이콘	오브젝트
	패키지
	네임스페이스
	쿼리 제목
	쿼리 항목
	측도 차원
	측도
	차원
	레벨 계층 구조
	레벨
	필터
	보고서
	독립형 계산

Analytic Server

새 데이터 소스 작성 시 Analytic Server 데이터 소스 유형을 선택할 수 없습니다. 이는 표시 전용인데, 이 경우 Analytic Server 데이터 소스에 반하여 IBM SPSS Modeler에서 작성된 모델을 열기 위해 IBM SPSS Modeler Advantage를 사용합니다.

IBM SPSS Modeler 16에 IBM SPSS Analytic Server이 도입되었으므로 데이터 소스로 HDFS(Hadoop Distributed File System)를 사용할 수 있습니다. IBM SPSS Modeler 사용자는 Hadoop 대형 데이터 시스

템에 기반하여 HDFS Analytic Server 소스 노드 및 빌드 모델을 사용하여 스트림을 작성할 수 있습니다. IBM Analytical Decision Management 사용자는 IBM SPSS Modeler Advantage의 해당 모델 스트림을 열 수 있습니다(Analytic Server 데이터 소스를 통해 새 프로젝트를 작성할 수 없음). 모델 유형 또는 대상 변경 등 특정한 조치는 수행할 수 없습니다.

데이터 탭에서 Analytic Server 데이터 소스를 확인할 때 다음 정보가 있습니다.

데이터 소스 이름 IBM SPSS Modeler의 데이터 소스에 제공된 이름입니다. 원하는 경우에는 이 이름을 변경할 수 있습니다. 뒤따르는 나머지 필드는 편집할 수 없습니다.

Analytic Server 소스 IBM SPSS Modeler에서 구성된 대로 데이터 소스의 경로입니다.

입력 필드 지정. 입력 필드가 사용되었습니다.

입력 필드 선택

많은 양의 데이터를 보유하고 있는 경우 소스에서 가져온 데이터를 줄이거나 세밀하게 선별할 수 있습니다. 예를 들어, 관심 분야와 관련된 필드만 사용할 수 있습니다.

1. 데이터 소스 편집기 대화 상자의 맨 아래에서 **입력 필드 지정**을 클릭하십시오. 대화 상자가 펼쳐지며 소스 내에 있는 모든 데이터 필드가 표시됩니다. 또한 작성하는 모든 표현식이 표시됩니다. 표현식 작성에 대한 자세한 내용은 48 페이지의 『표현식 빌더』의 내용을 참조하십시오.
2. 사용할 각 필드 및 표현식을 선택하십시오.

데이터를 선택하는 것 외에, 필요한 경우에는 하나 이상의 데이터 필드의 측정 수준을 변경할 수도 있습니다. 자세한 정보는 45 페이지의 『측정 수준』의 내용을 참조하십시오.

데이터 미리보기

올바른 데이터 소스를 선택했는지 확인하려면 **미리보기** 아이콘을 클릭하여 소스에 포함된 데이터의 샘플을 표시하십시오.

데이터 개요

데이터 개요 기능은 모든 애플리케이션의 데이터 탭, IBM SPSS Modeler Advantage의 모델링 탭 및 스코어 탭, 그리고 지금 스코어링 기능이 포함된 애플리케이션의 배포 탭에서 사용 가능합니다.

1. 데이터 소스의 개요를 실행하려면 애플리케이션 내에서 사용 가능한 데이터 개요 아이콘을 클릭하십시오.



그림 5. 데이터 개요 아이콘

2. 데이터 개요 대화 상자가 표시됩니다. 원하는 경우에는 개요 옵션 섹션의 드롭 다운에서 오버레이 필드를 선택하십시오.

개요를 실행한 후에는 기본 선택된 필드의 결과만을 표시하는 탭, 또는 이 드롭 다운에 지정된 필드로 기본 선택된 필드를 오버레이하는 탭을 사용할 수 있습니다. 예를 들어, 결과에서 고객의 나이에 대한 정보를 보고, 이를 성별과 같은 필드로 오버레이할 수 있습니다.

- 원하는 경우에는 개요 옵션 섹션에서 파티셔닝된 데이터 사용을 선택하십시오. 이 옵션은 모델링 탭의 선택적 설정에서 평가 및 테스트를 위해 빌드 데이터 소스에 대한 모형 평가를 할 수 있도록 자동으로 데이터를 파티셔닝 옵션을 선택한 경우 사용 가능합니다. 이 옵션은 훈련 및 모델의 테스트를 위해 데이터를 별도의 서브세트 또는 샘플로 분할합니다. 하나의 서브세트에 모델을 빌드하고 이를 다른 서브세트에 테스트하면 이것이 다른 데이터 세트에 어떻게 적용될지 알 수 있습니다.
- 개요에 포함시킬 데이터 필드를 선택하고 개요 실행을 클릭하십시오. 데이터 소스에서 사용 가능한 모든 필드가 표시됩니다. 기본적으로는 모든 모델 입력 필드 및 대상이 선택되어 있습니다.

데이터 개요가 실행되어 결과가 표시됩니다. 정보를 정렬하고 표시할 열을 선택할 수 있습니다.

- 필드의 세부사항을 보려면 해당 필드를 클릭하십시오. 선택한 각 필드에 대해 새 결과 탭이 열려, 차트 및 테이블을 보고 사용 가능한 경우에는 오버레이 필드를 선택할 수 있게 해 줍니다.

측정 수준

측정 수준은 각 데이터 필드가 나타내는 정보를 규정하는 데 도움을 주며 임의의 필드가 규칙, 모델링 또는 다른 애플리케이션에서 어떻게 사용될지 결정합니다. 측정 수준은 데이터 소스의 프로젝트 데이터 모델에서 지정할 수 있습니다. 예를 들어, 0 = *False*, 1 = *True*로 나타내기 위해 0과 1 값을 갖는 정수 필드의 측정 수준을 플래그로 설정할 수 있습니다. 또는, 사용될 입력 필드를 지정하는 경우에는 데이터 소스 편집기 대화 상자에서 수준을 변경할 수 있습니다. 자세한 정보는 44 페이지의 『입력 필드 선택』의 내용을 참조하십시오.

다음 측정 수준을 사용할 수 있습니다.

- 연속형 0 - 100 또는 0.75 - 1.25의 범위와 같은 숫자 값을 설명하는 데 사용됩니다. 연속형 값은 정수, 실수 또는 날짜/시간이 될 수 있습니다.
- 범주형 고유 값의 정확한 수치를 알 수 없을 때 문자열 값에 사용됩니다. 이는 인스턴스화되지 않은 데이터 유형이며 저장 공간과 데이터 사용에 대한 모든 가능한 정보가 아직 알려지지 않았음을 의미합니다. 데이터를 읽고 나면 측정 수준이 플래그, 명목 또는 유형 없음이 됩니다. 프로젝트 특성 대화 상자에서 지정된 명목 필드에 대한 최대 멤버 수에 따라 .
- 플래그 true와 false, Yes와 No 또는 0과 1처럼 특성의 유무를 나타내는 두 개의 고유 값을 갖는 데이터에 사용됩니다. IBM Analytical Decision Management에서는 데이터 탭에 표시된 값 위에 커서를 두면 표시되는 도구 팁에서 볼 수 있듯이, 마우스를 카테고리의 첫 번째 값이 "false" 값으로 사용되며 두 번째 값이 "true" 값으로 사용됩니다. 데이터는 텍스트, 정수, 실수, 날짜, 시간 또는 시간소인으로 표현될 수 있습니다.
- 명목 small/medium/large와 같이 여러 고유 값을 가지며 각각의 값이 세트의 멤버로 취급되는 데이터를 설명하는 데 사용됩니다. 명목 데이터의 저장 공간은 숫자, 문자열 또는 날짜/시간 중 어느 것도 될 수 있습니다. 측정 수준을 명목으로 설정했다고 해서 자동으로 값이 문자열 저장 공간으로 변경되는 것은 아니라는 점을 참고하십시오. 명목 필드에 대해 허용된 최대 멤버 수를 설정하는 방법에 대한 정보는 32 페이지의 『특성』의 내용을 참조하십시오.

- 순서 여러 고유 값을 가지며 본질적으로 순서가 있는 데이터를 설명하는 데 사용됩니다. 예를 들어, 봉급 카테고리나 만족 순위를 순서 데이터 유형으로 지정할 수 있습니다. 순서는 데이터 요소 본연의 정렬 순서에 따라 정의됩니다. 예를 들어, 1, 3, 5는 정수 세트의 기본적인 정렬 순서이며 HIGH, LOW, NORMAL(영문자 순으로 오름차순)은 문자열 세트의 순서입니다. 순서 측정 수준은 시각화, 모델 작성, 또는 순서 데이터를 고유 유형으로 인식하는 다른 애플리케이션(예: IBM SPSS Statistics)에 내보낼 수 있도록 범주형 데이터 세트를 순서 데이터로 정의할 수 있게 해 줍니다. 명목 필드가 사용될 수 있는 위치에서는 어디든 순서 필드를 사용할 수 있습니다. 또한 모든 저장 공간 유형(실수, 정수, 문자열, 날짜, 시간 등)의 필드는 순서로 정의할 수 있습니다.
- 유형 없음 위의 유형 중 어느 것도 따르지 않는 데이터, 하나의 값이 있는 필드 또는 세트에 정의된 최대값보다 많은 멤버가 있는 명목 데이터에 사용됩니다. 이는 측정 수준이 멤버가 많은 세트(예: 계정 번호)가 되는 경우에도 유용합니다. 필드에 대해 유형 없음을 선택하는 경우 역할은 자동으로 없음으로 설정되며 유일한 대안으로 레코드 ID가 설정됩니다. 세트의 기본 최대 크기는 250개의 고유 값입니다. 이 개수는 도구 메뉴에서 액세스할 수 있는 프로젝트 특성 대화 상자는 도구 모음 아이콘에서 액세스할 수 있습니다.
- 컬렉션 목록에 기록되는 비지형정보 데이터를 식별하는 데 사용됩니다. 컬렉션은 실질적으로 깊이가 0인 목록 필드이며 해당 목록의 요소에는 다른 측정 수준 중 하나가 있습니다.
- 지형정보 지형정보 데이터를 식별하기 위해 목록 저장 공간 유형과 함께 사용됩니다. 목록은 정수 목록 또는 목록 깊이가 0에서 2 사이인 실제 목록 필드 중 하나일 수 있습니다.

데이터 필드 값 편집

데이터 테이블에서 값 열은 데이터 소스에 있는 각 필드의 값을 표시합니다(예: 범위의 최대값과 최소값). 플래그 필드의 경우 필드에 대해 정의된 "true" 및 "false" 값을 표시하는 도구 팁을 표시하려면 값 위에 마우스 포인터를 놓으십시오. 값을 클릭하여 편집할 수도 있습니다. 대화 상자가 열리고 해당 콘텐츠는 선택한 데이터 필드의 유형에 따라 다릅니다. 예를 들어, 다음과 같습니다.

- 필드 유형이 연속형인 경우에는 상한 및 하한 값을 입력할 수 있습니다.
- 필드 유형이 플래그인 경우에는 true와 false 값을 스왑할 수 있습니다.
- 필드 유형이 명목인 경우에는 값을 추가하고, 제거하고, 다시 정렬하고, 값의 이름을 바꿀 수 있습니다.

필드 매핑

새 데이터 소스를 추가할 때는 프로젝트 데이터 모델과 호환되도록 하기 위해 여기에 속한 모든 필드를 매핑해야 할 수 있습니다. 예를 들어, 프로젝트 데이터 모델에서 예 또는 아니오(측정 수준 플래그)의 값을 갖는 구매라는 이름의 필드를 필요로 할 경우에는 사용되는 데이터 소스에 호환 가능한 필드가 있어야 합니다. 필드 이름이 동일하지 않은 경우에는 적절히 매핑할 수 있습니다. 입력, 그리고 매핑된 연관 필드의 데이터 유형은 같아야 한다는 점을 참고하십시오.

또는, 프로젝트 데이터 모델이 변경되는 경우에는 많은 데이터 소스가 더 이상 올바르지 매핑되지 않을 수 있습니다. 이러한 경우에는 프로젝트 데이터 소스 영역의 호환성 표시기 원이 이러한 각 데이터 소스에 대해 적절한 색상으로 변경됩니다. 이 때는 이러한 모든 데이터 소스를 새 프로젝트 데이터 모델로 다시 매핑해야 합니다.

1. 프로젝트 데이터 소스 영역에서 필드 매핑을 클릭하십시오. 대화 상자가 열리며 시스템에서 자동으로 필드 매핑을 수행합니다. 사용자는 언제든지 다시 설정을 클릭하여 시스템이 자동으로 수행한 필드 매핑으로 돌아갈 수 있습니다.
2. 사용 가능한 입력열에서 필드를 선택하고, 그에 해당하는 필드를 필수 입력 필드 열에서 선택한 후 매핑을 클릭하여 두 필드를 매핑하십시오. 사용 가능한 필드 열에 있는 모든 필드가 매핑될 때까지 이를 반복하십시오.
3. 완료한 후에는 확인을 클릭하십시오.

가능한 경우 IBM Analytical Decision Management에서는 사용자가 데이터 소스를 추가할 때 호환 가능한 스토리지 유형을 자동으로 매핑합니다. 호환 가능하지 않은 유형인 경우 사용자에게 유형 변환이 허용됩니다.

다음 테이블에는 허용된 매핑이 나열됩니다.

표 3. 변경할 수 있는 유형

To → From ↓	Int	Real	String	Date	Time	timestamp
Int	해당 사항 없음	자동	예	허용되지 않음	허용되지 않음	허용되지 않음
real	예(소수점 자리수 잘림)	해당 사항 없음	예	허용되지 않음	허용되지 않음	허용되지 않음
String	예	예	해당 사항 없음	예	예	예
Date	허용되지 않음	허용되지 않음	예	해당 사항 없음	허용되지 않음	예(시간 00:00)
Time	허용되지 않음	허용되지 않음	예	허용되지 않음	해당 사항 없음	예(기준 날짜)
Timestamp	허용되지 않음	허용되지 않음	예	자동(시간 잘림)	자동(날짜 잘림)	해당 사항 없음

필드 파생

대부분의 IBM Analytical Decision Management 애플리케이션에서 표현식, 모델 결과, 세그먼트 규칙에서 작성된 필드 또는 보조 테이블의 필드와 같은 추가 필드를 프로젝트 데이터 모델에 추가할 수 있습니다. 데이터 탭의 파생 탭은 데이터의 일부로 포함되는 모든 필드를 표시하며 이 필드는 정렬이 가능하므로, 누적하여 사용할 수 있습니다. 테이블에서의 순서는 계산되는 순서와 일치합니다. 예를 들어, 두 번째 파생 필드는 첫 번째 파생 필드를 입력으로 사용할 수 있습니다.

파생 탭은 IBM SPSS Rules Management 애플리케이션에 포함되지 않습니다.

- 데이터 탭에서 파생 탭을 클릭하십시오.

표현식을 추가하려면 다음과 같은 작업을 수행하십시오.

- 표현식을 사용하여 필드 파생 아이콘을 클릭하십시오. 표현식 빌더가 표시됩니다. 자세한 정보는 48 페이지의 『표현식 빌더』의 내용을 참조하십시오.

모델 결과를 추가하려면 다음 작업을 수행하십시오.

1. 모델 또는 규칙을 사용하여 필드 파생 아이콘을 클릭한 후 모델을 찾아보고 선택하십시오. 선택된 모델에 대한 입력 및 결과 대화 상자가 표시됩니다.

2. 모델 입력 필드가 호환 가능한 경우에는 대화 상자의 모델 입력 맵핑 섹션이 접혀져 있습니다. 모델 입력 필드가 호환 가능하지 않은 경우 이 섹션을 펼쳐 사용 가능한 테이블 입력을 선택된 모델의 입력에 맵핑해야 합니다.
3. 모델 결과 섹션에 선택된 모델의 사용 가능한 모든 모델 결과가 표시됩니다. 데이터에 추가할 모델 결과 필드를 선택하십시오. 모델 결과 이름 옆은 모델에 의해 생성된 결과 필드의 이름을 표시합니다.

모델 결과의 모든 필드는 선택할 수 있습니다. 모델 결과 필드는 스트림에서 나오면서 스트림의 입력이 아닌 모든 필드(스트림에 의해 추가된 모든 필드)입니다. 입력 필드는 프로젝트 데이터 모델 내의 모든 것을 포함하지 않을 수 있으므로 모델 결과 필드의 이름이 프로젝트 데이터 모델 필드의 이름과 같을 수 있습니다. 따라서 원하는 경우에는 데이터 필드 이름을 여기서 변경할 수 있습니다.

수치 옆은 적절한 유형 아이콘이 앞에 붙어 있는 모델 결과의 수치 유형을 표시하며 값 옆은 모델 결과의 값을 표시합니다. 이들 옆은 일부 필드의 결과를 표시하고 일부는 표시하지 않을 수 있음을 참고하십시오. 보통 수치 유형 및 값은 데이터 스캔 없이 실제 모델 결과(예측된 값, 확률 등)에 대해 알려져 있으므로 알려진 경우에는 이것이 표시됩니다. 그러나 스트림의 다른 결과 필드(예: 파생된 필드)에 대한 수치 유형 또는 값은 데이터 스캔 없이는 알려지지 않습니다.

4. 모델 매개변수 섹션은 선택된 스트림이 스코어링 매개변수를 포함할 경우 대화 상자의 맨 아래에 포함됩니다. 모델 매개변수에 대한 값을 입력하십시오.
5. 완료한 후에는 대화 상자에서 저장을 클릭하십시오. 선택된 모델 결과가 파생 탭의 추가 필드 섹션에 추가됩니다.

세그먼트 규칙을 사용하여 필드를 작성하는 방법

1. 세그먼트 규칙을 사용하여 필드 파생 아이콘을 클릭하십시오.
2. 새 세그먼트 규칙을 작성하십시오. 세그먼트는 true 또는 false 결과를 리턴하는 하나 이상의 표현식에 의해 정의됩니다. 자세한 정보는 53 페이지의 『세그먼트 규칙 정의』의 내용을 참조하십시오.
3. 대화 상자에서 완료했다면 확인을 클릭하십시오. 새 세그먼트 규칙이 파생 탭의 추가 필드 섹션에 추가됩니다.

보조 테이블에서 필드를 추가하려면 다음 작업을 수행하십시오.

1. 데이터 탭에서 일대일 관계로 보조 테이블을 정의한 경우 보조 테이블에서 필드 추가 아이콘을 클릭하십시오.
2. 드롭 다운에서 보조 테이블을 선택한 후 하나 이상의 테이블 필드를 선택하십시오.
3. 완료한 후에는 대화 상자에서 저장을 클릭하십시오. 필드가 파생 탭의 추가 필드 섹션에 추가됩니다.

표현식 빌더

애플리케이션에서 사용할 추가 필드나 속성을 파생하기 위해 표현식 빌더를 사용할 수 있습니다. 예를 들어, 은행 업무 데이터를 사용하고 있는 경우에는 고객의 수입과 해당 고객이 보유한 대출 계좌 수 사이의 비율을 보여주는 표현식을 작성하려 할 수 있습니다.

표현식은 운영 데이터와 동일한 방식(예: 규칙이나 모델 예측변수에서)으로 애플리케이션 전반에서 사용 가능합니다.

기존 표현식을 표시하거나 새 표현식을 작성하려면 데이터 탭으로 이동하여 **파생** 탭을 클릭하십시오.

표현식 작성

1. 표현식 빌더 대화 상자에 표현식의 고유 이름을 입력하십시오.
2. 함수 및 필드 목록을 참조로 사용하여 표현식 방정식 필드에 입력하십시오.

또는

3. 목록에서 원하는 필드 및 함수를 선택하십시오.
4. 파란색 화살표를 두 번 클릭하거나 클릭하여 표현식 방정식 필드에 필드 또는 함수를 추가하십시오.
5. 대화 상자 가운데의 피연산자 단추를 사용하여 표현식에 연산을 삽입하십시오.
6. 표현식의 용도를 설명하는 설명을 입력하십시오.
7. 확인을 클릭하여 표현식의 유효성을 검증한 후 저장을 클릭하십시오.

참고: IBM SPSS Modeler 클라이언트에서 표현식을 복사하여 붙여넣을 수 있지만 표현식이 스트림 종속적인 경우 결과가 항상 예상대로 나오는 것은 아닙니다. 예를 들어, 지원되는 @ 함수 또는 행 함수를 붙여넣을 경우 나머지 스트림이 이를 지원하도록 설정되어 있지 않았을 수 있으므로 원하는 결과가 나오지 않을 수 있습니다.

함수 선택

함수 목록은 사용 가능한 모든 함수 및 연산자를 표시합니다. 스크롤하여 목록에서 함수를 선택하거나 쉽게 검색하려면 드롭 다운 목록을 사용하여 함수 또는 연산자 서브세트를 표시하십시오. 사용 가능한 함수는 쉽게 검색할 수 있도록 범주로 그룹화됩니다.

다음은 범주에 대한 세부사항입니다. 자세한 정보는 IBM SPSS Modeler 문서를 참조하십시오.

표 4. 함수 범주

함수 유형	설명
최근 사용 함수	현재 세션 중 사용되는 함수 목록이 있습니다.
연산자	표현식을 빌드할 때 사용할 수 있는 모든 연산자를 표시합니다. 연산자는 대화 상자 가운데의 단추에서도 사용할 수 있습니다.
정보 함수	필드 값을 자세히 확인하는 데 사용됩니다. 예를 들어, is_string 함수는 유형이 문자열인 모든 레코드에 대해 True를 리턴합니다.
변환 함수	새 필드를 구성하거나 저장 공간 유형을 변환하는 데 사용됩니다. 예를 들어, to_timestamp 함수는 선택된 필드를 시간소인으로 변환합니다.
비교 함수	필드 값을 서로 비교하거나 지정된 문자열과 비교하는 데 사용됩니다. 예를 들어, <=는 두 필드의 값이 작거나 같은지 비교하는 데 사용됩니다.
논리 함수	if, then, else 연산과 같은 논리 연산을 수행하는 데 사용됩니다.
숫자 함수	필드 값의 자연 로그와 같은 숫자 계산을 수행하는 데 사용됩니다.
무작위 함수	무작위로 항목을 선택하거나 숫자를 생성하는 데 사용됩니다.
문자열 함수	지정된 문자를 제거할 수 있도록 stripchar과 같은 문자열에서 다양한 연산을 수행하는 데 사용됩니다.

표 4. 함수 범주 (계속)

함수 유형	설명
날짜 및 시간 함수	날짜, 시간, 시간소인 필드에서 다양한 연산을 수행하는 데 사용됩니다.
공백 및 널 함수	사용자 지정 공백 또는 시스템 누락 값에 액세스하고 플래그를 지정하고 자주 채우는 데 사용됩니다. 예를 들어, @BLANK(FIELD)는 공백이 있는 레코드에 대해 True 플래그를 지정하는 데 사용됩니다.
모든 함수	사용 가능한 전체 함수 목록이 있습니다.

필드 및 값 선택

필드 목록은 사용 가능한 모든 필드를 표시합니다. 스크롤하여 목록에서 필드를 선택하십시오. 파란색 화살표를 두 번 클릭하거나 클릭하여 표현식에 필드를 추가하십시오.

값 보기 및 선택

필드 값은 애플리케이션의 여러 위치에서 볼 수 있습니다. 표현식 빌더에서 필드의 값을 보려면 원하는 필드를 선택하고 표현식 입력 단추를 클릭하여 선택된 필드의 값을 표시하는 대화 상자를 여십시오. 그런 다음, 값을 선택하고 삽입을 클릭하여 현재 표현식에 값을 붙여넣을 수 있습니다.



그림 6. 표현식 입력 단추

플래그 및 명목 필드의 경우 정의된 모든 값이 표시됩니다. 연속형(숫자 범위) 필드의 경우 최소값과 최대값이 표시됩니다.

표현식 확인

표현식 빌더(오른쪽 하단)에서 확인을 클릭하여 표현식의 유효성을 검증할 수 있습니다. 확인되지 않은 표현식은 빨간색으로 표시됩니다. 오류를 찾은 경우 원인을 나타내는 메시지가 표시됩니다.

다음 항목을 확인합니다.

- 값 및 필드 이름의 올바른 인용
- 매개변수 및 글로벌 변수의 올바른 사용
- 연산자의 올바른 사용
- 참조된 필드의 존재
- 참조된 글로벌의 존재 및 정의

구문에서 오류가 발견된 경우 표현식을 수동으로 입력하지 말고 목록 및 연산자 단추를 사용하여 표현식을 작성하십시오. 이 방법은 필드 및 값의 적합한 인용을 자동으로 추가합니다.

IBM Analytical Decision Management에서 표현식을 빌드할 때 다음 제한사항을 주의하십시오. 표현식은 다음 항목을 포함할 수 없습니다.

- IBM SPSS Modeler 스트림 매개변수에 대한 참조

- IBM SPSS Modeler 스트림 글로벌에 대한 참조
- 데이터베이스 함수에 대한 참조
- 다음 특수 필드 또는 필드 값 @ 함수 중 하나에 대한 참조:
 - @TARGET
 - @PREDICTED
 - @FIELD
 - @PARTITION_FIELD
 - @TRAINING_PARTITION
 - @TESTING_PARTITION
 - @VALIDATION_PARTITION

제 5 장 규칙 작업

규칙은 비즈니스 논리나 예측 모델의 결과를 기반으로 하여 의사결정을 자동화하는 데 사용됩니다. 애플리케이션에 따라, 지불 히스토리를 기반으로 고객을 제외하거나, 사기 가능성이 있는 클레임을 특수 수사기관에 의뢰하거나, 예측 모델에 의해 식별된 바와 같이 이탈, 구매 또는 사기를 저지를 성향이 가장 높은 고객을 대상으로 하기 위해 규칙을 사용할 수 있습니다. 규칙은 자동화를 가능케 해 주는 반면 모델은 예측 능력을 제공하며, 이들을 하나로 결합시키는 기능은 다른 분석 도구로부터 예측 애플리케이션을 차별화시켜주는 주요 기능 중 하나입니다.

규칙은 각 애플리케이션 내의 다양한 위치에서 작성되고, 적용되고 저장될 수 있습니다. 또는, IBM SPSS Rules Management를 사용하여 공유 규칙을 작성하고 편집할 수 있습니다.

규칙 작성

- 애플리케이션 내에 규칙을 추가할 지점에서 새 규칙 작성 아이콘을 클릭하십시오.
- 또는, 기존 규칙 찾기 아이콘을 클릭하여 현재 애플리케이션에 추가하거나 재사용할 수 있는 기존 규칙을 찾아보십시오. 자세한 정보는 56 페이지의 『내보내기 및 재사용 규칙』 주제를 참조하십시오.
- 공유 규칙을 작성하려면 애플리케이션 시작 페이지에서 IBM SPSS Rules Management를 시작하십시오.

세그먼트 규칙 정의






세그먼트 규칙은 다른 애플리케이션에 있는 레코드를 적절히 선택하고 할당하기 위해 애플리케이션 전반에서 사용됩니다. 세그먼트는 true 또는 false 결과를 리턴하는 하나 이상의 표현식에 의해 정의됩니다. 예를 들어, 나이 필드의 값이 18 미만일 경우 true를 리턴하는 하나의 세그먼트가 있는 규칙이 있을 수 있습니다. 세그먼트는 데이터에 정의된 필드 또는 표현식을 기반으로 하거나 모델이 리턴한 성향 스코어 또는 예측을 기반으로 하여 정의할 수 있습니다. 여러 차원 계층 구조(요소라고 하는)를 갖도록 구성된 정의 탭이 있는 애플리케이션에서는 다른 차원 멤버에서 선택할 수도 있습니다.

각 세그먼트는 하나 이상의 표현식에 의해 정의되며 모든 조건이 만족되면 true 값을 리턴합니다. 예를 들어, 나이와 성별로 정의된 세그먼트는 두 개의 조건을 모두 만족시키는 레코드만 포함하며 규칙은 다른 모든 항목에 대해 false를 리턴합니다.

세그먼트 규칙을 작성하거나 편집하는 방법

1. 규칙을 추가할 애플리케이션에서 새 규칙 작성 아이콘을 클릭하십시오.
2. 규칙의 이름을 지정하십시오.
3. 입력 열의 경우에는 아래로 화살표 아이콘을 클릭하여 필드, 규칙 또는 모델 중 원하는 것을 지정하십시오.

표 5. 입력 열 옵션.

아이콘	사용법
	사용으로 설정되면 이전에 정의된 요소에서 차원 멤버를 선택할 수 있게 해 줍니다.
	현재 데이터 세트에서 필드를 선택할 수 있게 해 줍니다.
	리포지토리에서 기존 모델 또는 규칙을 선택할 수 있게 해 줍니다.
	새 모델을 작성할 수 있게 해 줍니다.
	표현식을 작성할 수 있게 해 줍니다.

4. 사용할 기능을 선택하십시오. 사용 가능한 옵션은 다음 표를 참조하십시오.

표 6. 사용 가능한 연산자.

값 유형	사용 가능한 연산자
시간소인, 날짜, 날짜 시간	<ul style="list-style-type: none"> • =(같음) • !=(같지 않음) • >(보다 큼) • <(보다 작음) • >=(이상임) • <=(이하임) • 사이 • ISNIL • IS NOT NIL
범위	<ul style="list-style-type: none"> • =(같음) • !=(같지 않음) • >(보다 큼) • <(보다 작음) • >=(이상임) • <=(이하임) • 사이 • ISNIL • IS NOT NIL • 중의 하나 • 중의 하나가 아님

표 6. 사용 가능한 연산자 (계속).

값 유형	사용 가능한 연산자
세트	<ul style="list-style-type: none"> • =(같음) • !=(같지 않음) • ISNIL • IS NOT NIL • 중의 하나 • 중의 하나가 아님
플래그	<ul style="list-style-type: none"> • =(같음) • !=(같지 않음) • ISNIL • IS NOT NIL

5. 입력 유형의 값을 적절히 지정하십시오. 숫자가 아닌 값의 경우에는 선택사항의 드롭 다운 목록이 제공됩니다.

또는, 기존 규칙 찾기 아이콘을 클릭하여 이미 정의된 규칙을 찾아보십시오. 자세한 정보는 56 페이지의 『내보내기 및 재사용 규칙』 주제를 참조하십시오.

Age BETWEEN 50 and 20 등의 세그먼트 규칙은 결과를 리턴하지 않습니다. 표현식 $x \leq \text{value and value} < y$ 와 동일한데, 이는 True일 수 없습니다. 이는 SQL Server 작동과 일치합니다. Age BETWEEN 20 and 50 이 올바른 형식입니다.

세그먼트 규칙 세트

- 세그먼트 규칙은 세트로 결합될 수도 있습니다. 세그먼트 규칙 세트는 개별 세그먼트 중 하나가 true인 경우 true 값을 리턴합니다. 이는 세그먼트 내부의 표현식과는 다르다는 점을 참고하십시오. 이는 모두 true여야 합니다.
- 세그먼트 규칙 세트는 규칙을 사용하는 애플리케이션에서 내보낼 수 있으며, 사용 가능한 경우에는 IBM SPSS Rules Management에서 이를 수행할 수 있습니다. 자세한 정보는 56 페이지의 『내보내기 및 재사용 규칙』 주제를 참조하십시오.

선택 규칙 정의

선택 규칙은 하나 이상의 세그먼트 규칙의 세트 또는 각각에 포함 또는 제외 속성이 지정된 규칙 세트입니다. 선택 규칙은 애플리케이션 전반에서 모델링, 스코어링 또는 기타 처리를 위해 레코드의 서브세트를 선택하는데 사용됩니다.

1. 규칙을 추가할 애플리케이션에서 새 규칙 작성 아이콘을 클릭하십시오. 이 아이콘은 선택 규칙을 적용할 수 있는 위치에서 사용할 수 있습니다.
2. 규칙의 이름을 지정하고 포함이나 제외 어느 쪽을 위해 규칙을 사용할 것인지 선택한 후(또는 이를 나중에 수행) 하나 이상의 표현식을 원하는 대로 추가하십시오. 자세한 정보는 53 페이지의 『세그먼트 규칙 정의』 주제를 참조하십시오.

3. 규칙을 저장하려면 **확인**을 클릭하십시오.

4. 각 규칙에 대해 이 설정을 전환하려는 경우에는 **포함/제외** 열의 아이콘을 클릭하십시오. 참조된 선택 규칙 세트의 경우에는 이 설정을 변경할 수 없습니다. 같은 규칙 내에서는 포함과 제외를 결합할 수 있습니다.

또는, 기존 규칙 찾기 아이콘을 클릭하여 이미 정의된 규칙을 찾아보십시오. 자세한 정보는 『내보내기 및 재사용 규칙』 주제를 참조하십시오.

선택 규칙 평가

선택이 이뤄지는 방법(논리 AND 대 OR)에 대한 세부사항을 보려면 선택 규칙 논리 보기 아이콘을 클릭하십시오.

- 선택 규칙을 평가할 때, 여러 포함 세그먼트가 지정된 경우 레코드는 하나의 세그먼트가 일치하면 포함됩니다. 예를 들어, 고객은 신용 등급 또는 가입 기간에 따라 포함될 수 있습니다.
- 여러 제외 세그먼트가 지정된 경우 레코드는 아무 세그먼트나 일치하면 제외됩니다. 예를 들어, 고객은 나이 또는 신용 등급에 따라 제외될 수 있습니다.
- 충돌이 있을 경우에는 제외 규칙이 우선순위를 갖습니다. 예를 들어, 제외된 세그먼트에 속하는 고객은 다른 세그먼트의 적격 여부에 관계없이 제외됩니다.
- IBM SPSS Rules Management에서 선택 규칙 세트를 작성할 때 모든 규칙의 유형은 포함 또는 제외로 같아야 합니다. 같은 규칙 세트 내에서는 포함과 제외 규칙을 결합할 수 없습니다.

내보내기 및 재사용 규칙

기본적으로 애플리케이션에서 작성된 규칙은 해당 애플리케이션에 저장됩니다. 또는, 여러 애플리케이션에서 사용할 수 있도록 규칙을 IBM SPSS Collaboration and Deployment Services Repository로 내보내거나 여기에 저장할 수 있습니다. 이 경우 각 애플리케이션은 특정, 레이블 지정된 공유 규칙 버전에 대한 참조를 저장합니다. 규칙에 대한 모든 변경사항은 해당 버전을 사용하는 모든 애플리케이션에 적용됩니다.

내보내기 규칙

공유하려는 로컬 규칙이 포함된 애플리케이션에서 내보내기 아이콘을 클릭하여 규칙을 필요할 때 재사용할 수 있는 IBM SPSS Collaboration and Deployment Services Repository에 저장하십시오.

- 여러 규칙을 내보낼 때는 하나의 규칙 세트에 이를 모두 내보낼지, 또는 개별 규칙으로 내보낼지 지정하십시오. 개별 규칙의 경우 각 세그먼트는 별도의 규칙으로 내보내게 됩니다.
- 여러 선택 규칙을 내보낼 때는 선택과 함께 저장을 선택하면 선택 규칙 세트를 내보냅니다. 이 경우에는 포함 또는 제외를 내보낼 수 있으나, 두 가지를 모두 내보낼 수는 없습니다. 포함과 제외가 모두 필요한 경우에는 이를 각각 별도의 규칙 세트로 내보내야 합니다. 선택과 함께 저장이 선택되지 않은 경우에는 세그먼트 규칙 세트를 내보냅니다.
- 통합 규칙 편집기에서 내보낼 때는 위험 점과 함께 내보내기를 선택하면 통합 규칙 세트가 작성됩니다. 이 옵션이 선택되지 않은 경우에는 세그먼트 규칙 세트를 내보냅니다.

기존 규칙 사용

- 공유 규칙을 사용할 애플리케이션에서 기존 규칙 찾기 아이콘을 클릭하여 현재 애플리케이션에서 재사용할 수 있는 기존 규칙을 IBM SPSS Collaboration and Deployment Services Repository에서 찾아보십시오.
- 특정 레이블 지정된 버전을 참조하려면 규칙 이름 옆에 있는 화살표를 클릭하고 원하는 버전을 선택하십시오. 버전이 선택되지 않으면 최신 버전이 사용됩니다. (최신은 각 새 버전이 작성되면 자동으로 해당 버전으로 이동되는 특수 레이블이라고 생각하십시오.) 규칙의 새 버전이 작성되었을 때 발생할 수 있는 예기치 않은 결과를 방지하려면 특정 레이블 지정된 버전을 사용하는 것이 좋습니다.

참조된 규칙은 여러 경우에 로컬 규칙과 결합될 수 있으나 쿼은체, 그리고 규칙 또는 규칙 세트의 유형을 나타내는 아이콘을 사용하여 로컬 규칙과 시각적으로 구분됩니다.

- 또는, 원하는 경우에는 참조된 규칙을 로컬 규칙으로 변환하여 현재 애플리케이션에 저장할 수 있습니다. 이를 수행하려면 참조된 규칙의 이름을 클릭한 후 규칙 미리보기 대화 상자에서 단일 규칙으로 변환을 선택하십시오.

공유 규칙에 대한 가이드라인

공유 규칙을 찾아보고 이에 대해 작업할 때는 다음 가이드라인이 적용됩니다.

- 어떤 애플리케이션에서든 재사용하기 위해 규칙을 내보낼 수 있으나 일단 내보낸 후에는 다른 애플리케이션에서 편집할 수 없으며 IBM SPSS Rules Management에서만 편집할 수 있습니다. IBM SPSS Rules Management가 모든 곳에 설치되어 있는 것은 아니며 이를 사용할 수 없는 사용자도 있다는 점을 참고하십시오.
- 최신 버전의 변경사항이 의도하지 않은 결과를 발생시킬 수 있으므로, 특히 생산 환경에서는 특정 레이블된 버전(최신 이외)을 사용하는 것이 좋습니다. 규칙 내에서 규칙을 참조할 때와 같은 특정 경우에는 반드시 레이블된 버전을 사용해야 합니다.
- 단일 규칙은 참조를 통해 IBM SPSS Collaboration and Deployment Services Repository에서 추가되거나 미리보기 대화 상자에서 로컬 규칙으로 변경될 수 있습니다.
- 단일 규칙은 규칙 편집기에서 참조하거나 로컬 규칙과 결합할 수 있습니다. 규칙 세트로 내보낸 경우에는 단일 규칙에 대한 중첩된 참조가 유지됩니다.
- 세그먼트 및 선택 규칙 세트는 대부분의 경우 단일 규칙과 같이 참조되고 결합될 수 있으나, 결합된 규칙 세트는 먼저 중첩된 규칙 세트를 단일 규칙으로 변환하지 않으면 내보낼 수 없습니다. 따라서 내보낸 규칙 세트에는 중첩된 규칙에 대한 참조가 포함되어 있을 수 있으나 중첩된 규칙 세트에 대한 참조는 없습니다.
- 통합 규칙 세트 또한 참조될 수 있으나 이는 다른 규칙과 결합할 수 없으며 다른 규칙에 대한 중첩된 참조를 포함할 수 없습니다. 통합 규칙 세트를 참조할 때 참조된 규칙은 통합 규칙 편집기의 전체 콘텐츠를 대체합니다. (또는, 통합 규칙 세트를 참조하지 않는 경우에는 통합 편집기는 로컬 규칙, 참조된 세그먼트 규칙 및 세그먼트 규칙 세트를 모두 포함할 수 있습니다.)
- IBM SPSS Rules Management를 사용할 때는 내보낸 규칙과 같은 제한사항이 적용됩니다. 중첩된 규칙은 허용되지만 중첩된 규칙 세트는 허용되지 않습니다. 이는 IBM SPSS Rules Management가 공유 규칙을 작성하고 편집하도록 설계되었기 때문입니다(다른 애플리케이션에서 내보내는 것과 유사).

- IBM SPSS Rules Management 또는 다른 위치에서, 규칙 세트는 참조된 규칙 세트의 제한사항을 적용받지 않도록 하기 위해 항상 개별, 로컬 규칙으로 변환될 수 있습니다.
- 공유 규칙은 외부 IBM Operational Decision Management 규칙에 대한 참조를 포함할 수 없습니다.

주석 추가

통합 또는 할당에서, 규칙 결과에 포함될 참고사항을 입력할 수 있는 자유 텍스트 필드의 옆을 추가하려면 주석 추가 아이콘을 클릭하십시오. 예를 들어, 이는 보험 클레임이 의외하도록 플래그 지정된 경우 누구에게 연락해야 하는지 콜 센터 직원에게 알려주는 추가 정보가 될 수 있습니다.

주석은 추가 세부사항을 제공하기 위해, 차원에 지정한 다른 대상과 함께 리턴 값과 조합하여 사용할 수 있습니다. 다른 대상과 함께 리턴 값은 리턴된 차원 또는 의사결정과 연관되어 있는 반면, 주석은 해당 의사결정의 입력으로 어떤 규칙이 실행되었는지 알려줍니다.

IBM Operational Decision Management의 규칙 사용

IBM Operational Decision Management와 같은 비즈니스 규칙 관리 시스템에서 작성된 규칙은 IBM Analytical Decision Management 애플리케이션에서 참조되고 사용됩니다. 이러한 외부 규칙은 애플리케이션의 원하는 위치에서 외부 규칙 찾기 아이콘을 클릭하여 다른 규칙과 같은 방식으로 선택하고 사용할 수 있습니다. 외부 규칙은 아이콘이 다르지만 표준 방식으로 사용할 수 있습니다.

IBM Analytical Decision Management에서 사용할 외부 규칙을 작성하는 데 대한 정보는 *애플리케이션 디자이너 안내서*를 참조하십시오.

이 규칙은 다른 규칙과 같은 방식으로 애플리케이션에 삽입됩니다. 규칙 결과는 규칙이 사용된 컨텍스트와 일치해야 한다는 점을 참조하십시오. 예를 들어, 할당은 의사결정의 입력으로 true/false 또는 예/아니오(부울) 값을 요구합니다.

다른 규칙과 마찬가지로 입력과 필드, 규칙 설명 및 기타 특성을 포함, 애플리케이션에 사용된 외부 규칙의 세부사항을 보려면 이 규칙의 이름을 클릭하십시오.

외부 규칙에 여러 결과 필드가 있는 경우에는 원하는 필드를 선택할 수 있게 해 주는 프롬프트가 표시됩니다.

모델 대신 외부 규칙이 사용된 경우에는 대상 필드가 표시되지 않으며 구성된 모든 결과가 드롭 다운 목록에 표시됩니다.

외부 규칙 사용에 대한 팁

- *true* 또는 *false* 값(또는 예/아니오나 0/1)을 리턴하는 규칙은 모든 선택 및 할당에서, 규칙 편집기에서 규칙의 입력으로, 또는 특정 차원 항목과 연관된 다른 대상과 함께 리턴 필드로 사용할 수 있습니다.
- 범주형 결과(적색, 흰색, 파란색 또는 북, 남, 동, 서 등)를 리턴하는 규칙은 정의 탭의 결정하는 데 모델 사용 섹션에서, 규칙 편집기에서 입력으로, 또는 차원 항목과 연관된 다른 대상과 함께 리턴 필드로 사용할 수 있습니다.

- 숫자 결과를 리턴하는 규칙은 정의 탭의 결정하는 데 모델 사용 섹션에서, 우선순위의 입력으로, 규칙 편집기에서 입력으로, 또는 차원 항목과 연관된 다른 대상과 함께 리턴 필드로 사용할 수 있습니다.
- 분석 데이터 보기 데이터 소스는 복합(계층 구조) 외부 규칙의 사용을 지원합니다. 복합 외부 규칙을 참조하려면 프로젝트 데이터 모델이 데이터 보기 소스여야 하며 기능을 사용 가능해야 합니다. 관리자는 데이터 탭에서 적절한 계층 구조 데이터에 기반하여 **ODM** 규칙 사용 허용 옵션을 선택해야 합니다.
- 외부 규칙에 대한 참조가 포함된 IBM Analytical Decision Management에서는 공유 규칙을 작성할 수 없습니다.

제 6 장 예측 모델 작성

예측 모델 정보

모델은 이전 데이터의 패턴을 기반으로 미래에 일어날 일을 예측하는 데 사용할 수 있습니다. 예를 들어, 모델을 사용하여 수입, 나이, 가입한 조직이나 멤버십과 같은 특성을 기반으로 이탈할 가능성이 가장 낮은 고객, 또는 특정 오피에 반응할 가능성이 가장 높은 고객을 예측하려 할 수 있습니다.

모델은 규칙과 같은 방식으로 사용될 수 있으나, 규칙은 회사 정책, 비즈니스 논리 또는 기타 가정을 기반으로 하는 반면 모델은 과거 결과의 실제 관측을 기반으로 작성되며 쉽게 파악할 수 없는 패턴을 발견할 수 있습니다. 규칙은 애플리케이션에 일반적인 비즈니스 논리를 부여하지만 모델은 통찰력과 예측 능력을 제공합니다.

모델 작성 및 스코어링에 필요한 데이터

모델링 프로세스에는 두 가지 서로 다른 유형의 데이터가 사용됩니다.

- 모델을 작성하려면 예측할 항목에 대한 정보가 있어야 합니다. 예를 들어, 이탈을 예측하려는 경우에는 과거에 이탈한 고객에 대한 정보가 필요합니다. 이는 보통 히스토리 또는 분석 데이터라 하며 프로젝트 데이터 모델에 있는 필드의 일부 또는 전부, 그리고 예측할 결과를 기록할 추가 필드를 포함해야 합니다. 이 추가 필드는 모델링의 대상으로 사용됩니다.
- 미래의 결과를 예측하는 데 모델을 사용하려면 잠재적 고객이나 발생 클레임과 같은, 대상이 되는 그룹이나 모집단에 대한 정보가 있어야 합니다. 이는 보통 운영 데이터 또는 스코어링 데이터라고 합니다. 프로젝트 데이터 모델은 보통 이 데이터를 기반으로 합니다.
- 대상 필드가 데이터 모델에 포함된 경우 운영 열은 모델을 스코어링할 때 사용할 수 없으므로 이 필드에 대해 선택하지 말아야 합니다. 히스토리 데이터를 사용하여 모델을 작성할 때 스코어링의 목표는 결과를 알 수 없는 새 데이터를 모델에 적용하는 것입니다.

모델 작성

모델링 탭을 사용하면 데이터 소스, 대상 필드 및 기타 옵션을 지정할 수 있습니다. 작성할 모델 유형도 선택할 수 있습니다.

IBM SPSS Modeler에서 매개변수가 애플리케이션(스트림 파일)에 추가된 경우에는 모델링 탭 및 스코어 탭에서 매개변수 링크를 사용할 수 있습니다. 매개변수 링크를 클릭하면 정의된 매개변수와 각 매개변수의 현재 값이 표시되며 모델 작성 또는 스코어링에 영향을 주기 위해 매개변수 값을 변경할 수 있게 해 줍니다.

예측 모델 작성

예측 모델을 작성할 때는 데이터 소스와 대상 필드만 지정되어 있으면 됩니다. 데이터 준비 및 파티셔닝, 필드 선택과 사용할 레코드에 대한 옵션도 선택할 수 있습니다. 모델은 대상 유형에 여러 방법을 적절히 적용하고 선택된 데이터에 대한 최적의 특정 기법 또는 조합을 자동으로 식별하여 생성됩니다.

모델은 스코어링되면 예측이 포함된 하나 이상의 필드, 그리고 해당 예측과 연관된 신뢰도를 리턴합니다. 예를 들어, 대상 필드가 이름이 반응인 플래그 필드인 경우 모델은 이 필드에 대한 예측된 값이 포함된 *XF*-반응이라는 필드를 리턴합니다.

1. 데이터 소스를 지정하십시오. 이는 예측할 결과를 기록하는 분석 데이터 소스가 될 수 있습니다.
2. 예측할 결과를 기록하는 대상 필드를 지정하십시오. 예를 들어, 대상 필드는 이탈한 고객, 이전 오피에 반응한 고객, 사기 클레임을 건 고객 등을 나타낼 수 있습니다.

다른 모든 필드는 자동으로 이 값을 예측하는 데 유용한 입력으로 포함됩니다.

3. 선택적 설정을 원하는 대로 지정하십시오. 자세한 정보는 68 페이지의 『선택적 모델 설정』 주제를 참조하십시오.
4. 원하는 경우에는 데이터 개요 아이콘을 클릭하여 현재 모델을 작성하는 데 사용될 데이터의 개요를 보십시오. 자세한 정보는 44 페이지의 『데이터 개요』 주제를 참조하십시오.
5. 모델 작성을 클릭하십시오.

모델을 작성하는 중에는 브라우저를 닫거나 다른 태스크에서 작업을 수행할 수 있습니다. 모델은 작성이 완료되면 갤러리에서 액세스할 수 있습니다. 자세한 정보는 27 페이지의 『갤러리』 주제를 참조하십시오. 모델 작성에 예상보다 오랜 시간이 걸리는 경우에는 중지 단추를 클릭하여 이전에 저장된 버전으로 되돌릴 수 있습니다.

모델 작성이 완료되면 결과가 표시됩니다. **결합된 모델 결과**(현재 작성된 결합된 모델에 대한 성능 차트) 또는 **개별 모델 결과**(결합된 모델을 구성하는 개별 모델에 대한 높은 수준의 세부사항)를 보도록 선택할 수 있습니다.

개별 모델 결과를 볼 때, 각 모델의 이름은 대상 필드나 입력 필드와 같은 모델에 대한 정보를 포함하고 있는 새 대화 상자를 여는 링크입니다. 선택된 모델이 트리 모델인 경우에는 트리 뷰어도 표시됩니다.

6. 선택적으로, 평가 및 테스트 기능을 사용하여 샘플 데이터에 대한 모델의 성능을 확인하십시오. 자세한 정보는 72 페이지의 『모델 평가』 주제를 참조하십시오.

예측 모델링 결과

예측 모델이 완료되면 결과 섹션을 통해 **결합된 모델 결과**와 **개별 모델 결과** 보기 간에 전환할 수 있습니다. **결합된 모델 결과**에는 빌드 시에 결합된 모델에 대한 분포도와 변수중요도 차트가 포함되어 있어 최종 결과를 얻는 데 있어 모델 성능 및 다른 예측변수의 중요도에 대한 빠른 시각적 평가를 제공합니다. **개별 모델 결과**에는 결합된 모델을 구성하는 모든 개별 모델에 대한 정보가 나열되므로 각 개별 모델을 클릭하면 해당 세부 사항을 볼 수 있습니다.

분포도

모델 분포도는 관측된 값을 가로축에 표시하고 예측된 값을 세로축 또는 오버레이에 표시합니다. 이는 각 값 또는 반응이 예측된 빈도와 해당 예측이 적절한 빈도를 알려줌으로써 모델 정확도에 대한 간략한 시각적 평가를 제공합니다.

- 플래그 또는 세트 대상의 경우 분포는 막대형 차트를 사용하여 표시되며 각 막대는 대상 필드의 각 카테고리 리를 나타냅니다. 각 막대의 높이는 해당 결과가 모델에 의해 예측된 횟수를 나타냅니다. 각 막대 내부의 색칠된 영역은 검정 데이터에서 일치하는 결과가 실제로 관측된 횟수를 나타냅니다. 모델이 충분히 정확한 경우에는 각 막대 내의 가장 큰 영역이 해당 반응과 일치해야 합니다.
- 범위 대상의 경우에는 분포가 구간화된 산점도로 표시됩니다. 마찬가지로 예측된 값은 가로축에 표기되며 관측된 값은 세로축에 표기됩니다. 이상적인 모습은 점들이 대각선을 따라 군집되어 있는 것입니다. 점(결과)이 더 멀리 산포되어 선형분포에서 멀어질수록 모델의 정확성은 떨어집니다.

예측변수중요도 차트

일반적으로 사용자는 가장 중요한 예측변수 필드에 모델링 노력을 집중하고 가장 쓸모 없는 예측변수 필드를 삭제하거나 무시하고자 합니다. 예측변수 중요도 차트는 모델 추정 시 각 예측변수의 상대적 중요도를 표시하여 이러한 작업에 도움을 줍니다. 값은 상대적이므로 표시된 모든 예측변수에 대한 값의 합계는 1.0이 됩니다. 예측변수 중요도는 모형 정확도와는 관련이 없습니다. 이는 예측의 정확성과 상관이 없으며, 예측을 수행하는 데 관련된 각 예측변수의 중요도하고만 관련이 있습니다.

예측변수 중요도를 계산하는 데 소요되는 시간은 모델을 작성할 때보다 훨씬 길며, 특히 대형 데이터 세트를 사용할 경우에는 더욱 오래 걸립니다.

모델이 작성되었을 때 자동으로 데이터를 파티셔닝하는 옵션이 선택된 경우에는 테스트하는 파티션이 평가 페이지에서 자동으로 선택됩니다.

파티셔닝이 켜지지 않은 경우에는 모델을 평가하는 데 필요한 데이터 소스를 선택해야 합니다. 모델을 작성하는 데 사용되지 않은 데이터 소스를 선택하는 것이 좋습니다. 파티셔닝은 기본적으로 켜져 있으므로 이 설정을 변경하지 않은 한 이러한 작업은 필요하지 않습니다.

개별 모델 결과

개별 모델 결과는 결합된 모델을 구성하는 각 개별 모델에 대한 정보를 표시합니다. 각 모델의 이름은 대상 필드 및 입력 필드와 같은 모델에 대한 세부사항을 포함하고 있는 새 대화 상자를 여는 링크입니다. 선택된 모델이 트리 모델인 경우에는 특수 트리 뷰어도 표시됩니다. 트리 뷰어는 대화형으로, 다음 기능을 제공합니다.

- 트리에서 확대 및 축소
- 테이블, 그래프 또는 결합된 테이블 및 그래프로 트리 항목 표시
- 위에서 아래로 왼쪽에서 오른쪽으로 오른쪽에서 왼쪽으로 트리 방향 변경
- 트리 맵 창 및 화살표 아이콘을 사용하여 보다 쉽게 큰 트리 탐색

예측 규칙 모델 작성

예측 규칙 모델은 특정 결과에 대해 높거나 낮은 가능성을 나타내는 세그먼트를 식별할 수 있게 해 줍니다. 예를 들어, 이탈할 가능성이 가장 낮은 고객, 또는 특정 오퍼에 예라고 반응할 가능성이 가장 높은 고객을 찾으려 할 수 있습니다. 결과 모델은 선택 규칙의 목록과 유사하며 각 세그먼트가 스코어링되는 방식을 지정하는 포함 또는 제외 속성이 있습니다.

예측 규칙 모델은 선택 규칙과 유사하지만 스코어링되는 방식에 차이가 있습니다. 선택 규칙의 경우에는 제외된 레코드가 제거되어 결과의 레코드 수가 더 적습니다. 모델의 경우에는 레코드가 제거되지 않지만 각 레코드에 대한 예측 및 성향 값을 나타내는 추가 필드가 추가됩니다. 포함된 세그먼트는 예측하려는 것에 따라 *True* 또는 *False*로 스코어링됩니다. 제외된 세그먼트는 규칙과 일치하지 않아 나머지에 속하는 레코드와 마찬가지로 *NULL*로 스코어링됩니다.

각 레코드에는 대상 반응의 가능성을 나타내는 성향 스코어가 주어지 가장 반응할 가능성이 높은 고객을 식별할 수 있게 해 줍니다. 이와 같이 예측 규칙 모델은 메일링 목록을 생성하는 데 적합하며, 콜 센터 또는 마케팅 애플리케이션과 같은 고객 관계 관리에 널리 사용되고 있습니다.

- 예측 규칙 모델링은 범주형 대상에만 사용 가능합니다.
- 세그먼트를 보고, 수정하고 재구성할 수 있으며 스코어링 목적에 따라 어느 것을 포함시키고 제외할지 선택할 수 있습니다. 예를 들어, 미래의 오피에서 한 그룹의 고객을 제외시키고 나머지를 포함시킨 후 이러한 조치가 전체 적중 비율에 어떤 영향을 주는지 확인할 수 있습니다.
- 선택 규칙과 마찬가지로, 필요에 따라 수동으로 세그먼트를 추가하거나 편집할 수 있습니다. 그러나 보통은 모델이 세그먼트를 찾아주는 것이 훨씬 쉬우며 결과를 세세히 설정할 수 있도록 도와주는 몇 가지 옵션이 제공됩니다.
- 규칙은 나열된 순서대로 적용되며 첫 번째로 일치하는 규칙이 해당 레코드의 결과를 결정합니다. 개별적으로 따져보면 규칙이나 조건은 겹칠 수 있지만 규칙의 순서가 호호성을 해결해 줍니다. 일치하는 규칙이 없는 경우 레코드는 나머지 규칙으로 지정됩니다.

예측 규칙 모델을 얻는 방법

1. 데이터 소스를 지정하십시오. 이는 예측할 결과를 기록하는 분석 소스가 될 수 있습니다.
2. 예측할 결과를 기록하는 대상 필드를 지정하십시오.
3. 찾을 특정 값이나 반응을 지정하십시오. 예를 들어, 대상 필드의 이름이 반응인 경우에는 반응하거나 하지 않은 고객을 찾기 위해 *True* 또는 *False*를 각각 선택할 수 있습니다.

예를 들어, 대상은 이탈한 고객, 이전 오피에 반응한 고객, 사기 클레임을 건 고객 등을 나타낼 수 있습니다. 다른 모든 필드는 자동으로 이 값을 예측하는 데 유용한 입력으로 포함됩니다.

4. 선택적 설정을 원하는 대로 지정하십시오. 자세한 정보는 68 페이지의 『선택적 모델 설정』 주제를 참조하십시오.
5. 원하는 경우에는 데이터 개요 아이콘을 클릭하여 현재 모델을 작성하는 데 사용될 데이터의 개요를 보십시오. 자세한 정보는 44 페이지의 『데이터 개요』 주제를 참조하십시오.
6. 빌드 시작을 클릭하십시오.
7. 지정된 대상보다 높거나 낮은 확률의 세그먼트 중 어느 것을 찾을지 지정하십시오. 예를 들어, 이탈할 가능성이 가장 높거나 가장 낮은 그룹 또는 고객을 찾아 적절히 포함시킬 수 있습니다.
8. 찾을 세그먼트의 최대 수를 지정하십시오. 단순함과 성능을 위해 보통 이 수는 작게 지정합니다.
9. 모델 성장을 클릭하여 더 높거나 낮은 확률의 그룹을 식별하는 하나 이상의 세그먼트 규칙 목록을 얻으십시오.

10. 포함/제외 아이콘을 클릭하여 규칙이 사용되는 방식과 관련 세그먼트가 스코어링되는 방식을 지정하십시오.

포함된 세그먼트에는 대상 값과 일치하는 스코어가 부여됩니다(예: 반응=true). 제외된 세그먼트는 NULL로 스코어링되지만 파일에서 제거되지는 않습니다.

11. 선택적으로, 원하는 대로 세그먼트를 추가하거나 편집할 수 있습니다. 자세한 정보는 53 페이지의 『세그먼트 규칙 정의』 주제를 참조하십시오.
12. 선택적으로, 평가 및 테스트 기능을 사용하여 샘플 데이터에 대한 모델의 성능을 확인하십시오.

답:

- 규칙을 위나 아래로 이동하려면 위/아래 화살표를 사용하십시오. 각 레코드는 첫 번째 적용에 따라 스코어링되므로 규칙의 순서는 결과를 변경시킵니다. 예를 들어, 여러 규칙과 일치하는 레코드는 첫 번째로 적용되는 규칙에 따라 스코어링됩니다.
- 추가 규칙을 찾으려면 하나 이상의 옵션을 변경하고 모델 성장을 다시 클릭하십시오. 예를 들어, 더 많은 세그먼트를 식별하기 위해 최소 세그먼트나 신뢰구간에 대한 임계값을 낮출 수 있습니다.
- 모델 성장을 클릭한다고 해서 테이블이 지워지거나 모델이 처음부터 다시 작성되지는 않으며, 테이블에 추가할 세그먼트 규칙을 더 찾을 뿐입니다.

군집 모형 작성

군집 모형은 특정한 군집 수로 레코드를 분류할 수 있게 해 줍니다. 이는 데이터에서 서로 어울리는 그룹을 식별할 수 있게 도와줍니다.

군집 모형은 유사한 레코드의 그룹을 식별하고 레코드가 속하는 그룹에 따라 레코드에 레이블을 지정하는 것에 중점을 둡니다. 이는 각 그룹과 그 특성에 대한 사전 지식 없이 수행됩니다. 실은 찾아야 하는 그룹의 수조차 모를 수도 있습니다. 군집 모형이 다른 시스템 학습 기법과 다른 점은 모델이 예측할 사전 정의된 결과 또는 대상 필드가 없다는 점입니다. 모델의 분류 성능을 평가할 외부 표준이 없으므로 이러한 모델은 보통 비감독 학습 모델이라고 합니다. 이러한 모델에는 옳은 또는 그른 응답이 없습니다. 이 모델의 가치는 데이터 내에서 관심을 끄는 그룹을 잡아내고 이러한 그룹화에 대해 유용한 설명을 제공하는 것에 달려있습니다.

군집화 방법은 각 레코드 간, 그리고 각 군집 간의 거리 측정을 기반으로 합니다. 레코드는 동일한 군집에 속하는 레코드 간의 거리가 최소화되는 방향으로 군집에 지정됩니다.

군집 모형은 보통 후속 분석에서 입력으로 사용되는 군집이나 세그먼트를 작성하는 데 사용됩니다. 이러한 경우의 흔한 예로는 마케팅 담당자가 전체 시장을 동종 하위 그룹들로 파티셔닝할 때 사용되는 시장 세그먼트가 있습니다. 각 세그먼트에는 이를 대상으로 한 마케팅의 성공에 영향을 주는 특성이 있습니다. 마케팅 전략을 최적화하기 위해 데이터 마이닝을 사용하고 있는 경우에는 적절한 세그먼트를 식별하고 해당 세그먼트 정보를 예측 모델에서 사용함으로써 모델을 크게 개선할 수 있습니다.

군집 모형을 얻는 방법

1. 데이터 소스를 지정하십시오. 이는 예측할 결과를 기록하는 분석 소스가 될 수 있습니다.

2. 선택적 설정을 원하는 대로 지정하십시오. 자세한 정보는 68 페이지의 『선택적 모델 설정』 주제를 참조하십시오.
3. 원하는 경우에는 데이터 개요 아이콘을 클릭하여 현재 모델을 작성하는 데 사용될 데이터의 개요를 보십시오. 자세한 정보는 44 페이지의 『데이터 개요』 주제를 참조하십시오.
4. 군집 찾기를 클릭하십시오.
5. 선택적으로, 수동 군집을 추가할 수 있습니다. 자세한 정보는 『수동 군집 사용』 주제를 참조하십시오.
6. 선택적으로, 평가 및 테스트 기능을 사용하여 샘플 데이터에 대한 모델의 성능을 확인하십시오.

수동 군집 사용

원하는 경우에는 수동으로 군집을 정의한 후, 자동화된 군집화를 사용하여 수동으로 정의된 군집에 의해 식별되지 않은 레코드의 군집을 찾을 수 있습니다. "첫 번째 적중" 세그먼트 기반 할당 규칙 세트를 통해 수동 군집을 작성하십시오. 작성한 규칙은 군집을 정의하며 나머지는 자동 군집 모델링에 사용됩니다.

수동 군집은 결과 요약 차트에 포함되지 않는다는 점을 참고하십시오.

1. 수동 군집 섹션에서 새 규칙 아이콘 작성을 클릭하고 하나 이상의 규칙을 추가하여 수동 군집을 정의하십시오.
2. 원하는 경우에는 군집 레이블 이름을 변경하십시오.
3. 수동 군집에 포함된 레코드의 총계를 세려면 수동 군집의 레코드 아이콘을 클릭하십시오. 정의된 각 수동 군집에 대해 개수 열이 업데이트됩니다.

연관 모델 작성

연관 모델은 함께 나타날 가능성이 가장 높은 항목을 예측할 수 있게 해 주며 이들 사이에 있는 관계의 강도를 예측합니다.

연관 규칙은 특정 결과(예: 특정 제품의 구매)를 조건 세트(예: 다른 여러 제품의 구매)와 연관시킵니다. 예를 들어, 규칙

```
beer <= cannedveg & frozenmeal (173, 17.0%, 0.84)
```

는 *cannedveg*와 *frozenmeal*이 함께 발생하면 *beer*가 가끔씩 발생함을 나타냅니다. 이 규칙의 신뢰도는 84%이며 데이터의 17%(173개의 레코드)에 적용됩니다. 연관 규칙 알고리즘은 웹 노드와 같은 시각화 기법을 사용하여 수동으로 찾을 수 있는 연관을 자동으로 찾습니다.

보편적인 의사결정 트리 알고리즘(C5.0 및 C&R 트리)에 비해 연관 규칙 알고리즘이 갖는 장점은 어느 속성 간에도 연관이 존재할 수 있다는 점입니다. 의사결정 트리 알고리즘은 하나의 결과를 갖는 규칙을 빌드하는 데 반해 연관 알고리즘은 여러 규칙을 찾으려 시도하며 이들에게는 각각 다른 결과가 있을 수 있습니다.

연관 알고리즘의 단점은 패턴을 찾는 대상이 잠재적으로 매우 큰 검색 공간이 될 수 있으며 따라서 의사결정 트리 알고리즘을 실행하는 것보다 훨씬 많은 시간이 필요할 수 있다는 것입니다. 이 알고리즘이 규칙을 찾는 데 사용하는 방법은 생성 및 테스트로, 간단한 규칙이 처음에 작성되어 데이터 세트에 대해 유효성 검증됩니다. 유용한 규칙은 저장되며 모든 규칙은 그 후 다양한 제한조건에 따라 특화됩니다. 특화는 규칙에 조건을 추

가하는 프로세스입니다. 이러한 새 규칙은 데이터에 대해 유효성 검증되며 프로세스는 찾아낸 최선의, 또는 가장 관심을 끌 만한 규칙을 저장하는 것을 반복합니다. 일반적으로 사용자는 규칙에 허용될 수 있는 조건의 수에 한계를 지정하며 정보 이론이나 효율적 색인 작성 스키마에 기반한 여러 기법이 잠재적으로 커질 수 있는 검색 공간을 줄이기 위해 사용됩니다.

연관 모델을 얻는 방법

1. 데이터 소스를 선택하십시오. 이는 예측할 결과를 기록하는 분석 소스가 될 수 있습니다.
2. 데이터 형식을 지정하십시오. 자세한 정보는 『표 형식 대 트랜잭션 데이터』 주제를 참조하십시오.
3. 선택적 설정을 원하는 대로 지정하십시오. 자세한 정보는 68 페이지의 『선택적 모델 설정』 주제를 참조하십시오.
4. 원하는 경우에는 데이터 개요 아이콘을 클릭하여 현재 모델을 작성하는 데 사용될 데이터의 개요를 보십시오. 자세한 정보는 44 페이지의 『데이터 개요』 주제를 참조하십시오.
5. 모델 작성을 클릭하십시오.
6. 선택적으로, 테스트 기능을 사용하여 샘플 데이터에 대한 모델의 성능을 확인하십시오. 연관 모델 유형의 경우에는 평가를 사용할 수 없다는 점을 참고하십시오.

다른 모델 유형을 통해 다른 모델링 기법 중에서 선택할 수 있습니다. 연관 모델의 경우 *Apriori* 모델 유형이 사용됩니다.

표 형식 대 트랜잭션 데이터

연관 모델을 작성할 때는 표 형식 또는 트랜잭션 데이터 형식이 사용 가능합니다. 기본적으로 표 형식 데이터가 선택되어 있습니다. 트랜잭션을 선택할 경우에는 나타나는 드롭 다운 메뉴에서 **ID** 및 콘텐츠에 대한 필드를 선택하십시오.

트랜잭션 형식

트랜잭션 데이터에는 각 트랜잭션이나 항목에 대한 별도의 레코드가 있습니다. 예를 들어, 고객이 여러 개의 상품을 구매하는 경우 각각은 별도의 레코드가 되며 여기에는 고객 ID로 링크된 연관 항목이 있습니다. 이는 영수증 형식이라고도 합니다.

고객	구매
1	잼
2	우유
3	잼
3	빵
4	잼
4	빵
4	우유

표 형식 데이터

표 형식 데이터(장바구니 또는 진리표 데이터라고도 함)에는 별도의 플래그로 표현된 항목이 있으며 각 플래그 필드는 특정 항목의 유무를 나타냅니다. 각 레코드는 연관 항목의 완전한 세트를 나타냅니다. 플래그 필드는 카테고리 또는 숫자가 될 수 있으나 특정 모델에는 특정 요구사항이 있을 수 있습니다.

고객	잼	빵	우유
1	T	F	F
2	F	F	T
3	T	T	F
4	T	T	T

선택적 모델 설정

모델을 작성할 때는 다음 선택적 설정을 선택할 수 있습니다. 모델의 유형과 선택된 대상에 따라 이들 설정이 모두 나오지 않을 수도 있다는 점을 참고하십시오. 기존 모델에 대해 이들 설정을 변경할 경우 변경사항을 적용하려면 해당 모델을 다시 시작해야 합니다.

신뢰할 수 있는 모델 빌드를 위해 자동으로 데이터 정리 및 준비. 모델링을 더 빠르고, 예측할 수 있으며 신뢰할 수 있도록 만들기 위해 데이터 문제를 식별하고 복구합니다. 결측 또는 극단값을 처리하거나, 적절한 경우에는 새 속성을 파생하거나, 지능적인 선별 및 샘플링 기법을 통하여 성능을 향상시킴으로써 문제가 있거나 사용하지 않은 필드를 선별합니다. 새 데이터 소스를 사용하여 처음으로 모델이 작성되면 문제와 수정사항을 식별하는 분석이 수행되므로 처음에는 실행 속도가 느립니다. 후속 실행에서는 수정사항이 적용되지만 분석은 데이터 소스가 변경되지 않는 한 반복되지 않습니다. 이 설정은 숙련된 사용자에게 의해 지정된 사용자 정의 데이터 준비 설정이 있는 모델을 포함하여, 일부 모델의 경우 사용 안함으로 설정될 수 있습니다.

참고: 자동 데이터 준비 설정은 새 모델을 작성할 경우에만 적용되며 구간화는 수행되지 않습니다.

평가 및 테스트를 위해 빌드 데이터 소스에 대한 모형 평가를 할 수 있도록 자동으로 데이터를 파티셔닝. 이 옵션을 선택하면 훈련 및 모델의 테스트를 위해 데이터를 별도의 서브세트 또는 샘플로 분할합니다. 하나의 서브세트에 모델을 빌드하고 이를 다른 서브세트에 테스트하면 이것이 다른 데이터 세트에 어떻게 적용될지 알 수 있습니다. 임의로 생성되는 훈련 백분율 값 및 테스트 파티션 크기의 백분율 값을 지정할 수도 있습니다.

최소 훈련 파티션 크기는 1이며 최대값은 100입니다. 최소 테스트 파티션 크기는 0이며 최대값은 100입니다. 최소 검증 파티션 크기는 0이며 최대값은 100입니다. 모델이 성공적으로 작성되려면 전체 백분율은 100 이하가 되어야 합니다.

이러한 파티셔닝 옵션은 연관 모델링을 제외한 모든 모델 유형에 대해 사용할 수 있습니다.

모델 작성을 위한 모델 기법 선택. 원하는 경우에는 선택을 클릭하여 모델을 작성할 때 포함시킬 모델링 기법을 선택하십시오. 선택한 대상을 기반으로 모든 사용 가능한 모델링 기법이 나열됩니다. 사용 가능한 모델링 기법과 기본적으로 선택되는 기법은 선택한 대상에 따라 달라집니다. 예를 들어, 의사결정 목록 알고리즘은 플래그 대상에 대해서만 나타납니다. 이 기능은 예측 모델에 대해서만 사용할 수 있습니다.

모델링 기법에 대한 전체 세부사항은 *IBM SPSS Modeler 알고리즘 안내서*, 그리고 IBM SPSS Modeler 제품과 함께 제공되거나 웹에서 사용 가능한 기타 문서를 참조하십시오.

결합할 수 있는 최대 모델 수. 보유하고 결합할 수 있는 최대 모델 수를 설정할 수 있습니다. 이 옵션을 1로 설정하면 하나의 모델을 작성하고 보유할 수 있습니다. 여기에 입력된 값보다 적은 수의 모델을 작성한 경우에는 작성된 모든 모델을 보유하고 결합할 수 있습니다. 이 옵션은 예측 모델에 대해서만 사용할 수 있습니다.

수익 기준(2진 대상이 있는 모델의 순위를 지정하는 데 사용). 모델을 작성하기 전에 수익 기준을 설정할 수 있게 해 줍니다. 여기에 설정된 값은 모델 평가의 기본값으로 설정됩니다. 이 옵션은 플래그 대상이 선택된 예측 모델에 대해서만 사용할 수 있습니다.

사용할 입력 지정. 사용할 필드를 선택할 수 있게 해 줍니다. 보통 이는 나이 또는 수입과 같이 사용자가 예측하려는 항목과 실제적인 관계가 있는 필드입니다. 대형 데이터 세트가 있는 경우 필드의 수를 제한하는 것은 모델을 단순화시키는 방법 중 하나입니다. 데이터가 고객 ID 또는 연락처와 같은 필드를 포함하고 있는 경우 이는 보통 모델링에 있어서 유용하지 않으며 따라서 선택하지 않아야 합니다. 다른 데이터를 복제하는 필드 또한 제외해야 합니다.

링크된 입력 필드(표현식)를 클릭하면 해당 표현식에 대해 표현식 뷰어가 열립니다. 표현식을 편집하려면 데이터 탭을 보십시오. 자세한 정보는 48 페이지의 『표현식 빌더』 주제를 참조하십시오.

사용할 선택 지정. 모델링 시 포함시키거나 제외할 레코드를 지정하십시오. 기존 규칙을 검색하거나, 적절한 경우에는 새 규칙을 작성할 수 있습니다. 자세한 정보는 55 페이지의 『선택 규칙 정의』 주제를 참조하십시오.

또한, 글로벌 선택선이 정의된 경우에는 여기에 표시되며 모델링 중에 이를 적용할지 지정할 수 있습니다. 글로벌 선택선 규칙은 모두 사용하거나 전혀 사용하지 않도록 하는 두 가지만 선택할 수 있습니다. 서브세트는 선택할 수 없습니다.

세그먼트 옵션

예측 규칙 모델을 작성할 때는 몇 가지 세그먼트 옵션을 사용할 수 있습니다. 이들은 다른 유형의 모델에 사용 가능한 옵션 이외의 추가 옵션입니다.

최소 세그먼트 크기. 이 설정은 최소 세그먼트 크기를 결정하며 두 값 중 큰 쪽이 우선순위를 갖습니다. 예를 들어, 백분율 값이 절대값보다 큰 수로 계산될 경우에는 백분율 설정이 우선순위를 갖습니다.

- **이전 세그먼트의 백분율.** 레코드의 백분율로 최소 그룹 크기를 지정합니다. 허용된 최소 설정은 0이며 최대 설정은 99.9입니다.
- **절대값.** 레코드의 절대 수로 최소 그룹 크기를 지정합니다. 허용된 최소 설정은 1이며 최대 설정은 없습니다.

최대 속성 수. 세그먼트 규칙 당 최대 조건 수를 지정합니다. 허용된 최소 설정은 1이며 최대 설정은 없습니다.

속성 재사용 허용. 사용으로 설정되면 각 주기에서는 이전 주기에 사용된 속성을 포함하여 모든 속성을 고려할 수 있습니다. 세그먼트에 대한 규칙은 주기 내에서 빌드되며 각 주기에서는 새 조건을 추가합니다. 주기의 수는 최대 속성 수 설정을 사용하여 정의됩니다.

새 조건에 대한 신뢰구간. 세그먼트 유의수준 테스트를 위한 신뢰수준을 지정합니다. 이 설정은 리턴되는 세그먼트(있을 경우)의 수와 세그먼트 규칙당 조건의 수에 큰 영향을 줍니다. 값이 클수록 리턴되는 결과 세트는 적어집니다. 허용된 최소 설정은 50이며 최대 설정은 99.9입니다.

자동 군집 옵션

군집 모형을 작성할 때는 몇 가지 자동 군집 옵션을 사용할 수 있습니다. 이들은 다른 유형의 모델에 사용 가능한 옵션 이외의 추가 옵션입니다.

평가 필드. 평가에 사용할 필드를 지정합니다. 평가 필드가 선택된 경우 IBM SPSS Modeler 자동 군집 작성 노드의 순위화 방법은 평가의 중요성으로 설정됩니다. 평가 필드가 선택되지 않으면 순위화 방법은 기본값으로 남아있게 됩니다. 여기서 분석 필드가 평가 필드로 선택될 수 있다는 점을 참고하십시오.

검색할 군집 수에 대한 원하는 범위 설정. 자동 군집 모델에 의해 검색할 군집의 수에 대해 원하는 최소 및 최대 범위를 지정하십시오.

연관 규칙 및 스코어링 옵션

연관 모델을 작성할 때는 몇 가지 연관 규칙 및 스코어링 옵션을 사용할 수 있습니다. 이들은 다른 유형의 모델에 사용 가능한 몇 가지 옵션 이외의 추가 옵션입니다.

연관 규칙 옵션

- 범위(%). 규칙이 얼마나 일반적으로 적용 가능한가를 제어하는 최소 범위 백분율을 지정하십시오. 예를 들어, 값이 80%인 설정은 모든 규칙이 적어도 케이스의 80%에 적용된다는 뜻입니다. 값이 더 높아질수록 규칙은 더 적어지며, 또는 전혀 없을 수도 있음을 참고하십시오.
- 신뢰도(%). 최소 신뢰도 백분율을 지정하십시오.
- 최대 전제조건 수. 최대 전제조건 수를 지정하십시오.
- 데이터가 ID로 이미 정렬됨(트랜잭션 형식 한정). ID가 연속적이 되도록 하기 위해, 모델 작성 시작 시 기본적으로 수행되는 정렬을 사용 안함으로 설정하려면 이 옵션을 선택하십시오. 이 옵션을 선택하면 처리 시간이 줄어듭니다. 데이터 형식이 표 형식일 경우 이 옵션은 영향을 주지 않습니다.

스코어링 옵션

- 최대 예측 수. 포함된 예측의 최대 수를 지정하십시오. 이 설정은 "최상위" 예측을 작성하기 위해 규칙 기준 드롭 다운과 함께 사용되며 여기서 최상위란 가장 높은 수준의 신뢰도, 지원, 리프트 등을 나타냅니다.
- 규칙 기준. 규칙의 강도를 결정하는 데 사용되는 측도를 선택하십시오. 규칙은 항목 세트에 대한 최상위 예측을 리턴하기 위해 여기서 선택된 기준의 강도에 따라 정렬됩니다. 사용 가능한 기준은 다음과 같습니다.
 - 신뢰도는 규칙 조건 지원에 대한 규칙 지원의 비율을 표시합니다. 이는 규칙 예측 또한 true인, 지정된 규칙 조건을 가진 레코드의 비율을 나타냅니다. 예를 들어, 훈련 데이터의 50%가 빵(규칙 조건 지원을 나타냄)을 포함하지만 20%만 빵과 치즈(규칙 지원을 나타냄)를 모두 포함할 경우 규칙 bread -> cheese에 대한 예측은 Rule Support / Rule Condition Support가 되며, 지금의 경우에는 40%가 됩니다.

- 범위(%)는 규칙 조건 지원, 즉 훈련 데이터를 기반으로 규칙 조건이 true인 ID의 비율을 나타냅니다. 예를 들어, 훈련 데이터의 50%가 빵 구매를 포함하는 경우 rule bread -> cheese의 규칙 조건 지원은 50%가 됩니다.
 - 규칙 지원(%)은 전체 규칙, 규칙 조건 및 규칙 예측이 true인 레코드의 비율을 나타냅니다. 예를 들어, 훈련 데이터의 20%가 빵과 치즈를 모두 포함할 경우 bread -> cheese에 대한 규칙 지원은 20%가 됩니다.
 - 리프트는 규칙 예측을 갖기 이전의 확률에 대한 규칙의 신뢰도의 비율을 나타냅니다. 예를 들어, 전체 모집단의 10%가 빵을 구매하는 경우 20% 신뢰도로 사람들이 빵을 구매할지 예측하는 규칙의 리프트는 $20/10 = 2$ 가 됩니다. 사람들이 빵을 구매한다는 데 대해 신뢰도가 11%인 다른 규칙이 있는 경우 이 규칙의 리프트는 1에 가까우며, 이는 규칙 조건이 갖고 있는 것이 규칙 예측을 갖고 있는 것과 확률에서 큰 차이가 없음을 의미합니다. 일반적으로 1에서부터 리프트 차이를 가지는 규칙은 1에 가까운 리프트가 있는 규칙보다 더 흥미있습니다.
 - 배포 가능성 (%) 규칙 조건은 충족하지만 규칙 예측은 충족하지 않는 훈련 데이터의 백분율을 측정합니다. 제품 구매 측면에서 이는 기본적으로 조건은 갖고 있으나(또는 구매함) 아직 규칙 예측은 구매하지 않은 총 고객 기반의 백분율을 의미합니다. 배포성 통계는 $((\text{Rule Condition Support in \# of Records} - \text{Rule Support in \# of Records}) / \text{Number of Records}) * 100$ 으로 정의되며 여기서 Rule Condition Support는 규칙 조건이 true인 레코드의 수를 의미하고 Rule Support는 규칙 조건과 예측이 모두 true인 레코드의 수를 의미합니다.
- 전체 트랜잭션 확인. 다음 옵션 중 하나를 선택하십시오.
 - 예측이 존재하지 않는지 확인은 규칙 예측 또한 없는지 확인합니다. 예를 들어 스코어링 목적이 가정용 가구 제품을 추천하기 위한 것일 경우 이미 식탁이 들어있는 장비구니는 다른 것을 구매할 가능성이 적습니다. 이런 경우에는 이 옵션을 선택해야 합니다. 반면 제품이 치즈, 유아용 유동식 또는 휴지처럼 잘 상하거나 일회용일 경우에는 결과가 이미 장비구니에 있는 규칙에 가치가 있을 수 있습니다. 후자의 경우 가장 유용한 옵션은 예측을 확인하지 않음일 수 있습니다.
 - 예측이 존재하는지 확인은 규칙 예측이 장비구니에 존재하는지도 확인합니다. 이 접근법은 기존 고객 또는 트랜잭션에 대해 통찰력을 얻고자 할 때 유용합니다. 예를 들어 최고의 리프트를 가진 규칙을 확인한 다음 어떤 고객이 이러한 규칙에 적합한지 탐색하고 싶을 수 있습니다.
 - 예측을 확인하지 않음은 장비구니에 규칙 예측이 있는지 없는지에 관계없이 스코어링 시 모든 규칙을 포함합니다.
 - 반복 예측 허용. 스코어링 시 같은 규칙 예측이 있는 여러 규칙을 포함하려면 이 옵션을 선택하십시오. 예를 들어, 이 옵션을 선택하면 다음 규칙을 스코어링할 수 있습니다.


```
bread & cheese -> wine
cheese & fruit -> wine
```

스코어링 시 반복 예측을 제외하려면 이 옵션을 끄십시오.
 - 일치하지 않는 항목 무시. 항목 세트에 있는 추가 항목을 무시하려면 이 옵션을 선택하십시오. 예를 들어, 이 옵션이 [tent & sleeping bag & kettle]을 포함하는 장비구니에 대해 선택된 경우 규칙 tent & sleeping bag -> gas stove는 여분의 항목(kettle)이 장비구니에 있어도 적용됩니다.

모델 평가

완료된 모델을 작성하거나 연 후에 분포, 오류 분포, 일치 교차, Gains, Lift, Response, Profit 및 ROI에 대한 차트를 얻고 가능한 최대의 수익을 시뮬레이션하려면 평가를 클릭하십시오. 차트 유형 및 사용 가능한 기타 옵션은 선택된 모델 유형에 따라 달라집니다. 평가 대화 상자는 선택된 모델이 완료된 경우에만 사용 가능합니다. 연관 모델 유형의 경우에는 평가를 사용할 수 없다는 점을 참고하십시오.

1. 상단 패널에서 평가 데이터 세트를 선택하십시오. 모델 작성과 같이 이는 실제 반응이 알려진 분석 데이터 세트여야 합니다. (그렇지 않은 경우에는 모델에 의해 예측된 반응을 실제 반응과 비교할 방법이 없습니다.)
2. 평가에 포함시키기 위해 모델을 수정하려는 경우에는 모델 선택 수정을 클릭하십시오. 선택한 후 원하는 경우에는 뒤에 있는 현재 모델 선택을 개별 모델 결과 테이블로 복사 단추를 클릭하십시오. 이들 옵션은 예측 모델 유형에서만 사용 가능합니다.



그림 7. 현재 모델 선택사항을 개별 모델 결과 테이블에 다시 복사

3. 빌드할 그래프를 선택하십시오.
4. 수익을 시뮬레이션하는 경우에는 최대 수익 시뮬레이션 상자를 선택하고 수익 계산, 예산 및 최소값의 기준이 될 모집단을 입력하십시오. 자세한 정보는 73 페이지의 『최대 수익 시뮬레이션』 주제를 참조하십시오.
5. Profit 또는 ROI 차트를 요청하거나 수익을 시뮬레이션하는 경우에는 비용 및 수입 값을 지정하십시오. 이 값은 각 오퍼의 비용, 그리고 각 호의적 반응으로부터 거둘 수 있을 것으로 예상되는 수입액을 나타냅니다.
6. "적중"으로 간주되는 반응 값을 선택하십시오. 이는 긍정적 또는 호의적 반응을 나타내는 반응 = true와 같이, 모델을 작성하는 데 사용된 대상 필드의 값 중 하나여야 합니다.
7. 성능 차트를 요청하는 경우에는 성능 차트를 위한 개별 모델 평가 포함 옵션을 선택할 수도 있습니다.
8. 원하는 경우 표 형식 결과 포함을 선택하십시오. 평가 결과를 리턴하는 데는 추가 시간이 필요합니다.
9. 실행을 클릭하여 결과를 보십시오.

분포도(모든 대상)

분포도는 예상 값과 관측 값을 대비하여 기록되며 모델이 완료되면 기본적으로 표시됩니다. 평가 대화 상자에서는 점정 데이터 세트가 아닌 데이터 세트의 분포도를 얻을 수 있습니다. 범위 대상의 경우에는 분포도가 사용 가능한 유일한 차트 유형입니다.

다음 추가 차트 중 하나를 선택할 수 있습니다.

오류 분포도

오류 분포도는 모델 결과에 있는 오류의 히스토그램을 표시하며 여기서 오류는 예측 값과 실제 값 간의 차이입니다. 양의 오류는 예측 값이 실제 값보다 큰 경우 발생하며 음의 오류는 예측 값이 실제 값보다 적은 경우 발생합니다. 이 차트는 모델 유형이 예측 모델인 연속형 대상에 대해서만 사용 가능합니다.

일치 교차표

일치 교차표에는 예측 결과와 실제 결과를 대비한 수가 기록됩니다. 숫자 대상의 경우 범위는 지정한 구간 수로 분할됩니다.

Gains 차트

Gains 차트는 모델을 사용함으로써 얻을 수 있는 이익 또는 "리프트"를 나타내며 사용 가능한 히트의 총 수에 대한 각 증분의 적중 비율로 정의됩니다. 모델이 사용되지 않은 경우에는 전체 샘플에 대해 곧은 대각선 도표가 예상됩니다. 이 경우 한 사람이 반응할 가능성은 다른 사람과 차이가 없으므로 반응률은 상수가 됩니다. 결과를 두 배로 만드려면 두 배로 많은 사람에게 물어야 합니다. 곡선은 반응할 가능성이 가장 높은 사람들을 대상으로 함으로써 이 결과를 얼마나 향상시킬 수 있는가를 나타냅니다. 곡선이 가파를수록 이익이 더 많아집니다.

Lift 차트

Lift 차트에는 각 증분에서 적중인 레코드의 백분율이 훈련 데이터 세트의 전체 적중 백분율과 함께 기록됩니다.

Response 차트

Response 차트에는 증분에서 적중인 레코드의 백분율을 기록합니다.

Profit 차트

수익은 각 레코드의 수입에서 해당 레코드의 비용을 뺀 것과 동일합니다. 분위수의 수익은 분위수에 있는 모든 레코드의 수익의 합입니다. 수입은 적중에만 적용되는 것으로 가정되지만, 비용은 모든 레코드에 적용됩니다. 자세한 정보는 『최대 수익 시뮬레이션』 주제를 참조하십시오.

ROI 차트

ROI(투자수익률)는 수입과 비용을 정의하는 부분을 포함하는 점에서 수익과 유사합니다. ROI는 수익을 각 구간의 비용과 비교합니다.

최대 수익 시뮬레이션

모델이 예측한 반응률을 기반으로, 임의의 모집단에서 거둘 수 있는 최대 수익을 시뮬레이션할 수 있습니다. 이를 수행하려면 비용과 수입 값, 그리고 모집단의 크기를 지정해야 합니다.

최대 수익을 예측하려면 다음 작업을 수행하십시오.

1. 예측에서 기반으로 삼을 모집단 또는 레코드의 총계를 지정하십시오. 참고: 모델을 스코어링할 때는 이것이 자동으로 데이터 소스 또는 서브세트의 레코드 총계로 설정됩니다.
2. 수익 시뮬레이션에 예산 매개변수를 포함시키려는 경우에는 예산의 값을 지정하고 최소값을 지정하십시오.
3. 각 레코드와 연관된 비용을 지정하십시오. 예를 들어, 이는 캠페인에 대한 홍보 책자의 한 사람당 인쇄 및 발송 비용이 될 수 있습니다.
4. 긍정적 반응이 발생시킬 기대 수입을 지정하십시오. 캠페인에 속한 항목의 소매가를 예로 들 수 있습니다.
5. 흥미가 있는 반응을 선택하십시오.
6. 실행을 클릭하십시오.

예측된 결과 그래프는 최대 수익에 대한 최선의 결과를 표시하며, 모델에 의해 식별된 반응 성향을 기반으로 이러한 결과를 성취하기 위해 목표로 해야 하는 레코드의 백분율을 알려줍니다.

그래프 아래에서는 슬라이더 막대를 목표 레코드의 0에서 100%까지 이동할 수 있으며 이러한 조작이 수익에 미치는 영향을 확인할 수 있습니다.

수치가 만족스러운 경우에는 이 설정을 레코드 스코어링에 사용하도록 선택할 수 있습니다. 이를 수행하려면 스코어링에 대해 사용을 클릭하십시오. 목표 레코드의 백분율은 스코어 탭의 성향 옵션 영역에 있는 최상위 % 필드에 표시됩니다.

수익 비교

선택된 레코드가 바람직한 선택인지 확인하기 위해 이 레코드에서 거둘 수 있는 기대 수익을 동일한 백분율의 레코드가 설정된 무작위 그룹, 그리고 레코드 전체 세트와 비교할 수 있습니다.

표시된 세부사항에는 목표로 삼은 레코드의 수, 예상 반응률, 비용, 수입 및 전체 수익이 포함되어 있습니다. 이 정보는 다른 애플리케이션에서 사용할 수 있도록 HTML 형식으로 내보낼 수 있습니다.

Profit 차트

선택된 레코드에서 예측된 누적 또는 비누적 수익을 그래프로 표시할 수 있습니다.

참고: 이 그래프는 평가를 처음 실행했을 때 생성되며 수익 시뮬레이션 섹션의 슬라이더를 통해 이루어진 변경에 대해 자동으로 업데이트되지 않습니다.

모델 테스트

테스트 대화 상자는 하나 이상의 샘플 레코드가 어떻게 스코어링될지 보여줍니다. 결과는 테스트가 실행되는 컨텍스트에 따라 달라지며 모델이나 규칙이 리턴한 스코어를 반영할 수 있습니다. 표시할 수 있는 결과 유형에 대한 예제는 애플리케이션 유형에 따라 다음과 같습니다.

- 몇 명의 특정 고객을 보고 각각에게 어느 오퍼가 할당되는지 확인할 수 있습니다.
- 몇 개의 고객 레코드를 보고 DM 캠페인에 반응할 가능성이 높은 고객을 확인할 수 있습니다.
- 몇 개의 특정 보험 클레임을 보고 각각에게 어느 조치가 할당되는지 확인할 수 있습니다.

• 몇 가지 특정 케이스를 보고 각 규칙이 어떻게 평가하는지 확인할 수 있습니다.

1. 모델링 탭에서 테스트를 클릭하십시오. 테스트 대화 상자가 표시됩니다.
2. 사용할 데이터 소스를 선택하거나, 레코드 섹션에서 사용자 정의 데이터 지정을 선택하여 테스트 레코드를 수동으로 입력하십시오.
3. 사용할 선택이 있는 경우에는 지정하십시오. 예를 들어, 테스트에 특정 레코드를 포함하거나 이를 테스트에서 제외할 수 있습니다. 기존 규칙을 검색하거나, 적절한 경우에는 새 규칙을 작성할 수 있습니다. 자세한 정보는 55 페이지의 『선택 규칙 정의』 주제를 참조하십시오.
4. 데이터 소스를 사용하는 경우에는 리턴할 레코드의 수를 지정하십시오. 데이터를 맨 위에서부터 읽거나, 레코드의 무작위 표본에서 처음 N 개의 레코드(최대 100개)를 리턴할 수 있습니다. 트랜잭션 데이터 데이터를 사용하여 연관 모델을 테스트하는 경우에는 입력 데이터의 샘플링은 레코드 레벨이 아닌 트랜잭션 레벨에서 이뤄집니다.
5. 사용자 정의 데이터를 사용하는 경우에는 사용자 정의 데이터 지정을 선택하고 사용자 정의 레코드 추가를 클릭하십시오. 각 필드에 값을 지정할 수 있도록 프로젝트 데이터 모델에 정의된 모든 필드가 나열됩니다. 완료한 후에는 저장을 클릭하십시오.
6. 표시할 필드 선택에서 테스트 결과에 표시할 필드를 선택하십시오.
7. 실행 클릭을 클릭하십시오. 테스트 요약이 각 레코드에 대해 성공한 할당의 목록을 표시합니다.
8. 각 행의 세부사항을 보려면 해당 행의 시작 부분에 있는 정보 아이콘을 클릭하십시오.

애플리케이션에서 모델 스코어 사용

모델을 스코어링한다는 것은 의사결정에 대한 입력으로 사용할 수 있는 결과나 예측을 얻기 위해 이 모델을 데이터나 모집단에 적용한다는 것을 의미합니다. 예를 들어, DM 캠페인의 전망을 알아보기 위해 데이터베이스에 있는 모든 고객을 스코어링하거나, 고객이 콜 센터에 문의해 올 때마다 이 고객에게 제안할 제품이나 서비스를 결정하기 위해 해당 고객을 스코어링할 수 있습니다. 애플리케이션에 따라, 스코어링 결과는 데이터베이스 테이블 또는 플랫폼 파일에 기록될 수 있으며, 또는 애플리케이션에서 의사결정을 내리는 세그먼트, 선택 및 할당 규칙에 대한 입력으로 사용될 수 있습니다.

모델 스코어는 다음과 같은 방식으로 사용될 수 있습니다.

- 규칙에 대한 입력으로. 자세한 정보는 53 페이지의 『규칙 작성』 주제를 참조하십시오.
- 일괄 스코어링을 지원하는 애플리케이션의 경우 모델 스코어는 데이터베이스 테이블 또는 플랫폼 파일에 기록될 수 있습니다. 자세한 정보는 77 페이지의 『데이터베이스 테이블, 파일, Analytic Server 또는 IBM Cognos BI 서버에 모델 스코어링』 주제를 참조하십시오.

모델 결과 필드

규칙에 따라 스코어링 결과 필드의 이름은 대상 필드의 이름을 기반으로 하지만, 여기에는 사용자 정의 접두부가 추가됩니다. 예를 들어, 대상 필드의 이름이 반응으로 지정된 경우 결과 필드는 $\$XF$ -반응이나 $\$XFRP$ -반응과 같은 이름을 갖게 됩니다. 특정 필드는 아래에 자세히 나타나 있는 바와 같이 모델의 유형과 선택된 대상 필드의 측정 수준에 따라 달라집니다. 자세한 정보는 45 페이지의 『측정 수준』 주제를 참조하십시오.

플래그 대상. 플래그 대상이 있는 자동화된 모델은 다음 필드와 같은 결과를 출력합니다.

- **\$XF-<대상>**. 각 레코드에 대해 예측된 값으로, 필드에 대해 정의된 "true" 및 "false" 값의 용어로 보고됩니다.
- **\$XFRP-<대상>**. 성향 스코어로, 각 레코드에 대해 "true" 값의 가능성을 나타냅니다. 성향 스코어는 보통 해당 레코드에 대한 예측으로 보고되는 신뢰도 스코어와는 다릅니다. 예를 들어, 신뢰도가 높은 "false" 예측은 **반응하지 않을 가능성이 높다는 것을 나타내므로 낮은 성향으로 해석됩니다.**

범주형 대상. 범주형 대상(명목 또는 순서)이 있는 자동화된 모델은 다음 필드와 같은 결과를 출력합니다.

- **\$XS-<대상>**. 각 레코드에 대해 예측된 값입니다.
- **\$XSC-<대상>**. 예측과 연관된 신뢰도입니다.

연속형 대상. 연속형 대상이 있는 자동화된 모델은 다음 필드와 같은 결과를 출력합니다.

- **\$XR-<대상>**. 각 레코드에 대해 예측된 값입니다.
- **\$XRE-<대상>**. 예측에 대한 표준 오차입니다.

예측 규칙(대화형) 모델. 예측 규칙 모델(플래그 또는 범주형 대상 한정)은 다음 필드와 같은 결과를 출력합니다.

- **\$D-<대상>**. 각 레코드에 대해 예측된 결과로, 원하는 반응을 나타내는 데는 필드에 대해 정의된 "true" 값이 사용되며 남은 그 외의 값을 나타냅니다.
- **\$DP-<대상>**. 각 레코드에 대해 원하는 결과를 얻을 확률입니다. 모델을 작성할 때 결과 테이블에서 세그먼트에 대해 표시된 확률과 일치합니다. 자세한 정보는 63 페이지의 『예측 규칙 모델 작성』 주제를 참조하십시오.
- **\$DI-<대상>**. 레코드가 속한 세그먼트를 식별하는 데 사용되는 색인 번호입니다(모델링 결과의 가장 왼쪽 열에 표시).
- **\$DRP-<대상>**. 원하는 반응의 가능성을 나타내는 성향 스코어입니다. 플래그 대상에만 사용할 수 있습니다.

제 7 장 스코어링 모형

데이터베이스 테이블, 파일, Analytic Server 또는 IBM Cognos BI 서버에 모델 스코어링

모델을 스코어링한다는 것은 이 모델을 관심이 가는 데이터 샘플이나 모집단에 적용한다는 것을 의미합니다. 예를 들어, 기존 고객을 사용하여 어느 고객이 캠페인에 반응할 가능성이 가장 높은지 예측하는 모델을 생성할 수 있으며, 그 후 이 모델을 사용하여 현재 고객은 아니지만 홍보 메일링에 포함시킬 사람에 대한 레코드를 스코어링할 수 있습니다.

모델 스코어링의 기본 단계는 다음과 같습니다.

1. 스코어링할 데이터 소스, 그리고 해당 데이터의 모든 서브세트를 선택하십시오.
2. 스코어링된 데이터의 대상을 선택하십시오.
3. 출력할 정보와 스코어링 방법을 선택하십시오.
4. 스코어링할 레코드를 지정하십시오. 스코어 드롭 다운에서 모든 레코드, 최상위 N%, 최상위 N개 레코드, 최소/최대 성향(플래그 대상에만 사용 가능) 또는 최소 스코어(연속형 대상에만 사용 가능) 중 어느 대상을 스코어링할지 선택하십시오.
5. 예측된 결과를 미리보고 원하는 경우에는 데이터 개요를 실행하십시오.
6. 사용자의 설정을 사용해 모델에 대해 스코어링을 실행하려면 지금 스코어링을 클릭하십시오. 완료되면 스코어링 요약 보고서 섹션을 검토하십시오. 자세한 정보는 84 페이지의 『스코어링 요약 보고서』 주제를 참조하십시오.

스코어는 데이터에 추가된 하나 이상의 필드에 기록됩니다. 규칙에 따라 이러한 필드의 이름은 대상 필드를 기반으로 하지만, 여기에는 접두부가 추가됩니다. 예를 들어, 대상 필드의 이름이 반응으로 지정된 경우 결과 필드는 $\$XF$ -반응이나 $\$XFRP$ -반응과 같은 이름을 갖게 됩니다. 자세한 정보는 82 페이지의 『결과 필드 선택』 주제를 참조하십시오.

참고: 데이터베이스에서 소스 데이터가 시작될 경우 가능한 한 언제든지 실행이 데이터베이스로 푸시백됩니다 ("SQL 푸시백"이라는 프로세스에서). 일반적으로 IBM Analytical Decision Management는 IBM SPSS Modeler와 동일한 레벨의 SQL 푸시백을 지원하지만 다음 제한사항이 있습니다.

- 프로젝트에 모델 또는 외부 규칙에 대한 참조를 가지는 선택 규칙 또는 추가 필드가 포함되어 있는 경우 이러한 참조는 데이터베이스로 푸시백되지 않으며, 이는 전체적으로 프로젝트 스트림에 대해 달성할 수 있는 푸시백의 차수를 제한할 수 있습니다.
- IBM SPSS Modeler Advantage 이외의 애플리케이션에서는 선택 규칙에 대해서만 푸시백을 사용할 수 있습니다(첫 번째 차원에서의 멤버 선택 및 글로벌 선택). 프로젝트 데이터 모델에서 엔티티 ID로 정수 필드를 선택할 경우 푸시백의 차수가 늘어날 수 있습니다.

데이터 및 서브세트 선택

스코어링할 데이터 소스 목록에서 필요한 데이터 소스를 선택하십시오.

데이터 서브세트

전체 데이터 소스를 스코어링하지 않으려는 경우에는 기존 규칙을 사용하여 데이터의 서브세트를 선택하거나 이 모델에만 사용될 새 규칙을 작성할 수 있습니다. 자세한 정보는 55 페이지의 『선택 규칙 정의』 주제를 참조하십시오.

스코어링될 레코드 선택

성향 스코어를 리턴하는 모델의 경우에는 이 값을 사용하여 스코어링될 레코드를 선택할 수 있습니다. 예를 들어, 반응 성향이 가장 높은 30%의 고객을 스코어링한 후 이를 사용하여 메일링 목록을 작성할 수 있습니다. 성향 스코어는 대상의 "true" 값에 대한 가능성을 나타내며 플래그 대상을 갖는 예측 모델과 예측 규칙(대화형) 모델에 사용 가능합니다. (예측 규칙 모델의 경우 선택된 반응이 "true" 값이며 다른 모든 값은 false입니다. 다른 유형의 모델에는 이 옵션을 사용할 수 없습니다.)

성향을 기반으로 레코드를 선택하기 위해서는 세 가지 방법 중 하나를 선택할 수 있습니다.

- **최상위 백분율.** 기본적으로 이 방법은 최상위 10퍼센트를 선택합니다. 이 옵션은 수익을 기반으로 스코어링 결과를 시뮬레이션할 수 있게 해 줍니다. 자세한 정보는 73 페이지의 『최대 수익 시뮬레이션』 주제를 참조하십시오.
- **최상위 레코드 개수.** 기본적으로 이 방법은 가능한 경우 최상위 1000개의 레코드를 선택합니다.
- **최소 및 최대 성향 값을 지정.** 반응 성향이 75%에서 100% 사이인 모든 레코드를 예로 들 수 있습니다.

스코어링 대상 선택

데이터베이스, 파일, Analytic Server 또는 Cognos BI 서버에 스코어링 데이터를 출력하도록 선택할 수 있습니다. 지원되는 파일 유형에는 텍스트(*.txt, *.csv), Excel, IBM SPSS Statistics에서 사용되는 *.sav 형식, SAS 및 IBM SPSS Data Collection에서 지원하는 파일 유형이 포함됩니다. 자세한 정보는 79 페이지의 『파일 대상』 주제를 참조하십시오.

데이터베이스 대상

스코어링 결과를 데이터베이스로 전송하도록 선택하면 세부사항을 입력하도록 프롬프트됩니다.

1. 데이터베이스 이름을 입력하거나 필요한 데이터베이스를 찾아보십시오.
2. 스코어가 저장되는 방식을 지정하십시오. 새 테이블을 작성하도록 선택할 수 있으며 이 경우 테이블의 이름을 입력하도록 프롬프트됩니다. 또는, 이를 기존 테이블에 추가하거나 기존 데이터 소스를 겹쳐쓸 수 있으며 이 경우에는 관련 테이블 또는 데이터 소스를 선택하도록 프롬프트됩니다.
3. 원하는 경우에는 기본 문자열 크기를 변경하십시오.

기존 데이터베이스에 추가하거나 이를 겹쳐쓰는 경우에는 스코어링 필드를 기존 대상의 필드에 맵핑할 수 있습니다. 자세한 정보는 81 페이지의 『스코어링 필드 맵핑』 주제를 참조하십시오.

파일 대상

스코어는 몇 가지 서로 다른 파일 유형의 결과일 수 있습니다. 각 유형에 대해 사용자는 스코어링 데이터가 추가될 기존 파일을 찾아보거나 새 파일을 작성합니다.

기존 파일에 추가하거나 이를 겹쳐쓰는 경우에는 스코어링 필드를 기존 대상의 필드에 맵핑할 수 있습니다. 자세한 정보는 81 페이지의 『스코어링 필드 맵핑』 주제를 참조하십시오.

Excel 파일

유형. 내보낼 Excel 파일 유형을 선택하십시오.

필드 이름 포함. 워크시트의 첫 번째 행에 필드 이름을 포함시킬지 지정하십시오.

플랫 파일

스코어링 결과를 플랫 파일에 전송하도록 선택한 경우에는 추가 세부사항을 입력하여 파일 콘텐츠를 제어할 수 있습니다.

쓰기 모드. 겹쳐쓰기가 선택된 경우 지정된 파일에 있는 모든 기본 데이터는 겹쳐쓰여집니다. 추가가 선택된 경우 결과는 기존 파일의 끝에 추가되어 이 파일에 포함된 기존 데이터가 유지됩니다.

- 필드 이름 포함. 이 옵션이 선택된 경우에는 필드 이름이 결과 파일의 첫 번째 행에 기록됩니다. 이 옵션은 겹쳐쓰기 쓰기 모드에서만 사용 가능합니다.

각 레코드 다음에 줄 바꾸기. 이 옵션이 선택된 경우에는 결과 파일에서 각 레코드가 새 줄에 기록됩니다.

필드 구분 문자. 생성된 텍스트 파일에 있는 필드 값 사이에 삽입할 문자를 지정합니다. 옵션에는 쉼표, 탭, 공백 및 기타가 있습니다. 기타를 선택한 경우에는 원하는 구분 문자를 텍스트 상자에 입력하십시오.

기호 따옴표. 기호 필드의 값에 사용할 따옴표 유형을 지정합니다. 옵션에는 없음(값을 따옴표로 묶지 않음), 작은따옴표('), 큰따옴표(") 및 기타가 있습니다. 기타를 선택한 경우에는 값을 묶는 데 사용할 문자를 텍스트 상자에 입력하십시오.

인코딩. 사용되는 텍스트 인코딩 메소드를 지정합니다. 시스템 기본값이나 UTF-8을 선택할 수 있습니다.

소수점이하자리 기호. 데이터에서 소수점이하자리가 어떻게 표현될지 지정합니다.

- 마침표(.). 마침표 문자가 소수점 구분 기호로 사용됩니다.
- 쉼표(.). 쉼표 문자가 소수점 구분 기호로 사용됩니다.

IBM SPSS Statistics 데이터 파일

필드 이름 내보내기. IBM SPSS Statistics .sav 파일로 내보내기를 수행한 후 변수 이름 및 레이블을 처리할 방법을 지정합니다. 모든 올바르지 않은 이름은 올바르지 않은 문자를 대체하여 자동으로 정정된다는 점을 참고하십시오.

- 이름 및 변수 레이블. 필드 이름과 필드 레이블을 모두 내보내려면 선택하십시오. 이름은 IBM SPSS Statistics 변수 이름으로, 레이블은 IBM SPSS Statistics 변수 레이블로 내보내게 됩니다.
- 이름을 변수 레이블로. 필드 이름을 IBM SPSS Statistics에서 변수 레이블로 사용하려면 선택하십시오. 올바르게 사용하지 않은 IBM SPSS Statistics 이름을 작성하는 일을 방지하려면 이름 및 변수 레이블을 대신 선택하십시오.

SAS 데이터 파일

유형. 작성할 SAS 파일 형식을 선택하십시오. 세 가지 SAS 파일 형식(Windows/OS2용 SAS(*.sd2), UNIX용 SAS(*.ssd) 또는 SAS 버전 7/8/9(*sas7bdat)) 중에 선택할 수 있습니다.

필드 이름 내보내기. SAS와 함께 사용할 필드 이름 및 레이블 내보내기에 대한 옵션을 선택하십시오.

- 이름 및 변수 레이블. 필드 이름과 필드 레이블을 모두 내보내려면 선택하십시오. 이름은 SAS 변수 이름으로, 레이블은 SAS 변수 레이블로 내보내게 됩니다.
- 이름을 변수 레이블로. 필드 이름을 SAS에서 변수 레이블로 사용하려면 선택하십시오. 올바르게 사용하지 않은 SAS 이름을 작성하는 일을 방지하려면 이름 및 변수 레이블을 대신 선택하십시오.

Analytic Server 대상

스코어링 결과를 Analytic Server로 전송하도록 선택하면 세부사항을 입력하도록 프롬프트됩니다.

1. 데이터베이스 소스를 지정하십시오.
2. 스코어가 저장되는 방식을 지정하십시오. 이를 기존 테이블에 추가하거나 기존 데이터 소스를 겹쳐쓸 수 있으며 이 경우에는 관련 테이블 또는 데이터 소스를 선택하도록 프롬프트됩니다.
3. 스코어링 결과와 함께 포함할 결과 필드를 선택하십시오.
4. 스코어링할 레코드를 지정하십시오. 스코어 드롭 다운에서 모든 레코드, 최상위 N%, 최상위 N개 레코드, 최소/최대 성향(플래그 대상에만 사용 가능) 또는 최소 스코어(연속형 대상에만 사용 가능) 중 어느 대상을 스코어링할지 선택하십시오.

Cognos BI 서버 대상

스코어링 결과를 Cognos BI 서버에 전송한다고 선택하면 세부사항을 입력하도록 프롬프트됩니다.

참고: OLAP 데이터가 아닌, 관계형 데이터만 내보낼 수 있습니다.

데이터를 Cognos BI에 내보내려면 다음 항목을 지정해야 합니다.

- **Cognos 연결.** Cognos BI 서버와의 연결입니다.
- **ODBC 연결.** Cognos BI 서버가 사용하는 Cognos 데이터 서버와의 연결입니다.

이 연결들은 같은 데이터베이스를 가리켜야 하며 ODBC에 연결하기 위한 Cognos 서버의 사용자 이름 및 비밀번호는 ODBC 세부사항과 동일해야 합니다.

실제 데이터는 데이터 서버로, 패키지 메타데이터는 Cognos BI 서버로 내보내게 됩니다.

1. Cognos BI 서버 대상을 선택하고 연결을 클릭하십시오. 서버 연결 세부사항을 입력하도록 프롬프트됩니다.
 - a. 데이터를 내보내거나 가져올 IBM Cognos 서버의 서버 URL을 입력하십시오. 어느 URL을 사용할지 모르겠는 경우에는 Cognos 시스템 관리자에게 문의하십시오.
 - b. 연결에 사용할 모드를 선택하십시오. 특정 사용자(예: 관리자)로 로그인하려면 신임 정보 설정을 선택하고 IBM Cognos 네임스페이스, 사용자 이름 및 비밀번호를 입력하십시오.
 - 서버에 로그인하는 데 사용되는 IBM Cognos 보안 인증 제공자 네임스페이스를 입력하십시오. 인증 제공자는 사용자, 그룹 및 역할을 정의하고 유지하며 인증 프로세스를 제어하는 데 사용됩니다.
 - 서버에 로그인하는 데 사용할 IBM Cognos 사용자 이름을 입력하십시오.
 - 지정된 사용자 이름과 연관된 비밀번호를 입력하십시오.

또는

IBM SPSS Collaboration and Deployment Services Repository에 저장된 IBM Cognos 신임 정보를 찾아보고 선택하십시오.

- c. 또는, 사용자 신임 정보 없이 로그인하려면(이 경우 네임스페이스, 사용자 이름 및 비밀번호 필드를 지정할 수 없음) 익명 연결 사용을 선택하십시오.

참고: 일부 서버 연결에서는 익명 연결을 허용하지 않습니다.

2. 데이터 소스를 선택하십시오.
3. 내보내기 패키지가 작성될, Cognos BI 서버에 있는 폴더의 경로와 이름을 입력하십시오. 또는, 필요한 폴더를 찾아보기로 찾을 수도 있습니다.
4. 내보낸 메타데이터를 포함할 패키지의 패키지 이름을 입력하십시오. 이는 새 패키지여야 하며 기존 패키지로는 내보낼 수 없습니다. 자세한 정보는 42 페이지의 『Cognos 오브젝트 세부사항 선택』 주제를 참조하십시오.
5. ODBC 대상을 선택하려면 데이터베이스 이름을 입력하거나 찾아보기로 필요한 대상을 찾으십시오.
6. 스코어가 저장되는 방식을 지정하십시오. 새 테이블을 작성하도록 선택할 수 있으며 이 경우 테이블의 이름을 입력하도록 프롬프트됩니다. 또는, 이를 기존 테이블에 추가하거나 기존 데이터 소스를 겹쳐쓸 수 있으며 이 경우에는 관련 테이블 또는 데이터 소스를 선택하도록 프롬프트됩니다.

기존 데이터베이스에 추가하거나 이를 겹쳐쓰는 경우에는 스코어링 필드를 기존 대상의 필드에 맵핑할 수 있습니다. 자세한 정보는 『스코어링 필드 맵핑』 주제를 참조하십시오.

스코어링 필드 맵핑

기존 데이터베이스 테이블 또는 파일에 스코어를 추가하거나 겹쳐쓸 때는 스코어링 필드를 기존 테이블 또는 파일에 맵핑해야 할 수 있습니다. 결과, 그리고 맵핑된 연관 필드의 데이터 유형은 같아야 한다는 점을 참고하십시오.

- 기존 데이터베이스 테이블에 스코어를 기록할 때, 새 스코어를 겹쳐쓰는 경우나 추가하는 경우 모두 스코어링이 진행되려면 모든 필드가 기존 필드에 맵핑되어 있어야 합니다. 필요한 경우에는 필드 맵핑을 클릭하여 남아있는 모든 필드에 대해 결과를 지정하십시오.
 - 기존 파일에 기록할 때는 결과 파일 옵션에 추가가 선택되어 있는 경우에만 맵핑이 필요합니다. 기존 파일을 겹쳐쓰는 경우에는 새 파일이 이전 파일을 대체하므로 맵핑이 필요하지 않습니다.
1. 필드 맵핑을 클릭하십시오. 대화 상자가 열리며 시스템에서 자동으로 필드 맵핑을 수행합니다. 사용자는 언제든지 다시 설정을 클릭하여 시스템이 자동으로 수행한 필드 맵핑으로 돌아갈 수 있습니다.
 2. 사용 가능한 결과 대상을 선택하고 맵핑을 클릭하여 이를 필요한 스코어링 필드에 맵핑하십시오.
 3. 완료되면 저장을 클릭하십시오.

결과 필드 선택

스코어링 결과에 포함시킬 필드를 선택하려면 스코어 탭에서 결과 필드 지정을 선택하십시오. 보통은 실제 스코어 외에 고객 ID 필드와 같이 각 레코드를 식별할 수 있게 해 주는 필드가 하나 이상 필요합니다. 선택적으로, 모델링에 사용된 입력 필드를 일부 또는 전부 포함시킬 수 있습니다.

모델 결과 필드

규칙에 따라 스코어링 결과 필드의 이름은 대상 필드의 이름을 기반으로 하지만, 여기에는 사용자 정의 접두부가 추가됩니다. 예를 들어, 대상 필드의 이름이 반응으로 지정된 경우 결과 필드는 $\$XF$ -반응이나 $\$XFRP$ -반응과 같은 이름을 갖게 됩니다. 특정 필드는 아래에 자세히 나타나 있는 바와 같이 모델의 유형과 선택된 대상 필드의 측정 수준에 따라 달라집니다. 자세한 정보는 45 페이지의 『측정 수준』 주제를 참조하십시오.

플래그 대상. 플래그 대상이 있는 자동화된 모델은 다음 필드와 같은 결과를 출력합니다.

- $\$XF$ -<대상>. 각 레코드에 대해 예측된 값으로, 필드에 대해 정의된 "true" 및 "false" 값의 용어로 보고됩니다.
- $\$XFRP$ -<대상>. 성향 스코어로, 각 레코드에 대해 "true" 값의 가능성을 나타냅니다. 성향 스코어는 보통 해당 레코드에 대한 예측으로 보고되는 신뢰도 스코어와는 다릅니다. 예를 들어, 신뢰도가 높은 "false" 예측은 반응하지 않을 가능성이 높다는 것을 나타내므로 낮은 성향으로 해석됩니다.

범주형 대상. 범주형 대상(명목 또는 순서)이 있는 자동화된 모델은 다음 필드와 같은 결과를 출력합니다.

- $\$XS$ -<대상>. 각 레코드에 대해 예측된 값입니다.
- $\$XSC$ -<대상>. 예측과 연관된 신뢰도입니다.

연속형 대상. 연속형 대상이 있는 자동화된 모델은 다음 필드와 같은 결과를 출력합니다.

- $\$XR$ -<대상>. 각 레코드에 대해 예측된 값입니다.
- $\$XRE$ -<대상>. 예측에 대한 표준 오차입니다.

예측 규칙(대화형) 모델. 예측 규칙 모델(플래그 또는 범주형 대상 한정)은 다음 필드와 같은 결과를 출력합니다.

- **\$D-<대상>**. 각 레코드에 대해 예측된 결과로, 원하는 반응을 나타내는 데는 필드에 대해 정의된 "true" 값이 사용되며 넓은 그 외의 값을 나타냅니다.
- **\$DP-<대상>**. 각 레코드에 대해 원하는 결과를 얻을 확률입니다. 모델을 작성할 때 결과 테이블에서 세그먼트에 대해 표시된 확률과 일치합니다. 자세한 정보는 63 페이지의 『예측 규칙 모델 작성』 주제를 참조하십시오.
- **\$DI-<대상>**. 레코드가 속한 세그먼트를 식별하는 데 사용되는 색인 번호입니다(모델링 결과의 가장 왼쪽 열에 표시).
- **\$DRP-<대상>**. 원하는 반응의 가능성을 나타내는 성향 스코어입니다. 플래그 대상에만 사용할 수 있습니다.

모델 스코어링

스코어링 세부사항을 지정한 후에는 모델 스코어의 샘플을 미리볼 수 있습니다. 미리보기는 데이터 세트가 큰 경우 특히 유용한데, 이는 전체 데이터 세트가 스코어링되기를 기다릴 필요 없이 기대했던 대로 스코어링 결과가 나타나는지 빠르게 확인할 수 있기 때문입니다.

참고: 데이터가 올바르지 않거나 불완전한 경우에는 미리보기를 시도할 때 경고가 표시됩니다.

미리보기는 행에 레코드, 열에 변수가 있는 테이블로 데이터를 표시합니다.

스코어링

스코어링되는 레코드의 수를 설정하려면 새로 고치기 아이콘을 클릭하십시오. 이 아이콘을 클릭하면 데이터 소스에 있는 레코드의 총계와 사용자가 지정한 선택적 설정을 고려하여 모델을 작성하는 데 사용될 레코드의 수를 알려줍니다.

지금 스코어링을 클릭하면 진행 표시줄이 탭의 맨 아래에 표시됩니다. 이 표시줄이 표시되어 있는 동안은 언제든지 스코어링을 중지할 수 있습니다.

참고:

데이터가 올바르지 않거나 불완전한 경우에는 스코어링을 시도할 때 경고가 표시됩니다.

데이터베이스에 대해 스코어링할 경우 가능한 한 언제든지 실행이 데이터베이스로 푸시백됩니다("SQL 푸시백이라 함"). 프로젝트에 모델에 대한 참조가 포함된 규칙이 포함된 경우 실행이 데이터베이스로 푸시백되지 않습니다.

데이터 스코어링이 완료되면 결과가 저장된 위치(지정한 데이터베이스 또는 파일)를 알리는 메시지가 표시되며 스코어링 요약 보고서에서 요약 내용이 제공됩니다. 자세한 정보는 84 페이지의 『스코어링 요약 보고서』 주제를 참조하십시오. 스코어링 결과를 플랫폼 파일에 전송하도록 선택했으며 관리자가 이 기능을 사용으로 설정한 경우에는 이 파일의 로컬 사본을 다운로드할 수도 있습니다.

스코어링 요약 보고서

프로젝트를 스코어링하면 스코어링 요약 보고서 섹션에서 데이터가 스코어링된 위치, 파일 크기(적용 가능한 경우), 데이터베이스 및 테이블 이름(적용 가능한 경우), 모델 이름 또는 프로젝트 이름, 스코어링된 레코드의 수, 그리고 어떤 경우에는 모델을 요약하는 차트 및/또는 테이블 등의 세부사항을 제공합니다. 스코어링 요약 보고서 섹션은 IBM SPSS Modeler Advantage와 지금 스코어링 기능을 포함하는 애플리케이션의 배포 탭에서 사용 가능합니다.

스코어링 결과를 플랫폼 파일에 전송하도록 선택한 경우에는 필요한 경우 이 파일의 로컬 사본을 다운로드할 수도 있습니다. 관리자는 브라우저 기반 IBM SPSS Collaboration and Deployment Services Deployment Manager를 사용하여 이 기능을 사용으로, 또는 사용 안함으로 설정하거나 허용된 최대 파일 크기 한계를 늘리고 줄일 수 있습니다. 파일 다운로드 단추가 비활성 상태인 경우에는 파일 크기가 관리자가 설정한 최대 파일 한계를 초과하거나 관리자가 이 기능을 사용 안함으로 설정한 것입니다. 이 기능은 기본적으로 사용 안함으로 설정되어 있습니다.

제 8 장 애플리케이션 관리

관리자는 비즈니스 사용자가 IBM Analytical Decision Management 애플리케이션과 상호작용하는 방식을 제어할 수 있습니다. 이러한 수준의 제어는 비즈니스 사용자의 작업을 용이하게 해주어 비즈니스 문제점을 해결하는 데 도움을 받기 위해 애플리케이션을 더 빠르고 효율적으로 사용할 수 있게 해 줍니다. 또한 관리자가 특정 기능에 대한 액세스를 제한할 수 있게 해 줍니다.

필요에 따라 특정 사용자 또는 그룹에 관리 권한을 부여하려면 IBM SPSS Collaboration and Deployment Services Deployment Manager를 사용하십시오. *Decision Management* 관리 조치를 부여받은 사용자만 이 섹션에 설명된 기능에 대한 액세스 권한을 갖습니다.

1. IBM SPSS Collaboration and Deployment Services Deployment Manager를 여십시오.
2. 도구 메뉴에서 서버 관리를 선택하십시오.
3. 서버 관리 탭에서 사용자의 서버에 연결하십시오.
4. 사용자 및 그룹에서 로컬 사용자 리포지토리를 선택하십시오.
5. 새 그룹을 클릭하여 관리자, 그리고 비즈니스 사용자의 새 그룹을 작성하십시오. 예를 들어, 다음과 같습니다.

- Decision Management 관리자
- Decision Management 사용자

이 때 원하는 경우에는 새 사용자를 작성할 수도 있습니다. 그룹을 작성할 때는 각 그룹에 적절한 사용자를 추가하십시오. 예를 들어, 관리자 그룹의 구성원은 조직에서 IBM Analytical Decision Management 애플리케이션을 관리할 권한을 부여받습니다.

6. 역할으로 이동한 후 새 역할을 클릭하여 새 관리 역할과 새 비즈니스 사용자 역할을 작성하십시오. 예를 들어, 다음과 같습니다.
 - IBM Analytical Decision Management 관리자
 - IBM Analytical Decision Management 사용자

새 역할을 작성할 때는 최소한 이를 다음 조치에 지정하십시오.

표 7. 역할 조치.

역할	필요한 조치
Decision Management 관리자	컨텐츠 및 폴더에 액세스
	Decision Management 관리
	작업 편집
	동적으로 보고서 실행
	모델 스코어링
	작업 제출
	모든 버전 표시

표 7. 역할 조치 (계속).

역할	필요한 조치
Decision Management 사용자	컨텐츠 및 폴더에 액세스
	작업 편집
	동적으로 보고서 실행
	모델 스코어링
	작업 제출
	최신 또는 모든 버전 표시*

*참고: 사용자가 자신이 작성하지 않은 프로젝트에 액세스해야 하는 경우에는 최신 버전 표시 또는 모든 버전 표시 조치 중 하나가 필요합니다. 조치가 없을 경우 프로젝트는 읽기 전용 모드로 열리며 현재 사용자에게 최신 버전에 대한 액세스 권한이 주어지지 않습니다.

- Decision Management 관리자 역할을 선택하고 사용자 및 그룹 편집을 클릭하십시오. 방금 작성한 Decision Management 관리자 그룹을 추가하십시오.
- Decision Management 사용자 역할을 선택하고 사용자 및 그룹 편집을 클릭하십시오. 방금 작성한 Decision Management 사용자 그룹을 추가하십시오.

일반 옵션

특정 일반 옵션은 관리자가 IBM Analytical Decision Management 사용자 인터페이스의 모든 탭에서 사용할 수 있습니다. 이러한 옵션은 관리자가 비즈니스 사용자로부터 탭을 숨기거나 잠글 수 있게 해 줍니다.

- 각 탭의 맨 위에는 관리자가 아닌 사람으로부터 <탭 이름> 숨기기 옵션이 있으며 여기서 <탭 이름>은 해당 탭의 이름입니다. 이는 비즈니스 사용자로부터 해당 탭 전체를 숨깁니다. 예를 들어, 사용자가 애플리케이션에서 데이터 탭을 볼 수 없게 하려면 애플리케이션에 로그인하여 데이터 탭으로 이동한 후 관리자가 아닌 사람으로부터 데이터 탭 숨기기를 선택하십시오.
- 각 탭의 맨 위에는 모든 <탭 이름> 잠금 옵션이 있으며 여기서 <탭 이름>은 해당 탭의 이름입니다. 이는 탭에 있는 모든 제어를 잠궈 비즈니스 사용자가 읽기만 할 수 있도록 합니다. 예를 들어, 비즈니스 사용자가 현재 데이터 설정을 볼 수 있으나 이를 변경할 수 없도록 하려면 애플리케이션의 데이터 탭에서 이 옵션을 설정할 수 있습니다. 이 설정은 특정 부분을 미리 설정하여 비즈니스 사용자의 작업을 간단하게 하려는 경우에도 매우 유용합니다.
- 관리자는 사용 가능한 경우 패드락 아이콘을 클릭하여 각 탭의 개별 설정을 잠그거나 잠금 해제할 수 있습니다.



예를 들어, 정의 탭이 있는 애플리케이션에서 관리자는 비즈니스 사용자가 변경할 수 없도록 차원 계층 구조 트리를 잠글 수 있습니다. 패드락 아이콘을 클릭하면 연관된 제어가 잠기거나 잠금 해제됩니다. 비즈니스 사용자는 관리자가 잠근 제어의 잠긴 패드락 아이콘은 볼 수 있지만 잠금 해제된 패드락 아이콘은 볼 수 없습니다.

잠긴 개별 항목은 선택할 수 없음을 나타내기 위해 회색으로 칠해진 배경 또는 어두워진 색상으로 표시됩니다. 사용자가 이를 선택하려 시도하면 오류 메시지가 표시됩니다. 전체 탭이 잠긴 경우 이러한 메시지는 탭의 맨 위에 표시됩니다.

애플리케이션 단축키 제어

애플리케이션 시작 페이지는 비즈니스 사용자에게 애플리케이션을 열고 애플리케이션의 특정 레이블된 버전에 대한 사용자 정의 단축키를 추가하는 손쉬운 방법을 제공합니다. 관리자는 모든 사용자의 시작 페이지에 어느 애플리케이션 단축키가 표시될지 지정할 수 있습니다. 관리자 권한으로 로그인한 후 다음 작업을 수행하십시오.

- 원하는 경우에는 비즈니스 사용자가 애플리케이션 단축키를 시작 페이지에서 제거할 수 없도록 해당 애플리케이션 단축키의 패드락 아이콘을 클릭하십시오. 비즈니스 사용자는 여전히 다른 애플리케이션 단축키를 추가하고 제거할 수 있다는 점을 참고하십시오. 사용자는 잠긴 단축키의 경우 닫기 아이콘 대신 잠긴 패드락 아이콘을 보게 됩니다. 자세한 정보는 25 페이지의 『애플리케이션 시작』 주제를 참조하십시오.

홈 페이지 숨기기

각 애플리케이션은 홈 페이지에 나타나 있는 그래픽과 같이 단계별 워크플로우를 제공합니다. 그래픽을 클릭하면 사용자 인터페이스의 해당 섹션으로 점프합니다. 그러나 비즈니스 사용자가 애플리케이션의 워크플로우에 있는 모든 단계에 액세스할 필요가 있는 것은 아닙니다.

관리자는 비즈니스 사용자로부터 홈 페이지를 숨길 수 있습니다. 예를 들어, 특정 애플리케이션에서는 비즈니스 사용자가 보고서 탭만을 사용할 수도 있습니다. 이러한 경우 관리자는 홈 페이지와 다른 모든 탭을 비즈니스 사용자로부터 숨길 수 있습니다. 이렇게 하면 비즈니스 사용자가 로그인했을 때 이 사용자는 바로 보고서 탭으로 이동되며 다른 탭을 보는 일이 없게 됩니다.

1. 홈 페이지에서 애플리케이션 홈 페이지 숨기기를 선택하십시오.
2. 사용자가 애플리케이션을 처음 열었을 때 표시할 탭을 드롭 다운에서 지정하십시오.

이 선택된 탭을 이후에 숨기는 경우에는 애플리케이션에서 이 다음으로 숨겨지지 않은 탭이 표시됩니다.

데이터 옵션 잠금

데이터

데이터 탭은 애플리케이션이 분석, 시뮬레이션 및 테스트, 스코어링 또는 기타 운영용으로 사용하는 데이터 세트를 정의합니다. 이러한 데이터 소스는 사용자의 서버 설치에 따라 정의됩니다. 따라서 데이터 파일, ODBC 소스 등에 대한 모든 경로는 서버에 따라 달라집니다.

관리자는 비즈니스 사용자가 애플리케이션을 사용하기 전에 데이터의 일부 또는 전부를 설정해 둘 수 있습니다. 관리자는 다음과 같은 작업을 수행할 수 있습니다.

- 새 데이터 소스를 작성할 때 데이터 소스 잠금을 선택하여 이 소스를 작성한 후 잠글 수 있습니다.

- 테이블 또는 데이터 소스 편집기 대화 상자에서 데이터 소스 옆에 있는 패드락 아이콘을 클릭하여 개별 데이터 소스를 잠그거나 잠금을 해제할 수 있습니다.
- 프로젝트 데이터 모델 드롭 다운에 있는 데이터 모델을 선택하고 패드락 아이콘을 클릭하여 애플리케이션에서 사용되는 데이터 모델을 설정하고 잠글 수 있습니다.

비즈니스 사용자는 잠긴 데이터 소스를 프로젝트 데이터 세트에서 내 데이터 세트 섹션으로 복사할 수 있으나 그 후에도 이 소스는 두 세트 모두에서 잠긴 상태로 남아있게 된다는 점을 참고하십시오. 자세한 정보는 33 페이지의 제 4 장 『데이터 소스 관리』 주제를 참조하십시오.

- 프로젝트 데이터 모델이 데이터 보기인 경우, 계층 구조 데이터에 기반하여 ODM 규칙 사용 허용 드롭 다운을 사용하여 계층 구조 데이터 모델이 있는 IBM ODM(Operational Decision Management) 규칙의 사용을 제어하십시오. 해당 규칙을 사용하면 성능이 저하될 수 있습니다. 다음 옵션이 드롭 다운에서 사용 가능합니다.
 - 최상의 성능을 위해 허용 금지가 권장됩니다. 복합 데이터 구조가 포함된 ODM 규칙이 프로젝트의 임의의 위치에서 사용되지 않도록 합니다. 애플리케이션의 임의의 지점에서 복합 규칙을 선택하면 오류가 표시됩니다.
 - 모델링 탭에서만은 복합 규칙이 모델링 탭에서 모델을 빌드할 때 참조되도록 허용합니다. 이는 유연성을 증가시키고 성능에 영향을 미칩니다. 이 옵션은 IBM SPSS Modeler Advantage에만 적용됩니다.

스코어링 대상 옵션 잠금

지금 스코어링 섹션(배포 또는 스코어 탭)을 사용하도록 구성된 애플리케이션의 경우에는 비즈니스 사용자가 스코어 데이터를 저장하기 위한 대상(데이터베이스 또는 파일)을 지정할 수 있게 해 주는 섹션이 있습니다. 예를 들어, 데이터베이스에 스코어를 저장할 경우에는 데이터베이스 이름과 스코어를 저장하는 방법(데이터베이스에 새 테이블 작성 등)을 지정합니다.

관리자는 스코어링에 대한 대상 옵션을 잠글 수 있습니다. 예를 들어, 비즈니스 사용자에게 기존 데이터를 겹쳐쓰거나, 기존 테이블에 새 데이터를 추가하거나, 새 테이블을 작성하거나, 파일에 데이터를 저장하는 것과 같은 옵션을 제공하지 않고, 스코어를 특정 데이터베이스에 있는 기존 테이블에 추가만 하도록 할 수 있습니다.

- 모든 스코어링 대상 옵션을 잠그려면 배포 탭의 대상 드롭 다운 옆에 있는 패드락 아이콘을 클릭하십시오. 비즈니스 사용자가 대상 옵션을 볼 수 있으나 편집할 수는 없게 됩니다.

외부 규칙 작성에 사용하기 위한 프로젝트 메타데이터 다운로드

IBM Operational Decision Management와 같은 비즈니스 규칙 관리 시스템에서 작성된 규칙은 현재 IBM Analytical Decision Management 프로젝트에서 사용된 것과 같은 데이터 모델을 지원하도록 개발된 경우 IBM Analytical Decision Management 애플리케이션에서 참조하고 사용할 수 있으며, 애플리케이션에서 사용할 수 있도록 웹 서비스로 배포될 수 있습니다. 이를 위해 IBM Analytical Decision Management 관리자는 관련 메타데이터를 다운로드해야 합니다. 메타데이터를 다운로드하는 데는 두 가지 방법이 있으며, 데이터 모델의 복잡도와 잠재적인 결과 규칙의 복잡도는 다음에 따라 다릅니다.

IBM Analytical Decision Management에서 프로젝트 메타데이터 다운로드

관리자는 IBM Analytical Decision Management의 메타데이터 다운로드 아이콘을 클릭하여 현재 프로젝트에 대한 메타데이터가 포함된 .ZIP 파일을 다운로드할 수 있습니다. 다운로드된 파일은 관리자가 다운로드를 트리거하는 데 사용한 시스템에 저장됩니다.



그림 8. 메타데이터 다운로드 아이콘

이 방법으로 다운로드한 메타데이터는 프로젝트 데이터 모델을 기반으로 규칙 작성을 지원합니다. 프로젝트 데이터 모델은 플랫폼이 되므로(계층 구조가 아님) 이 메타데이터를 기반으로 작성된 규칙은 플랫폼 데이터 모델만 이용할 수 있습니다. 이 플랫폼 데이터 모델을 기반으로 작성된 규칙은 프로젝트 데이터 모델 데이터 소스의 유형과 관계 없이 IBM Analytical Decision Management 프로젝트에서 참조될 수 있습니다.

.ZIP 파일에는 다음 파일이 포함되어 있습니다.

- **XML 스키마 정의(*.XSD)**. 현재 IBM Analytical Decision Management 프로젝트의 데이터 모델에 있는 필드 유형의 정의를 포함합니다. IBM Operational Decision Management Rule Designer와 같은 외부 개발 도구에서는 현재 프로젝트에 사용할 규칙을 개발하기 위해 이 파일을 가져올 수 있습니다.
- **웹 서비스 설명 언어 템플릿(*.WSDL)**. 데이터 모델을 사용하는 웹 서비스 개발을 돕기 위해 포함되었습니다. .WSDL 파일은 IBM Analytical Decision Management 애플리케이션과 외부 규칙 서비스 사이의 메시지를 중개하는 웹 서비스를 빌드하는 데 사용될 수 있는 템플릿으로 제공되었습니다. 중개 웹 서비스에 대한 세부사항은 작성자가 직접 제공할 수 있습니다.

IBM SPSS Collaboration and Deployment Services 분석 데이터 보기용 데이터 모델 메타데이터 다운로드

IBM SPSS Collaboration and Deployment Services Deployment Manager에서는 특수 분석 데이터 보기용 메타데이터를 내보낼 수 있습니다.

이 방법으로 다운로드한 메타데이터는 분석 데이터 보기의 논리 데이터 모델을 기반으로 규칙 작성을 지원합니다. 분석 데이터 보기의 논리 데이터 모델은 계층 구조일 수 있습니다. IBM Analytical Decision Management 데이터 보기 소스는 각 레코드에 대한 계층 구조 데이터의 XML 표현 생성을 지원하므로, 생성 시에 외부 규칙으로 전달될 수 있습니다. 이 메타데이터를 기반으로 작성된 규칙은 복잡할 수 있으며 계층 구조 데이터 모델을 이용할 수 있습니다.

IBM Analytical Decision Management에서는 프로젝트 데이터 모델이 데이터 보기 소스를 사용하는 프로젝트에서 계층 구조 데이터 모델을 기반으로 한 외부 규칙만 참조될 수 있습니다. 호환성을 위해서는 데이터 보기 소스 및 외부 규칙이 동일한 분석 데이터 보기의 논리 데이터 모델을 기반으로 해야 합니다. 데이터 보기 소스만이 참조하는 분석 데이터 보기를 기반으로 계층 구조 데이터를 생성할 수 있습니다. 달리 말해서, 외부 규칙을 사용하려면 IBM Analytical Decision Management 프로젝트의 프로젝트 데이터 모델이 규칙 작성에 사용되는 모델과 일치해야 하며 분석 데이터 보기를 통해 이를 일치시킬 수 있습니다.

분석 데이터 보기의 논리 데이터 모델에 대한 메타데이터와 함께 .ZIP 파일을 다운로드하려면 다음을 수행하십시오.

1. IBM SPSS Collaboration and Deployment Services 배포 관리자에서 리포지토리의 분석 데이터 보기를 열고 논리 탭으로 이동하십시오.
2. 표를 마우스 오른쪽 단추로 클릭하고 XOM 메타데이터를 내보내십시오.
3. 파일을 저장할 위치를 선택하고 저장을 클릭하십시오.

.ZIP 파일에는 다음 파일이 포함되어 있습니다.

XML 스키마 정의(*.XSD). 둘 이상의 XML 스키마 정의가 포함되어 있습니다. 하나에는 규칙 반응 유형에 유용할 수 있는 정의가 포함되어 있고, 다른 스키마에는 분석 데이터 보기의 논리 데이터 모델에 있는 계층 구조 관계 세부사항 및 필드 유형을 비롯하여 표의 정의가 포함되어 있습니다.

프로젝트 데이터 모델이 동일한 분석 데이터 보기 인스턴스를 기반으로 데이터 보기 소스를 사용하는 프로젝트와 함께 사용할 규칙을 개발하기 위해 IBM Operational Decision Management Rule Designer와 같은 외부 개발 도구로 이러한 스키마 파일을 가져올 수 있습니다.

IBM Analytical Decision Management에서 사용할 외부 규칙을 작성하는 데 대한 정보는 *애플리케이션 디자이너 안내서*를 참조하십시오.

이러한 복합 외부 규칙을 사용할 경우 성능을 저하시킬 수 있기 때문에 관리자는 해당 규칙의 사용을 제한할 수 있습니다. 세부사항은 87 페이지의 『데이터 옵션 잠금』의 내용을 참조하십시오.

XML 스키마 정의 편집

IBM Analytical Decision Management에서 다운로드한 .XSD 파일은 두 가지 복합 데이터 유형을 포함하며, 하나는 IBM Operational Decision Management 규칙 프로젝트의 입력 매개변수로 사용되어야 하는 IBM Analytical Decision Management 프로젝트 데이터 모델을 나타내고, 다른 하나는 IBM Operational Decision Management 규칙 프로젝트의 결과 매개변수로 사용되어야 하는 규칙 리턴 유형을 나타냅니다.

IBM SPSS Collaboration and Deployment Services Deployment Manager 분석 데이터 보기 메타데이터에는 둘 이상의 .XSD 파일이 포함되어 있습니다. 하나는 IBM Operational Decision Management 규칙 프로젝트의 입력 매개변수로 사용되어야 하는 논리 데이터 모델과 관련된 파일이고, 다른 하나에는 IBM Operational Decision Management 규칙 프로젝트의 결과 매개변수로 사용되어야 하는 규칙 리턴 유형을 나타내는 복합 유형이 포함되어 있습니다.

숙련된 사용자는 생성된 .XSD 스키마 정의를 작성하려는 특정 규칙에 맞추어 편집하기로 선택할 수 있습니다. 사용자가 규칙을 작성하기 전에 관련 .XSD 스키마 정의에 정의된 입력 및/또는 반응 관련 유형을 편집하려는 경우가 있습니다. 예를 들어, 전문 사용자가 작성한 규칙을 허용하여 널값인 입력 필드의 가능성을 처리하기 위해 입력 매개변수 관련 스키마를 편집해야 할 수 있습니다. 전문 사용자는 .XSD 스키마 정의를 통해 필수 및 선택 입력 매개변수 필드/요소를 표시함으로써 이를 수행할 수 있습니다.

입력 매개변수 유형 편집

각 필드 관련 요소가 발생 가능한 수는 .XSD 스키마에 명시적으로 설정되어 있지 않습니다(즉 속성 minOccurs 및 maxOccurs는 설정되어 있지 않으며 각 속성의 기본값은 1). 이와 같이, 스키마를 IBM Operational Decision Management 규칙을 생성하기 위해 사용하기 전에 편집하지 않은 경우 생성된 규칙에서는 입력 매개변수 유형의 모든 필드를 필요로 합니다.

프로젝트 데이터 모델 관련 입력 매개변수 유형 편집

입력 매개변수 유형을 정의하는 스키마가 IBM Analytical Decision Management 프로젝트의 메타데이터를 다운로드하여 생성된 경우 입력 매개변수의 데이터 모델은 플랫폼이 됩니다. 이 경우에는 IBM Analytical Decision Management가 실행 시 규칙 서비스에 필요한 페이로드를 직접 작성합니다. 이를 통해 전문 사용자는 호환성을 유지하면서도 좀 더 자유롭게 입력 매개변수 유형을 편집할 수 있습니다. 예를 들면, 규칙 작성에 필요하지 않은 필드/요소를 제거하거나 필수/요소가 선택사항인지 표시할 수 있습니다.

분석 데이터 보기 논리적 데이터 모델 관련 입력 매개변수 유형 편집

입력 매개변수 유형을 정의하는 스키마가 IBM SPSS Collaboration and Deployment Services Deployment Manager 클라이언트로부터 분석 데이터 보기의 논리 데이터 모델을 내보내어 생성된 경우, 이 입력 매개변수의 데이터 모델은 계층 구조가 될 수 있습니다. 계층 구조인 경우 IBM Analytical Decision Management는 실행 시 규칙 서비스에 필요한 페이로드를 직접 작성하지 않습니다. 페이로드의 주요 파트는 분석 데이터 보기 소스에 의해 생성됩니다. 이렇게 되면 전문 사용자는 입력 매개변수 유형을 편집하고 호환성을 유지하는 방법에 제한을 받게 됩니다. 예를 들면, 필드/요소가 선택사항이지만 표시할 수 있습니다.

일반 정보

IBM Operational Decision Management와 같은 비즈니스 규칙 관리 시스템에서 작성한 규칙을 호출하기 전에, IBM Analytical Decision Management는 규칙의 서비스가 입력 매개변수 데이터 모델이 계층 구조인지 플랫폼인지를 묻습니다. 입력 매개변수가 플랫폼이면 규칙의 서비스에 필요한 입력도 확인합니다.

입력 매개변수의 데이터 모델이 플랫폼이고 스코어링될 레코드가 규칙 서비스에서 필요로 하는 필드에 대한 널값을 포함하고 있는 경우, IBM Analytical Decision Management에서는 규칙을 호출하지 않으며 규칙의 결과/반응 필드가 기본값으로 채워집니다. 결과/반응 필드의 유형에 따른 값은 다음과 같습니다.

표 8. 반응 필드 유형의 값

반응 필드 유형	값
xsd:string	""
xsd:boolean	""
xsd:int	0
xsd:long	0
xsd:byte	0
xsd:double	0.0
xsd:float	0.0
xsd:date	0.0

표 8. 반응 필드 유형의 값 (계속).

반응 필드 유형	값
xsd:time	0.0
xsd:datetime	0.0

이는 규칙 서비스를 통한 레코드 스코어링의 결과에 영향을 줄 수 있으며 .XSD 스키마의 입력 매개변수를 편집할 때는 주의해야 합니다.

입력 매개변수의 데이터 모델이 계층 구조이고 스코어링될 레코드가 규칙 서비스에서 필요로 하는 필드에 대한 널값을 포함하고 있는 경우, IBM Analytical Decision Management에서 규칙을 호출하지 않지만 널값이 있는 필드는 데이터 보기 소스 노드에서 작성된 페이로드에서 제외됩니다. 제외된 필드가 필수 필드인 경우 규칙 호출에 실패하게 되므로 .XSD 스키마에서 입력 매개변수를 편집할 때는 주의를 기울여야 합니다.

스키마를 편집하려면 작성할 규칙 서비스에 대한 지식이 있어야 합니다. 규칙에서 특정 입력 필드를 필요로 하지 않고 규칙의 입력 매개변수 데이터 모델이 플랫폼인 경우, 필드 관련 요소를 스키마의 입력 매개변수에서 제거하거나 선택사항으로 표시할 수 있습니다. 규칙에서 특정 필드를 필요로 하지 않고 규칙의 입력 매개변수가 계층 구조인 경우, 필드 관련 요소를 제거할 수 없지만 선택사항으로 표시할 수 있습니다. 이는 IBM Analytical Decision Management가 계층 구조 데이터 모델을 기반으로 규칙에 대한 페이로드를 직접 생성하지 않기 때문입니다.

입력 매개변수 필드를 제거할 경우 규칙을 호출하면 필드 값이 무시되며 규칙은 필드를 참조할 수 없습니다. 입력 매개변수 필드를 선택사항으로 만들 경우에는 널값이 허용되며 규칙은 이를 확인할 수 있도록 방어적으로 작성되어야 합니다. 예를 들어, 규칙의 입력 매개변수 데이터 모델이 플랫폼이고 규칙에서 Age 필드를 사용하지 않는다는 것을 사용자가 알고 있는 경우 사용자는 이 필드가 필수가 아니라는 것을 나타내기 위해 관련 요소를 제거할 수 있습니다. 다음 .XSD 스키마를 예로 들면, 필드 Age가 필수가 아닌 경우에는 전체 섹션 `<xsd:element name="Age">`를 스키마에서 제거할 수 있습니다.

```

<xsd:complexType name="drugs">
<xsd:annotation>
  <xsd:appinfo>
    <dmname>drugs</dmname>
  </xsd:appinfo>
</xsd:annotation>
<xsd:sequence>
  <xsd:element name="Age">
<xsd:annotation>
  <xsd:appinfo>
    <dmname>Age</dmname>
  </xsd:appinfo>
</xsd:annotation>
  <xsd:simpleType>
    <xsd:restriction base="xsd:long">
      <xsd:minInclusive value="15"/>
      <xsd:maxInclusive value="74"/>
    </xsd:restriction>
  </xsd:simpleType>
</xsd:element>
  <xsd:element name="Sex">
<xsd:annotation>
  <xsd:appinfo>
    <dmname>Sex</dmname>
  </xsd:appinfo>
</xsd:annotation>

```

```

    <xsd:simpleType>
      <xsd:restriction base="xsd:string">
        <xsd:enumeration value="F"/>
        <xsd:enumeration value="M"/>
      </xsd:restriction>
    </xsd:simpleType>
  </xsd:element>
  ...
</xsd:sequence>
</xsd:complexType>

```

또는 규칙의 입력 매개변수 데이터 모델이 플랫폼 또는 계층 구조인 경우 규칙에서 필요하지 않으면 요소의 minOccurs 속성 값을 0으로 설정하여 입력 필드 요소를 선택사항으로 지정할 수 있습니다.

```

<xsd:complexType name="drugs">
  <xsd:annotation>
    <xsd:appinfo>
      <dmname>drugs</dmname>
    </xsd:appinfo>
  </xsd:annotation>
  <xsd:sequence>
    <xsd:element name="Age" minOccurs="0">
      <xsd:annotation>
        <xsd:appinfo>
          <dmname>Age</dmname>
        </xsd:appinfo>
      </xsd:annotation>
      <xsd:simpleType>
        <xsd:restriction base="xsd:long">
          <xsd:minInclusive value="15"/>
          <xsd:maxInclusive value="74"/>
        </xsd:restriction>
      </xsd:simpleType>
    </xsd:element>
    <xsd:element name="Sex">
      <xsd:annotation>
        <xsd:appinfo>
          <dmname>Sex</dmname>
        </xsd:appinfo>
      </xsd:annotation>
      <xsd:simpleType>
        <xsd:restriction base="xsd:string">
          <xsd:enumeration value="F"/>
          <xsd:enumeration value="M"/>
        </xsd:restriction>
      </xsd:simpleType>
    </xsd:element>
    ...
  </xsd:sequence>
</xsd:complexType>

```

계층 구조 데이터 모델의 예:

```

<xsd:complexType name="Customer">
  <xsd:sequence>
    <xsd:element name="FirstName" type="xsd:string" minOccurs="0"/>
    <xsd:element name="Surname" type="xsd:string" minOccurs="0"/>
    <xsd:element name="Age" type="xsd:long"/>
    <xsd:element name="Sex" type="xsd:string"/>
    <xsd:element name="CustID" type="xsd:string"/>
    <xsd:element maxOccurs="unbounded" minOccurs="0" name="Orders" type="xom:Order"/>
  </xsd:sequence>
</xsd:complexType>
<xsd:complexType name="Order">
  <xsd:sequence>
    <xsd:element name="CustID" type="xsd:string"/>
    <xsd:element name="NoOfItems" type="xsd:long"/>
  </xsd:sequence>
</xsd:complexType>

```

```

    <xsd:element name="TotalCost" type="xsd:long" minOccurs="0"/>
    <xsd:element name="Date" type="xsd:date"/>
  </xsd:sequence>
</xsd:complexType>

```

규칙은 널값을 처리하기 위해 방어적으로 작성될 수 있으며, 이 경우 필드는 스키마에서 optional로 표시됩니다. 다음 내용은 널값을 처리하기 위해 작성된, 이름이 *currentOfferIsReplacePhone*인 IBM Operational Decision Management 규칙입니다.

```

if
    the CURRENT OFFER of 'the customer' is not null
    and the CURRENT OFFER of 'the customer' is "Replace Phone"
then
    make it true that 'the response' is value ;
else
    make it false that 'the response' is value ;

```

반응 매개변수 유형 편집

다양한 유형(예: 선택에 사용할 부울, 통합 규칙에 사용할 숫자, 문자열 등)을 IBM Analytical Decision Management에 리턴하는 IBM Operational Decision Management 규칙을 작성하고 이용할 수 있습니다. 그러나 문자열 데이터 반응 규칙 외에 다른 것을 작성하려면 반응 유형이 포함된 XML 스키마 정의(.XSD)를 수동으로 편집해야 합니다.

또한 규칙 프로젝트는 각각 서로 다른 유형을 리턴하는 여러 규칙을 포함할 수 있습니다. 이를 지원하려면 .XSD의 여러 버전을 유지해야 합니다(예: 데이터 반응 변수가 숫자로 설정된 버전, 문자열로 설정된 다른 버전, 부울로 설정된 또 다른 버전 등).

IBM Analytical Decision Management에서 다운로드한 .XSD 파일은 두 가지 복합 데이터 유형을 포함하며, 하나는 IBM Operational Decision Management 규칙 프로젝트의 입력 매개변수로 사용되어야 하는 IBM Analytical Decision Management 프로젝트 데이터 모델을 나타내고, 다른 하나는 IBM Operational Decision Management 규칙 프로젝트의 결과 매개변수로 사용되어야 하는 규칙 리턴 유형을 나타냅니다.

IBM SPSS Collaboration and Deployment Services Deployment Manager에서 다운로드된 반응 관련 .XSD 파일에는 반응 유형 정의가 포함되어 있으며 복합 데이터 유형 하나만 포함됩니다.

위의 경우 모두에서 리턴 유형(이름에 항상 "response" 접미부가 있음)은 기본적으로 문자열인 value 요소를 정의합니다. 다른 종류의 리턴 값을 갖는 규칙 프로젝트를 작성하려는 경우에는 리턴 유형의 value 요소가 원하는 대로 정의될 수 있도록, IBM Operational Decision Management 규칙 프로젝트를 가져오기 전에 다운로드한 .XSD를 수동으로 편집해야 합니다. 다음 내용은 몇 가지 예제입니다.

```

<xsd:complexType name="claim_dataresponse">
  <xsd:annotation>
    <xsd:appinfo>
      <dmname>claim_dataresponse</dmname>
    </xsd:appinfo>
  </xsd:annotation>
  <xsd:sequence>
    <xsd:element name="value" type="xsd:string"/>
  </xsd:sequence>
</xsd:complexType>

```

```
<xsd:complexType name="claim_dataresponse">
  <xsd:annotation>
    <xsd:appinfo>
      <dmname>claim_dataresponse</dmname>
    </xsd:appinfo>
  </xsd:annotation>
  <xsd:sequence>
    <xsd:element name="value" type="xsd:boolean"/>
  </xsd:sequence>
</xsd:complexType>
```

```
<xsd:complexType name="claim_dataresponse">
  <xsd:annotation>
    <xsd:appinfo>
      <dmname>claim_dataresponse</dmname>
    </xsd:appinfo>
  </xsd:annotation>
  <xsd:sequence>
    <xsd:element name="value" type="xsd:long"/>
  </xsd:sequence>
</xsd:complexType>
```

제 9 장 도움말 가져오기

다음과 같은 세 가지 수준의 사용자 지원이 제공됩니다.

- 사용자 정의 도움말은 IBM Analytical Decision Management와 함께 패키지된 각 사전 빌드된 애플리케이션에 대해 제공됩니다. 도움말을 시작하려면 화면 또는 대화 상자에서 도움말 아이콘을 클릭하십시오.



- 작은 팝업 창에 표시되는 코치 텍스트는 현재 애플리케이션 또는 비즈니스 문제점에 대한 세부사항을 제공합니다. 코치 텍스트는 특정 사이트 또는 애플리케이션에 대해 손쉽게 구성될 수 있는 사용자 정의된, 컨텍스트 도움말의 추가 레이어를 제공합니다. 코치 텍스트는 사용자 인터페이스 전반에서 작은 코치 텍스트 아이콘을 클릭하면 표시됩니다.



- 표준 도구 팁은 페이지 또는 대화 상자 내의 개별 제어에 대해 제공됩니다. 도구 팁을 표시하려면 단추, 링크 또는 기타 제어 위에 커서를 두십시오.

사용 가능한 문서

IBM Analytical Decision Management 문서는 두 가지 형식으로 사용할 수 있습니다. 애플리케이션의 도움말에 액세스하면 온라인 IBM Knowledge Center가 실행되고 관련 도움말 항목이 표시됩니다. 도움말 실행에 문제가 있는 경우 외부 Knowledge Center URL에 액세스할 수 있는 권한이 없다면 관리자에게 문의하십시오.

Knowledge Center는 모든 문서의 중앙 집중화된 온라인 컬렉션입니다. 이는 제품이 실행 중이 아니더라도 항상 사용 가능합니다. 브라우저에서 책갈피 기능을 사용하는 것이 좋습니다. 또한 기타 IBM 제품의 문서도 있습니다(예: IBM SPSS Modeler, IBM SPSS Collaboration and Deployment Services 및 Statistics).

인터넷이 연결되어 있지 않은 경우 또는 유지보수를 위해 Knowledge Center를 임시로 사용할 수 없는 경우에도 모든 문서는 PDF(*.pdf) 형식으로 사용할 수 있습니다. PDF 파일은 설치 매체에서 사용 가능하며 IBM SPSS Collaboration and Deployment Services 설치 디렉토리(예: *C:\Program Files\IBM\SPSS\Deployment\7.0\Server\documentation\DecisionManagement*)에 설치할 수도 있습니다.

웹 <http://www-01.ibm.com/support/docview.wss?uid=swg27038937>에 있는 PDF 문서도 사용할 수 있으며, 릴리스 정보는 <http://www-01.ibm.com/support/docview.wss?uid=swg27038946>에 있습니다.

부록 A. 스트림 공유

IBM Analytical Decision Management에서 작성된 모델과 프로젝트는 스트림 파일로 저장되며 사용 가능한 경우에는 IBM SPSS Modeler에서 수정하고 사용할 수 있습니다. 예를 들어, IBM SPSS Modeler Advantage 또는 다른 IBM Analytical Decision Management 애플리케이션을 사용하여 모델 또는 프로젝트 스트림을 작성한 후 이 스트림을 IBM Analytical Decision Management를 통해 배포하기 전에 IBM SPSS Modeler에서 사용자 정의하려 할 수 있습니다. 또는, 배포 설정이 이 부록에 설명된 바와 같이 지정된 경우에는 IBM SPSS Modeler에서 작성된 스트림을 IBM Analytical Decision Management에서 열 수 있습니다.

이 부록에 설명된 형식을 따르는 모델은 IBM SPSS Modeler Advantage를 사용하여 보고, 편집하고, 스코어링하고, 새로 고칠 수 있습니다. 다른 모델도 IBM SPSS Modeler Advantage에서 스코어링할 수 있으나 올바르게 표시되지 않을 수 있으며 새로 고칠 수 없습니다.

IBM Analytical Decision Management에서 스트림 열기

IBM Analytical Decision Management 또는 IBM SPSS Modeler Advantage에서 스트림을 열 때는 다음 사항에 유의하십시오.

- 배포 유형이 지정되지 않은 경우(없음), 스트림은 열 수 있으나 데이터 탭만 사용 가능합니다.
- 배포 유형이 스코어링 전용일 때는 스코어링 분기가 IBM SPSS Modeler Advantage 또는 IBM Analytical Decision Management에서 지원하는 유형의 단일 데이터 소스 노드로 시작할 경우, 스코어 탭도 사용 가능합니다. 즉 IBM SPSS Modeler에서 지원하는 모든 소스 유형은 스트림에 하나의 소스 노드만 있는 경우(병합된 데이터 세트 없음) IBM SPSS Modeler Advantage 및 IBM Analytical Decision Management에서도 지원됩니다.
- 배포 유형이 모델 새로 고치기일 때는 모델링 분기가 IBM SPSS Modeler Advantage 또는 IBM Analytical Decision Management에서 지원하는 유형의 단일 데이터 소스 노드로 시작하며 선택된 모델링 노드가 자동화된 모델링 노드(예측 모델링을 위해 지원), 의사결정 목록 노드(예측 규칙 모델링을 위해 지원), 군집 노드(군집 모델링을 위해 지원) 또는 연관 규칙 노드(연관 모델링을 위해 지원)인 경우, 모델링 탭도 사용 가능합니다. 또한, 대상 필드와 모델링에 사용되는 기타 정보를 정의하는 유형 노드가 있어야 합니다.

IBM SPSS Modeler에 배포 설정 지정

IBM SPSS Modeler에서, 스트림 특성 대화 상자에 다음과 같이 배포 설정을 지정하십시오.

1. IBM SPSS Modeler 메뉴에서 다음을 선택하십시오.

도구 > 스트림 특성 > 배포

2. 배포 유형 목록에서 스코어링 전용 또는 모델 새로 고치기 중 적절한 것을 선택하십시오.
3. 스코어링 노드, 모델링 노드 및 모델 노드를 필요에 따라 선택하십시오.
4. 검사를 선택하여 모든 필수 옵션이 지정되었는지 확인하거나 저장을 선택하여 스트림을 IBM SPSS Collaboration and Deployment Services Repository에 저장하십시오.

IBM SPSS Modeler Advantage 또는 IBM Analytical Decision Management에서 작성된 스트림의 경우 배포 정보는 자동으로 설정됩니다.

IBM Analytical Decision Management에서 IBM SPSS Modeler 스트림을 사용하는 데 대한 팁

- IBM SPSS Modeler에서 처음부터 작성하는 것보다 IBM SPSS Modeler Advantage 또는 IBM Analytical Decision Management를 사용하여 스트림을 작성한 후 IBM SPSS Modeler에서 이를 수정하는 것이 더 쉽습니다. 이 방식으로 작성된 스트림은 IBM Analytical Decision Management에서 완전히 작동할 수 있도록 구조화됩니다.
- IBM SPSS Modeler Advantage에서 작성된 모든 스트림에서, 수퍼노드는 숙련된 사용자가 사용자 정의할 수 있는 분기 내의 특정 위치에 포함됩니다. 스트림에 대한 사용자 정의나 추가는, 스트림에 유지되면서 IBM SPSS Modeler Advantage에서 이를 인식할 수 있도록 이러한 수퍼노드 내에 추가하는 것이 좋습니다.
- 각 소스 노드에 인접한 다운스트림인 표현식 규칙을 포함하는 로컬 규칙 노드의 필드는 추가/편집 추가 필드 대화 상자에 포함됩니다. 로컬 규칙은 IBM SPSS Modeler가 아닌, IBM Analytical Decision Management 또는 IBM SPSS Modeler Advantage에서만 작성할 수 있다는 점을 참고하십시오.
- 모델링 노드의 파티션 노드 업스트림(그리고 유형 노드의 다운스트림)이 있을 경우에는 평가 및 테스트를 위한 데이터 소스 빌드 시 모델 평가를 수행할 수 있도록 자동으로 데이터 파티션이 IBM SPSS Modeler Advantage에서 사용으로 설정되며 파티션 노드의 사용 설정 여부에 따라 선택란 상태가 설정됩니다.
- 유형 노드의 ADP(Automated Data Preparation) 노드 다운스트림이 있을 경우에는 자동으로 정리 후 신뢰할 수 있는 모델 작성을 위한 데이터 준비가 사용으로 설정되며 ADP 노드의 사용 설정 여부에 따라 선택란 상태가 설정됩니다.
- 규칙 또는 선택 노드가 모델링 노드의 현재 업스트림인 경우에는 이들 노드가 사용할 선택사항 지정 섹션을 채우는 데 사용됩니다.
- 스트림 매개변수가 정의된 경우(스트림 특성 대화 상자의 매개변수 탭) 매개변수 링크가 IBM SPSS Modeler Advantage의 모델링 및 스코어 탭에 표시됩니다. 사용자는 이 링크를 클릭하여 이들 값을 지정할 수 있습니다.

IBM SPSS Modeler에서 IBM SPSS Modeler Advantage 시작

IBM SPSS Modeler 시작 화면에서 IBM SPSS Modeler Advantage로 모델을 작성하도록 선택할 수 있습니다.

부록 B. 내게 필요한 옵션

내게 필요한 옵션 기능은 거동이 불편하거나 시각 장애 등의 신체적 장애가 있는 사용자가 IT 제품을 사용하는 데 도움을 줍니다. IBM은 연령 또는 신체적 능력과 상관 없이 모든 사용자가 사용할 수 있는 제품을 제공하기 위해 노력하고 있습니다. 이 제품은 표준 웹 브라우저 탐색 키를 사용합니다.

IBM Analytical Decision Management는 브라우저 기반 애플리케이션이므로 시각 장애와 같은 장애가 있는 사용자를 위한 내게 필요한 옵션은 웹 브라우저 설정에서 제어됩니다. 키보드 단축키, 스크린 리더 등을 포함한 사용자 인터페이스 조치에 대한 세부사항은 웹 브라우저의 문서를 참조하십시오.

벤더 소프트웨어

IBM Analytical Decision Management에서는 IBM 라이선스 계약에 포함되지 않는 몇 가지 벤더 소프트웨어를 사용해야 할 수 있습니다. IBM은 이러한 제품의 내게 필요한 옵션 기능에 대해 대변하지 않습니다. 해당 제품의 내게 필요한 옵션 정보에 대해서는 해당 벤더에 문의하십시오.

IBM과 내게 필요한 옵션(접근성)

접근성과 관련한 IBM의 약속과 이행에 대해 자세히 알아보려면 IBM Human Ability and Accessibility Center를 참조하십시오.

주의사항

이 정보는 미국에서 제공되는 제품 및 서비스용으로 작성된 것입니다.

IBM은 다른 국가에서 이 책에 기술된 제품, 서비스 또는 기능을 제공하지 않을 수도 있습니다. 현재 사용할 수 있는 제품 및 서비스에 대한 정보는 한국 IBM 담당자에게 문의하십시오. 이 책에서 IBM 제품, 프로그램 또는 서비스를 언급했다고 해서 해당 IBM 제품, 프로그램 또는 서비스만을 사용할 수 있다는 것을 의미하지는 않습니다. IBM의 지적 재산권을 침해하지 않는 한, 기능상으로 동등한 제품, 프로그램 또는 서비스를 대신 사용할 수도 있습니다. 그러나 비IBM 제품, 프로그램 또는 서비스의 운영에 대한 평가 및 검증은 사용자의 책임입니다.

IBM은 이 책에서 다루고 있는 특정 내용에 대해 특허를 보유하고 있거나 현재 특허 출원 중일 수 있습니다. 이 책을 제공한다고 해서 특허에 대한 라이선스까지 부여하는 것은 아닙니다. 라이선스에 대한 의문사항은 다음으로 문의하십시오.

135-700

서울특별시 강남구 도곡동 467-12, 군인공제회관빌딩

한국 아이.비.엠 주식회사

고객만족센터

전화번호: 080-023-8080

2바이트(DBCS) 정보에 관한 라이선스 문의는 한국 IBM 고객만족센터에 문의하거나 다음 주소로 서면 문의하시기 바랍니다.

Intellectual Property Licensing

Legal and Intellectual Property Law

IBM Japan Ltd.

1623-14, Shimotsuruma, Yamato-shi

Kanagawa 242-8502 Japan

다음 단락은 현지법과 상충하는 영국이나 기타 국가에서는 적용되지 않습니다. IBM은 타인의 권리 비침해, 상품성 및 특정 목적에의 적합성에 대한 묵시적 보증을 포함하여(단, 이에 한하지 않음) 묵시적이든 명시적이든 어떠한 종류의 보증 없이 이 책을 "현상태대로" 제공합니다. 일부 국가에서는 특정 거래에서 명시적 또는 묵시적 보증의 면책사항을 허용하지 않으므로, 이 사항이 적용되지 않을 수도 있습니다.

이 정보에는 기술적으로 부정확한 내용이나 인쇄상의 오류가 있을 수 있습니다. 이 정보는 주기적으로 변경되며, 변경된 사항은 최신판에 통합됩니다. IBM은 이 책에서 설명한 제품 및/또는 프로그램을 사전 통지 없이 언제든지 개선 및/또는 변경할 수 있습니다.

이 정보에서 언급되는 비IBM의 웹 사이트는 단지 편의상 제공된 것으로, 어떤 방식으로든 이들 웹 사이트를 옹호하고자 하는 것은 아닙니다. 해당 웹 사이트의 자료는 본 IBM 제품 자료의 일부가 아니므로 해당 웹 사이트 사용으로 인한 위험은 사용자 본인이 감수해야 합니다.

IBM은 귀하의 권리를 침해하지 않는 범위 내에서 적절하다고 생각하는 방식으로 귀하가 제공한 정보를 사용하거나 배포할 수 있습니다.

(i) 독립적으로 작성된 프로그램과 기타 프로그램(본 프로그램 포함) 간의 정보 교환 및 (ii) 교환된 정보의 상호 이용을 목적으로 본 프로그램에 관한 정보를 얻고자 하는 라이선스 사용자는 다음 주소로 문의하십시오.

135-700

서울특별시 강남구 도곡동 467-12, 군인공제회관빌딩

한국 아이.비.엠 주식회사

고객만족센터

이러한 정보는 해당 조건(예를 들면, 사용료 지불 등)하에서 사용될 수 있습니다.

이 정보에 기술된 라이선스가 부여된 프로그램 및 프로그램에 대해 사용 가능한 모든 라이선스가 부여된 자료는 IBM이 IBM 기본 계약, IBM 프로그램 라이선스 계약(IPLA) 또는 이와 동등한 계약에 따라 제공한 것입니다.

본 문서에 포함된 모든 성능 데이터는 제한된 환경에서 산출된 것입니다. 따라서 다른 운영 환경에서 얻어진 결과는 상당히 다를 수 있습니다. 일부 성능은 개발 단계의 시스템에서 측정되었을 수 있으므로 이러한 측정치가 일반적으로 사용되고 있는 시스템에서도 동일하게 나타날 것이라고는 보증할 수 없습니다. 또한 일부 성능은 추정을 통해 추측되었을 수도 있으므로 실제 결과는 다를 수 있습니다. 이 책의 사용자는 해당 데이터를 본인의 특정 환경에서 검증해야 합니다.

비IBM 제품에 관한 정보는 해당 제품의 공급업체, 공개 자료 또는 기타 범용 소스로부터 얻은 것입니다. IBM에서는 이러한 비IBM 제품을 반드시 테스트하지 않았으므로, 이들 제품과 관련된 성능의 정확성, 호환성 또는 기타 주장에 대해서는 확인할 수 없습니다. 비IBM 제품의 성능에 대한 의문사항은 해당 제품의 공급업체에 문의하십시오.

IBM이 제시하는 방향 또는 의도에 관한 모든 언급은 특별한 통지 없이 변경될 수 있습니다.

이 정보에는 일상의 비즈니스 운영에서 사용되는 자료 및 보고서에 대한 예제가 들어 있습니다. 이들 예제에는 개념을 가능한 완벽하게 설명하기 위하여 개인, 회사, 상표 및 제품의 이름이 사용될 수 있습니다. 이들 이름은 모두 가공의 것이며 실제 기업의 이름 및 주소와 유사하더라도 이는 전적으로 우연입니다.

이 정보를 소프트카피로 확인하는 경우에는 사진과 컬러 삽화가 제대로 나타나지 않을 수도 있습니다.

상표

IBM, IBM 로고 및 ibm.com은 전세계 여러 국가에 등록된 International Business Machines Corp.의 상표 또는 등록상표입니다. 기타 제품 및 서비스 이름은 IBM 또는 타사의 상표입니다. 현재 IBM 상표 목록은 웹 "저작권 및 상표 정보"(www.ibm.com/legal/copytrade.shtml)에 있습니다.

Adobe, Adobe 로고, PostScript 및 PostScript 로고는 미국 또는 기타 국가에서 사용되는 Adobe Systems Incorporated의 상표 또는 등록상표입니다.

Intel, Intel 로고, Intel Inside, Intel Inside 로고, Intel Centrino, Intel Centrino 로고, Celeron, Intel Xeon, Intel SpeedStep, Itanium 및 Pentium은 미국 또는 기타 국가에서 사용되는 Intel Corporation 또는 그 계열사의 상표 또는 등록상표입니다.

Linux는 미국 또는 기타 국가에서 사용되는 Linus Torvalds의 등록상표입니다.

Microsoft, Windows, Windows NT 및 Windows 로고는 미국 또는 기타 국가에서 사용되는 Microsoft Corporation의 상표입니다.

UNIX는 미국 및 기타 국가에서 사용되는 The Open Group의 등록상표입니다.

Java 및 모든 Java 기반 상표와 로고는 Oracle 및/또는 그 계열사의 상표 또는 등록상표입니다.

기타 제품 및 서비스 이름은 IBM 또는 타사의 상표입니다.

색인

[가]

- 개요
 - 관리 85
- 갤러리
 - 완료되지 않은 작업 27
 - 장기간 작업 27
- 공유 규칙 56
- 관리
 - 개요 85
 - 데이터 옵션 87
 - 스코어링 대상 옵션 88
 - 애플리케이션 단축키 87
 - 일반 옵션 86
 - 홈 페이지 87
- 관리 개요 85
- 군집 모델링 예제 19
- 군집 모형 65, 70
- 규칙 53
 - 공유 56, 58
 - 내보내기 56
 - 선택 55
 - 선택 논리 55
 - 세그먼트 53
 - 세그먼트 규칙과 함께 모델 사용 53
 - 외부 58
 - 작성 53
 - 재사용 56, 58
 - 주석 58
- IBM Operational Decision Management 58
- 규칙 제외 55
- 규칙 포함 55
- 글로벌 값
 - 표현식에서 50

[나]

- 날짜 형식 27
- 내 데이터 소스 33
- 내게 필요한 옵션 101
- 내보내기 규칙 56

[다]

- 단축키
 - 애플리케이션 25
- 대상 필드
 - 모델 작성 15, 61
- 대화형 목록 모델 63, 69
- 데모 아티팩트 12
- 데이터 개요 44
- 데이터 미리보기 44
- 데이터 보기 40
- 데이터 소스
 - 데이터 개요 44
 - 데이터 보기 40
 - 데이터베이스 39
 - 모델 결과 추가 47
 - 미리보기 44
 - 세그먼트 규칙을 사용한 필드 작성 47
 - 입력 필드 44
 - 정의 33, 35
 - 측정 수준 45
 - 텍스트 기반 36
 - 파생 탭 47
 - 파일 소스 36
 - 표현식 빌더 48
 - 표현식 추가 47
 - 필드 값 편집 46
 - 필드 맵핑 46
 - 필드 추가 47
 - 필드 파생 47
- Analytic Server 43
- Excel 36
- Hadoop 43
- HDFS 43
- IBM Cognos BI 41
- IBM SPSS Collaboration and Deployment Services 엔터프라이즈 뷰 40
- IBM SPSS Statistics 36
- 데이터 옵션 87
- 데이터 옵션 잠금 87
- 데이터 유형 45
- 데이터 파일 12
- 데이터베이스 소스 39

- 도구 팁 97
- 도움말 97
- 도움말 가져오기 97
 - 사용 가능한 문서 97

[마]

- 매개변수
 - 표현식에서 50
- 명목 필드의 최대 멤버 수 32
- 모델
 - 결과 62
 - 결과 필드 선택 82
 - 군집 65
 - 수동 군집 66
 - 규칙에서 사용 53
 - 대상 필드 15, 61
 - 대화형 목록 63
 - 데이터 선택 78
 - 데이터베이스 대상 선택 78
 - 레코드 선택 78
 - 선택적 설정 68, 69, 70
 - 수익 시뮬레이션 73
 - 스코어링 18, 77, 83, 84
 - 스코어링 대상 선택 78
 - 스코어링 요약 보고서 84
 - 스코어링 필드 맵핑 81
 - 시간 빌드 제한 27
 - 연관 66
 - 예측 61
 - 예측 규칙 63
 - 작성 15, 61
 - 차트 62
 - 테스트 18, 74
 - 파일 대상 선택 79
 - 평가 16, 72
 - 환경 설정 27
 - Analytic Server 대상 선택 80
 - Cognos BI 서버 대상 선택 80
- 모델 결과 추가 47
- 모델 내 결과 필드 82
- 모델 내 데이터 선택 78
- 모델 내 데이터베이스 대상 78
- 모델 내 스코어링 대상 78

- 모델 내 차트 62
- 모델 내 파일 대상 79
- 모델 내 Cognos BI 서버 대상 80
- 모델 내의 레코드 선택 78
- 모델 스코어링 18, 77, 83
 - 결과 필드 선택 82
 - 데이터 선택 78
 - 데이터베이스 대상 78
 - 스코어링 대상 선택 78
 - 파일 대상 79
 - 필드 맵핑 81
 - Analytic Server 대상 80
 - Cognos BI 서버 대상 80
- 모델 작성
 - 대상 필드 15
- 모델 테스트 18, 74
- 모델 평가 16
 - 군집 모형 19, 20
 - 대화형 목록 모델 17
 - 예측 규칙 모델 17
- 모델에서의 수익 시뮬레이션 73
- 모델의 평가 72
- 모델의 Analytic Server 대상 80

[바]

- 버전 레이블 31
- 버전 레이블 지정 31
- 변수중요도 차트 62
- 분포도 62, 72
- 브라우저 언어 변경 25
- 비즈니스 규칙 53

[사]

- 사용 가능한 문서 97
- 사용자 환경 설정 27
- 사전 빌드된 애플리케이션 12
- 새 기능 1, 2
- 새로운 기능 1, 2
- 샘플 데이터 12
- 샘플 애플리케이션 12
- 샘플 애플리케이션용 데이터 12
- 샘플 애플리케이션용 아티팩트 12
- 샘플 애플리케이션용 파일 12
- 샘플 파일 12
- 선택 규칙 55
- 설명 32

- 세그먼트 규칙
 - 세그먼트 규칙과 함께 모델 사용 53
- 세그먼트 규칙을 사용한 필드 작성 47
- 세그먼트 규칙을 사용한 필드 추가 47
- 수동 군집 66
- 수익
 - 시뮬레이션 16
 - 최대화 11
- 스코어링 대상 옵션 88
- 스코어링 대상 옵션 잠금 88
- 스코어링 요약 보고서 84
- 스코어링 필드 맵핑 81
- 스트림
 - 업로드 29
 - 스트림 업로드 29
 - 스트림 파일 업로드 29
- 시간 형식 27
- 시뮬레이션
 - 기본 날짜 27
 - 시뮬레이션 날짜 27
- 시작 페이지 13, 25, 87
 - 사용자 정의 25
 - 애플리케이션 추가 25

[아]

- 아이콘, IBM Cognos BI 43
- 애플리케이션
 - 시작 13
 - 테스트 74
- 애플리케이션 단축키 25, 87
- 애플리케이션 단축키 제어 87
- 애플리케이션 시작 페이지 25, 87
 - 사용자 정의 25
 - 애플리케이션 추가 25
- 애플리케이션 테스트 74
- 애플리케이션 홈 페이지 27
- 애플리케이션 URL 25
- 언어 25
- 언어 변경 25
- 연관 모델 66, 70
 - 트랜잭션 데이터 67
 - 표 형식 대 트랜잭션 데이터 67
 - 표 형식 데이터 67
- 연관 모델링 예제 20
- 연산자
 - 표현식에서 49
- 예측 규칙 모델 63, 69

- 예측 모델 61
- 예측변수중요도 차트 62
- 오류 분포도 72
- 온라인 도움말 97
- 완료되지 않은 항목
 - 갤러리 27
- 외부 규칙
 - 애플리케이션에서 사용 58
- 일반 옵션 86
- 일치 교차표 차트 72
- 입력 필드
 - 맵핑 46
 - 정의 44

[자]

- 제어
 - 일반 옵션 86
- 주석
 - 규칙에 추가 58
- 직접 메일 예제 11

[차]

- 추가 필드 추가/편집 47
- 추가 필드 편집 47
- 측정 수준 45

[카]

- 코치 텍스트 97
- 키워드 29, 32

[타]

- 텍스트 기반 데이터 파일 36
- 통화 형식 27
- 특성
 - 리포지토리 32
 - 명목 필드의 최대 멤버 수 32
 - 설명 32
 - 키워드 32
 - 프로젝트 32

[파]

- 파생 탭 47
- 파생 필드 추가/편집 47

파생 필드 편집 47
 파일 열기 29
 파일 저장 29, 30, 31
 표현식 빌더 47, 48
 표현식 추가 47
 표현식 편집기 48
 표현식 확인 50
 프로젝트 데이터 모델 33
 프로젝트 데이터 소스 33
 프로젝트 잠금 31
 필드
 맵핑 46
 파생 48
 표현식에서 50
 필드 값 편집 46
 필드 맵핑 46, 81
 필드 추가 47
 필드 파생 48

[하]

함수
 표현식에서 49
 홈 페이지 27, 87
 홈 페이지 숨기기 87
 환경 설정 27

A

Analytic Server 데이터 43

C

Cognos BI 소스 41

E

Excel 데이터 파일 36

G

Gains 차트 16, 72

H

Hadoop 데이터 43
 HDFS 데이터 43

I

IBM Analytical Decision Management와
 IBM SPSS Modeler 간 스트림 공유 99
 IBM Cognos BI 소스 41
 IBM Cognos BI 소스 노트
 아이콘 43
 IBM Operational Decision Management 규
 칙
 애플리케이션에서 사용 58
 IBM SPSS Collaboration and Deployment
 Services 리포지토리
 저장 29, 30, 31
 IBM SPSS Collaboration and Deployment
 Services 엔터프라이즈 뷰 데이터 소스 40
 IBM SPSS Modeler Advantage
 예제 11
 IBM SPSS Statistics 데이터 파일 36

L

Lift 차트 72

O

Operational Decision Management
 규칙 58

P

Profit 차트 16, 72

R

Response 차트 72
 ROI 차트 72

U

URL
 애플리케이션 25

