SAS® Enterprise Miner™

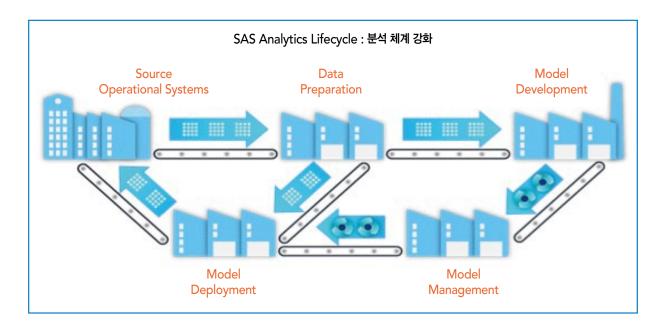
> fact sheet

신뢰성 높은 예측을 위한 보다 정확한 분석 모델 생성

대량의 데이터를 유용한 정보로 바꾸는 것은 대부분의 조직에 어려운 일입니다. 주요 기회를 식별해 낼 해답과 상관 관계가 그 많은 데이터 어딘가에 숨어 있습니다.

- 어느 고객이 어떤 제품을 언제 구입하게 될까요?
- 어느 고객이 이탈하며 이들의 이탈을 막기 위해 할 수 있는 일은 무엇일까요?
- 수익성을 보장하려면 보험 비율을 어떻게 설정해야 할까요?
- 고장을 예측하고 불필요한 유지관리를 줄이고 가동 시간을 늘리고 자산 성능을 최적화할 수 있는 방법은 무엇일까요?

복잡한 질문에 대한 답을 얻고 경쟁이 치열한 오늘날의 시장에서 우위를 점하려면 성능이 뛰어난 고급 분석 솔루션이 필요합니다. 이전에 알려지지 않은 패턴을 발견하면 기업 전반의 의사 결정자들이 효과적인 전략을 수립하는 데 도움이 될 수 있습니다. SAS® 데이터 마이닝으로 비즈니스 프로세스를 구현하는 기업은 오늘날 빠르게 변화하는 시장에서 경쟁력을 유지할 수 있습니다.



SAS[®] Enterprise Miner™ 주요 특징

• 핵심 관계 파악 및 직관적이고 신속한 모델 개발

그래픽 인터페이스를 사용해 분석 전문가들은 모델링 괴정에서 정보와 손쉽게 인터랙션할 수 있습니다. 데이터 마이닝 프로세스의 일반적이고 해석하기 쉬운 시각적 뷰는 분석 전문가와 비즈니스 분석가 모두 간편하게 사용할 수 있으며 협업을 통해 어려운 문제를 해결해 낼 수 있습니다.

• 다목적 데이터 마이닝 워크벤치로 더욱 효과적인 모델을 효율적으로 구축

인터랙티브 방식의 자체 문서화 프로세스 플로우 다이어그램 환경이 모델 개발 시간을 단축시킵니다. 데이터 마이닝 프로세스를 효율적으로 매핑하여 최상의 결과를 이끌어냅니다.

• 자동화된 방식으로 손쉬운 인사이트 도출

통계학적 지식이 적은 비즈니스 분석가와 업무 전문가도 SAS® Rapid Predictive Modeler를 사용하여 모델을 자동으로 생성하고 신속하게 조치를 취할 수 있습니다. 분석 결과는 이해하기 쉬운 도표에 표시되어 의사 결정 능력을 향상시켜 줍니다.

• 예측의 정확도를 끌어올려 올바른 결정 지원

우수한 성과를 내는 모델은 예측의 안정성과 정확성을 높여줍니다. 이는 시각적 모델 평가 및 유효성 검사 지표에 의해 쉽게 확인할 수 있습니다. 모델링에 사용된 타겟변수에 예측 변수가 어떤 영향을 미치는지 파악할 수 있도록 모델 프로파일링도 제공됩니다.

• 보다 빠른 결과를 위한 간편한 모델 구축 및 스코어링 프로세스

SAS[®] Enterprise Miner[™]는 새로운 데이터를 스코어링하는 지루한 작업을 자동화하고 모델 개발의 모든 단계를 커버하는 스코어링 코드를 제공합니다. 다양한 실시간 환경 혹은 배치(batch) 환경에 스코어링 코드를 배포할 수 있습니다. 이렇게 하여 시간을 단축하고 더 정확한 결과를 얻을 수 있기 때문에 사용자는 가치를 극대화하는 의사 결정을 내릴 수 있게 됩니다.



SAS® Enterprise Miner™ 주요 기능

누구나 안전하고 확장 가능한 방식의 통합 분석을 통해 이점을 누릴 수 있습니다. 그러나 이러한 통합 분석을 사용하려면 조직 전반에 걸쳐 협업이 이루어져야 하고 다양한 요구를 충족하도록 맞춤화할 수 있는 뛰어난 성능의 범용 데이터 마이닝 솔루션이 필요합니다.

하나의 분석 접근법이 특정 데이터 수집 방식과 잘 맞더라도 새로운 데이터 소스와는 원활히 작동하지 않을 수있고 새로운 비즈니스 질문에 응답하지 못하는 경우가 발생할 수도 있습니다. 따라서 다양한 분석 도구를 손쉽게 선택할 수 있도록 하는 것이 매우 중요합니다. 다른 도구가다른 모델을 생성하면 도출된 모델의 예측 결과와 평가통계를 나란히 표시하여 어떤 데이터 마이닝 접근법이 최상의 "적합성"을 나타내는지 비교할 수 있습니다. 분석도구가 제한적인 워크벤치로 시작하게 되면 도출된 모델도 제한된 예측 값을 가질 수 있습니다. (예: 회귀 분석 혹은 의사 결정 트리만 있는 경우)

SAS® Enterprise Miner는 분산 클라이언트 / 서버 시스템으로 전달됩니다 최적화된 아키텍처를 제공하기 때문에 데이터 마이너와 비즈니스 분석가가 보다 신속하게 작업하여 정확한 예측 및 기술 분석 모델을 생성하고 결과를 도출할 수 있으며, 이 결과를 다른 사람들과 공유하고 비즈니스 프로세스에 통합할 수 있습니다. 데이터 마이닝 프로세스를 향상시키기 위해 이 소프트웨어는데이터 통합, 분석 및 보고서 작성 등의 다른 SAS 기술과도 원활하게 작동할 수 있도록 설계되어 있습니다.

데이터에 대한 통합되고 완전한 뷰 제공

데이터 마이닝은 매우 다양한 기업 소스에서 수집된 데이터를 다루는 통합된 정보 전달 전략에 사용하는 경우 가장 효과적입니다. 콜센터 로그, 설문 조사 결과, 고객 피드백 양식, 웹 데이터, 시계열 데이터 및 판매 시점 거래 데이터는 모두 결합될 수 있으며 업계에서 가장 수준 높은 데이터 마이닝 패키지로 분석될 수 있습니다. 여기에 SAS® Text Miner를 추가로 사용하면 정형 및 비정형데이터를 함께 분석하여 보다 정확하고 완전한 결과를 얻을 수 있습니다.

사용하기 쉬운 GUI

분석 전문가의 편의성을 위해 사용하기 쉬운 드래그 앤드롭 (drag-and-drop) 인터페이스가 설계되어 있습니다. 고급 분석 알고리즘은 성공적인 데이터 마이닝 작업에서 수행된 핵심 작업에 따라 구성되어 있습니다. SAS 데이터 마이닝 프로세스는 샘플링, 탐색, 수정, 모델링, 평가 (SEMMA)라는 다섯 가지 주요 단계로 이루어져 있습니다. 사용자는 각 단계에서 데이터 마이닝 프로젝트가 개발됨에 따라 일련의 작업을 수행하게 됩니다. SEMMA 도구 모음에서 노드를 배치하여, 사용자는 고급 통계를 적용하고 가장 중요한 변수를 식별하고 수식 작성기로 데이터를 변환하고 모델을 개발하여 결과를 예측하고 정확도를 검증하고 예측 값이 포함되어 있는 스코어링된 데이터 세트를 생성하여 운영 애플리케이션에 배포합니다.

쉽고 빠르며 스스로 할 수 있는 모델 생성 방법

SAS® Rapid Predictive Modeler를 사용하면 비전문가도 데이터 마이닝 작업 흐름 (예: 데이터 변환, 변수 선택, 다양한 알고리즘 적용, 모델 평가) 을 통해 광범위한 비즈니스 문제에 대한 예측 모델을 자동으로 신속하게 생성할 수 있습니다. SAS® Rapid Predictive Modeler는 SAS® Enterprise Guide®에서 실행되거나 Microsoft Office (Microsoft Excel 전용)의 SAS 추가 기능으로 작동하며 기본 제공되는 SAS® Enterprise Miner 모델링 단계를 사용합니다 협업 접근법을 통해 SAS® Enterprise Miner를 사용하는 분석 전문가는 SAS® Rapid Predictive Modeler로 개발된 모델을 수정하여 사용할 수 있습니다.

기존 방식과 최신 모델링 기법 모두 지원

SAS® Enterprise Miner는 통계, 데이터 마이닝, 기계학습 알고리즘 등을 활용하여 깊이 있는 뛰어난 분석을 제공합니다. 여기에는 의사 결정 트리, 배깅과 부스팅, 시계열 데이터 마이닝, 뉴럴 네트워크, 메모리 기반 추론 (MBR), 계층 구조 클러스터링, 선형 및 로지스틱 회귀 분석, 연관성 분석, 시퀀스 및 웹 경로 분석이 모두 포함되어 있습니다. 뿐만 아니라, 지원되는 분석 알고리즘의 폭이 Credit Scoring 등의 산업에 특화된 알고리즘이나 그래디언트 부스팅 및 Least Angle Regression Spline(LARS) 등의 최신 기법으로도 확장됩니다.

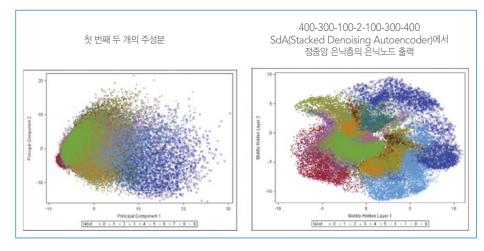


그림 1. 데이터 마이닝의 중간 과정에서 나오는 은닉층을 위한 신경망과 차원 축소를 위한 주성분 분석을 수행합니다

정교한 데이터 준비, 요약 및 탐색

데이터 준비는 모든 데이터 마이닝 작업 중에서 많은 시간이 소요되는 부분입니다. 인터랙티브 방식의 데이터 준비를 위한 강력한 도구 세트는 결측값을 다루고 이상점을 필터링하고 세그멘테이션 규칙을 개발하는 데 유용합니다. 코어 데이터 준비 도구에는 파일 가져오기와 추가하기, 병합하기 그리고 변수 버리기가 포함되어 있습니다. 광범위한 기술 통계량 요약 기능과 인터랙티브 방식의 탐색 도구는 초보 사용자도 동적으로 연결된 다차원 플롯에서 대량의 데이터를 탐색할 수 있도록 지원합니다 따라서 특정 비즈니스 문제에 맞춤화되고 가장 적합한 고품질의 데이터 마이닝 결과를 얻을 수 있습니다.

비즈니스 기반의 모델 비교, 리포팅 및 관리

평가 기능을 사용하면 모델을 비교하여 최상의 리프트와 평균 ROI를 얻을 수 있는 모델을 식별할 수 있습니다. 또한 시각화가 뛰어난 평가 인터페이스를 사용하여 서로 다른 알고리즘으로 생성된 모델을 일관성 있게 평가할 수 있습니다 데이터 마이너는 결과에 대해 비즈니스 도메인 전문가와 함께 논의함으로써 협업을 강화하고 보다 나은 결과를 얻을 수 있습니다 뿐만 아니라 혁신적인 컷오프 노드는 비즈니스 문제를 해결할 수 있는 최적의 액션을 정의하기 위해 사후 확률 분포를 이용합니다.

개방형의 확장 가능한 설계를 통해 유연성 제공

SAS® Enterprise Miner의 맞춤형 환경에서는 도구를 추가하고 개인화된 SAS 코드를 포함할 수 있습니다.
SAS® Enterprise Miner 이외의 환경에서 개발된 기존 SAS 모델이 프로세스 흐름 환경에 쉽게 통합되고 각 구문에 대한 완전한 제어를 관리합니다. 확장 노드에는 트레이닝과 스코어 코드에 대한 인터랙티브 방식의 편집기 기능이 포함되어 있습니다. 사용자는 로그와 출력 목록을 보면서 코드를 인터랙티브 방식으로 편집하고 실행할 수 있습니다 또한 기본 선택 목록은 SAS 코드나 XML 로직으로 작성한 사용자 정의 개발 도구로 확장되어 분석가가 SAS의 전체 제품군에서 확인할 수 있습니다.

오픈소스 통합 노드

이제 SAS® Enterprise Miner 프로세스 흐름도 안으로 R 언어 코드를 쉽게 가져와서 통합할 수 있습니다. 이를 통해 사용자는 데이터 변환과 탐색뿐만 아니라 R의 지도 (Supervised) 및 비지도 (Unsupervised) 모델에 대한 학습과 스코어링을 수행할 수 있습니다. 그리고 나서 원활하게 결과를 통합하고 R 모델을 평가하고 이것을 SAS® Enterprise Miner로 생성된 모델과 비교할 수 있습니다.

인데이터베이스 (In-database)와 인하둡 (In-Hadoop) 스코어링으로 보다 빠른 결과 제공

스코어링은 운영 환경에서 구현 가능하도록 모델을 정기적으로 새 데이터에 적용하는 프로세스입니다. 이 프로세스는 지루한 작업이 될 수 있습니다. 특히 코드를 수동으로 작성하거나 변환해야 할 경우, 모델 구현 속도가 지연되고 잠재적으로 비용이 많이 드는 오류가 발생할 수도 있습니다. SAS® Enterprise Miner는 SAS, C, Java 및 PMML에서 스코어 코드를 자동으로 생성합니다. 스코어링 코드는 SAS 내에서, 웹 상에서 또는 직접적으로 관계형 데이터베이스나 하둡에서 다양한 실시간 환경 또는 배치(batch) 환경에 배포할 수 있습니다.

SAS® Scoring Accelerator (Hadoop, Pivotal Greenplum, DB2, IBM Netezza, Oracle, Teradata 및 SAS® Scalable Performance Data Server에서 사용 가능) 와 결합된 SAS® Enterprise Miner 모델은 데이터베이스에서 직접 실행 가능한 데이터베이스별 스코어링 기능으로 게시될 수 있습니다. 그 결과를 다른 SAS 솔루션에 전달하여 실시간 운영 환경에 데이터 마이닝 결과를 배포할 수 있습니다.

병렬형 그리드 지원 워크벤치

단일 사용자 시스템에서 자바 클라이언트와 SAS 서버 아키텍처가 있는 대규모 엔터프라이즈 솔루션으로 규모를 변경할 수 있습니다. 강력한 서버가 컴퓨팅에 집중하는 한편, 사용자는 사무실에서 집으로, 또는 원격 장소로 이동하면서 마이닝 프로젝트나 서비스에 대한 액세스를 유지할 수 있습니다. 데이터 정렬, 요약, 변수 선택, 회귀 모델링 등 수많은 프로세스 집약적인 작업이 멀티스레드로 처리되고, 프로세스는 서버 그리드 전반적으로 균형 있는 워크로드와 분포를 위해 병렬로 실행되거나 배치 프로세스로 예약될 수 있습니다.

대기업에 적합한 분산형 데이터 마이닝 시스템

SAS® Enterprise Miner는 씬 (thin) 클라이언트 웹 포털을 통해 배포되는데, 최소한의 클라이언트 관리로 여러 사용자에게 배포할 수 있습니다. 또는 독립형 PC에 전체 시스템을 구성할 수도 있습니다. SAS® Enterprise Miner는 Windows 서버와 UNIX 플랫폼을 지원하므로 대규모 데이터 마이닝 프로젝트를 진행하는 조직이 선택할 수 있는 소프트웨어입니다 모델 결과 패키지는 SAS® Model Manager, SAS Data Integration Studio (SAS Data Integration 구성요소) 및 SAS® Enterprise Guide에서 사용하기 위해 SAS® Metadata Server에 생성하고 등록합니다.

하이퍼포먼스 데이터 마이닝

하이퍼포먼스 데이터 마이닝 노드의 일부 세트가 SAS Enterprise Miner에 포함되어 있습니다. 따라서 데이터와 분석 복잡도에 따라, 사용자는 싱글머신 SMP 모드에서 성능을 높일 수 있습니다. 앞으로는 보다 빠르게 빅데이터를 처리해야 하기 때문에 사용자는 정확한 예측모델을 적시에 개발할 수 있는 별도의 라이선스 대상 제품, SAS® High-Performance Data Mining이 필요합니다. 코딩 환경을 선호하는 사람들은하이퍼포먼스 데이터 마이닝 프로시저 (SAS® Procedures)를 사용할 수 있습니다. 뿐만 아니라,데이터 마이닝 프로그램의 완벽한 사용자 정의를 위해제공되는 옵션도 많습니다. 보다 자세한 내용은sas.com/hpdatamining에서 확인할 수 있습니다.

직관적인 인터페이스

- 프로세스 흐름도를 만드는 데 사용되는 간편한 GUI:
- 더 많고 더 우수한 모델을 더 빠르게 구축
- 웹을 통한 결과 제공
- SAS 프로그래밍 환경에 액세스
- XML 다이어그램 변환 제공
- 다른 프로젝트나 사용자를 위해 다이어그램을 템플릿으로 재사용
- 특정 데이터 마이닝 프로젝트나 다이어그램을 직접 로드, 또는 최신 프로젝트나 다이어그램을 포함하는 프로젝트 내비게이터 트리에서 선택
- 배치 처리 (프로그램 개발 인터페이스):
- GUI의 모든 기능 캡슐화
- SAS 매크로 기반
- 사용자 정의 애플리케이션에 포함되는 학습 및 스코어링 프로세스

확장 가능한 프로세싱

- 서버 기반 프로세싱
- 그리드 컴퓨팅, 인데이터베이스 및 인메모리 프로세싱 옵션
- 비동기식 모델 학습 (Training)
- 프로세싱을 완전히 중지하는 기능
- 병렬 프로세싱 여러 도구와 다이어그램을 동시에 실행.
- 멀티스레드 처리된 예측 알고리즘
- 서버에 있는 모든 스토리지

데이터 액세스 및 관리

- 정형 데이터와 비정형 데이터 소스를 액세스하고 통합 (후보 예측 변수로서 시계열 데이터, 마켓 바스켓, 웹 경로 및 서베이 데이터 포함)
- Microsoft Excel, 쉼표로 구분된 파일 (CSV), SAS 및 기타 일반 파일 형식에 쉽게 액세스하기 위한 파일 가져오기 노드
- 특수 문자가 있는 변수 지원
- SAS® Library Explorer 및 Library Assignment 마법사
- 테이블 목록을 신속히 찾아 보거나 인터랙티브 방식의 그래프 구성요소를 사용하여 플롯을 개발하기 위한 향상된 탐색기 창
- 변수 제거 노드
- 데이터 병합 노드
- 추가 노드
- 이상치 필터링:

- 극단적인 구간 값을 제거하기 위해 다양한 분포 임계치 적용
- 발생 빈도가 n보다 적은 클래스 값 통합
- 클래스 값과 숫자 값을 인터랙티브 방식으로 필터링
- 역할, 측정 레벨, 순서 등의 열 메타데이터를 수정하기 위한 메타데이터 노드
- SAS 메타데이터 서버를 통한 Microsoft Office용 SAS 추가 (AMO), SAS® Data Integration Studio, SAS® Enterprise Guide, SAS® Model Manager와 통합:
- 마이닝을 위한 트레이닝 테이블 만들기
- 스코어링 코드 배포

샘플링

- 단순 랜덤
- 층화
- 가중
- 집락
- 계통 - First N
- 오버 샘플링(Rare Event Sampling)
- 층화 및 이벤트 레벨 샘플링

데이터 분할

- 분석용, 검증용, 평가용 데이터셋 생성
- 홀드이웃 데이터 사용을 통해 모델의 우수한 일반화 보장
- 클래스 타겟별 기본적인 계층화
- 모든 클래스 변수별 균형 잡힌 분할
- SAS 테이블 또는 뷰 출력

변환

- 단순: 로그, 로그 10, 제곱근, 역, 제곱, 지수 및 표준화
- 구간화 (Binning) : 타겟과의 관계에 따른 버킷, 분위수 및 최적의 구간화
- 최고 검정력: 정상 상태 최대화, 타겟과의 상관 관계 최대화, 타겟 레벨 분배 동등화
- 교호작용 (interation) 편집기 : 다항식과 n번째 등급 교호작용 효과 정의
- 인터랙티브 방식으로 변환 정의:
- 표현식 작성기 (Expression Builder) 또는 SAS 코드 편집기를 사용하여 맞춤형 변환 정의
- 새 변수와 원래 변수의 분포 비교
- 재사용을 위해 전역 변환 코드 미리 정의

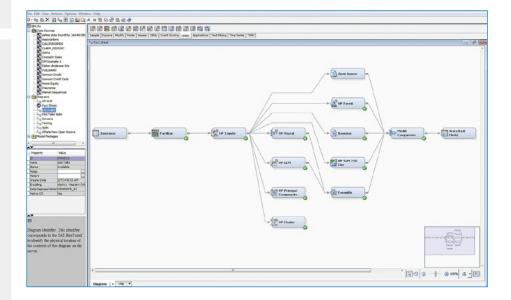


그림 2.

SAS Enterprise Miner GUI에서 프로세스 흐름도는 자체 문서화 템플릿으로, 새로운 문제에 쉽게 업데이트 및 적용할 수 있고 다른 모델러나 분석가와 공유할 수 있습니다.

인터랙티브 방식의 변수 구간화

- 분위수 또는 버킷
- 지니 (Gini) 통계량에 의한 변수 선택
- 결측값을 별도 그룹으로 처리
- 세밀한 (Fine) 구간화와 굵은 (Coarse) 구간화 (재구간화)
- 타겟별 구간그룹의 프로파일
- 인터랙티브 방식으로 그룹 수정
- 구간화 정의 저장

Rules Builder 노드

- 즉석 (ad hoc) 데이터 기반의 규칙과 정책 생성
- 결과에 대한 경로와 타겟 변수의 값을 인터랙티브 방식으로 정의

그래프/시각화

- 배치 및 인터랙티브 방식의 플롯: 산점도, 행렬플롯, 상자그림, constellation, 등고선, 바늘, 격자, 밀도 및 다차원 플롯; 3D, 파이 및 영역 막대형 차트; 히스토그램
- 세그먼트 프로파일 플롯:
- 클러스터링 및 모델링에 의해 생성된 데이터의 세그먼트를 인터랙티브 방식으로 프로파일링
- 그룹 간의 차이와 프로파일을 결정하는 변수를 쉽게 식별
- 사용하기 쉬운 Graphics Explorer 마법사와 Graph Explore 노드:

- 제목과 각주 만들기
- WHERE 절 적용
- 색 구성표에서 선택
- 축 크기를 쉽게 조정
- 표준 SAS[®] Enterprise Miner 결과에서 기본 데이터를 표면화하여 맞춤형 그래프 개발.
- 플롯과 테이블이 인터랙티브 방식으로 연결되고, 브러싱과 밴딩 등의 작업 지원
- 데이터와 그래프를 쉽게 복사하여 다른 애플리 케이션에 붙여 넣거나 BMP 파일로 저장할 수 있음
- 인터랙티브 방식의 그래프가 자동으로 노드 결과 창에 저장됨

데이터 대체

- 중심성 지수
- 배포 기반
- 대체분할을 사용한 트리방식의 결측값 대체
- 중간 매체 간격
- 강건 M-추정량
- 기본 상수
- 대체 편집기:
- 클래스 변수에 대한 새로운 값 지정
- 알 수 없는 값에 대해 대체 값 할당
- 연속형 변수의 대체 임계값 상한설정에 인터랙티브 방식 적용

기술 통계

- 단변량 통계량 및 플롯:
- 연속형 변수: 관측치 수(n), 평균, 중앙값, 최소값, 최대값, 표준 편차, 배율 편차 및 결측 퍼센트
- 클래스 변수: 범주 수, 통계, 모드, 퍼센트 모드 및 누락 퍼센트
- 배포 플롯
- 클래스 대상의 각 수준에 대한 통계 분석
- 이변량 통계량 및 플롯:
- 순서화된 Pearson 및 Spearman 상관 관계 플롯
- 연속형 입력변수를 n 개로 구간화하기 위한 옵션을 갖춘 순서화된 카이 제곱 플롯
- 변동계수 (CV) 플롯
- Logworth를 통한 변수 선택
- 기타 인터랙티브 방식의 플롯:
- 타겟을 설명하는 입력변수의 가치를 기반으로 그들의 순위를 정하는 변수 가치 플롯
- 타겟이나 세그먼트 변수에 걸친 범주형 변수의 분포
- 배율 평균 편차 (Scaled mean deviation) 플롯

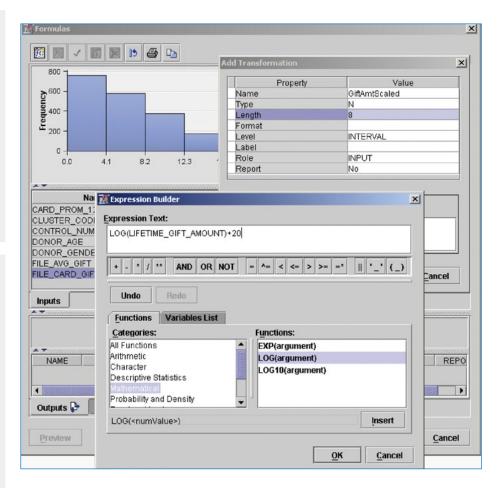


그림 3.

인터랙티브 방식의 변환 변수 노드에서 표현식 작성기를 사용하여 변수를 맞춤형으로 변환합니다.

클러스터링 및 자기 조직화 지도

- 클러스터링:
- 사용자 정의 또는 자동으로 최적의 클러스터를 선택
- 클래스 변수를 분석에 인코딩하는 데 효과적인 몇 가지 전략
- 결측값 처리
- 세그먼트 프로파일 변수 플롯은 각 클러스터 별로 입력변수 및 기타 팩터들의 분포를 보여줌
- 의사 결정 트리 프로파일은 입력을 사용하여 클러스터 멤버십을 예측
- PMML 스코어 코드
- 자기 조직화 지도:
- 배치 SOM은 Nadaraya-Watson 또는 로컬 선형 평활 옵션을 가짐
- 코호넨 네트워크
- 다른 변수의 분포를 지도에 오버레이
- 결측값 처리

마켓 바스켓 분석

- 연관성 및 시퀀스 발견:
- 신뢰도 기준으로 정렬된 규칙들의 그리드 플롯
- 기대 신뢰도 vs 신뢰도의 산점도
- 규칙들에 대한 향상도, 신뢰도, 기대 신뢰도 및 지지도의 통계량 라인 플롯
- 주어진 지지도 및 신뢰도 범위에 대한 빈도수의 히스토그램
- 규칙 설명 테이블
- 규칙에 대한 네트워크 플롯
- 향상도, 신뢰도, 지지도, 연결항목 수 등에 근거하여 인터렉티브하게 규칙들을 서브셋
- 강화된 예측 모델링을 대해 다른 입력변수와 규칙을 원활하게 통합
- 계층 구조 연관성:
- 여러 레벨에서 규칙들을 유도
- 다차원 입력테이블에서 계층을 정의하는 상위 (부모) 및 하위 (자식)를 지정

웹 경로 분석

- 클릭스트림 데이터에서 가장 빈번하게 탐색된 경로의 확장 가능하고 효율적인 마이닝.
- 시퀀스 데이터의 모든 유형에서 자주 연속되는 서브시퀀스 마이닝.

링크 분석

- 효과의 네트워크로서 시각화할 수 있는 상호 연결된 개체의 집합으로 데이터 변환.
- 트랜잭션 데이터 내 두 항목의 동시성 사이에서 또는 관계형 데이터 내에서 두 변수 레벨이 어떻게 연결되는지 보여주는 시각적 모델 제공.

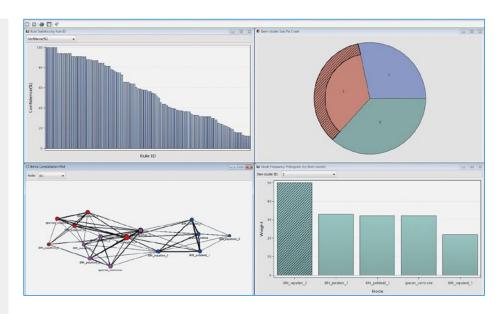


그림 4.

새 패턴을 시각적으로 발견하기 위해 노드 사이의 관계를 평가하는 링크 분석을 사용합니다.

- 중심성 지수와 커뮤니티 정보를 제공하여 연결 그래프 파악
- 가중 신뢰도 통계량를 제공하여 NBO (Next Best Offer)의 정보 제공
- 데이터 축소 및 세그멘테이션에 대한 클러스터 스코어 생성

차원 축소

- 변수 선택:
 - 카이 제곱 또는 R2 선택 기준을 기반으로 타겟과 관련 없는 변수 제거
 - 계층 구조에서 변수 제거
 - 결측값이 많은 변수 제거
 - 레벨 (범주)의 수가 많은 범주형 변수 제거
- 비선형관계를 파악하기 위하여 연속형 입력변수를 구간화 (bin)
- 교호작용 감지
- LARS (Least Angle Regression) 변수 선택:
- AIC, SBC, Mallows C (p), 교차 타당성 및 기타 선택 기준 제공
- 플롯 포함 내용: 파라미터 추정, 계수 경로, 반복 플롯, 스코어 랭킹 등
- LASSO (Least Absolute Shrinkage and Selection Operator) 지원을 위한 일반화

- 범주형 변수와 타겟뿐만 아니라 연속형 변수도 지원
- 스코어 코드 생성
- 주성분 분석:
- 상관 및 공분산 행렬의 고유값과 고유벡터 계산
- 플롯 포함 내용: 주성분 계수, 주성분 행렬, 고유값, 로그 고유값 및 누적 비례 고유치
- 유지할 성분의 수를 인터랙티브 방식으로 선택
- 예측 모델링 기법을 사용하여 선택한 주성분을 마이닝
- 변수 클러스터링:
- 변수를 배반 클러스터 또는 계층 구조 클러스터로 분할
- 고유값 또는 주성분 학습
- 범주형 변수 지원
- 클러스터의 덴드로그램 트리
- 클러스터 및 상관계수가 있는 선택 변수 테이블
- 클러스터 네트워크 및 R-제곱 플롯
- 선택 변수의 인터랙티브 사용자 재정의
- 시계열 마이닝:
- 여러 가지 누적 방법과 변환을 사용하여 시계열 데이터로 만드는 트랜잭션 데이터 축소
- 분석 방법에는 계절별, 추세, 시간, 계절적 분해가 포함됨
- 클러스터링 및 예측 모델링 기법을 사용하여 축소된 시계열을 마이닝

SAS 코드 노드

- 쉬운 데이터부터 복잡한 데이터까지 다양하게 준비하고 변환 작업을 하기 위한 SAS 코드 작성
- 다른 SAS 제품의 프로시저 통합
- 사용자 정의 모델 개발
- SAS Enterprise Miner의 확장 노드 (Extension Node) 생성
- 스코어 코드 로직 확대
- SAS 프로시저 지원
- 배치 코드에서 서로 다른 이름과 위치의 입력 테이블을 사용할 수 있음
- 배치 코드가 라이브러리와 옵션을 정의하는 데
 사용할 수 있는 프로젝트 시작 코드와 통합됨
- 사용하기 쉬운 프로그램 개발 인터페이스:
- 데이터 소스, 변수 등을 참조하기 위한 매크로 변수
- 인터랙티브 방식의 코드 편집기 및 실행
- 분석, 스코어링 및 리포팅 코드를 별도 관리
- SAS 결과 및 SAS 로그
- 그래프 생성

일관된 모델링 기능

- 이익/손실, AIC, SBC, 평균제곱오차, 오분류율, ROC, Gini, KS(Kolmogorov-Siminov) 등의 여러 기준을 사용하여 분석용, 검증용, 평가용 데이터 기반으로 모델 선택
- 모델 개발 프로세스에 사전 확률 통합
- 이진형, 명목형, 순서형 및 구간형 입력변수와 타겟변수 지원
- 스코어 코드와 모든 분할된 데이터 소스에 쉽게 액세스
- 모델 성능 평기를 보다 효과적으로 할 수 있도록 여러 결과를 하나의 창에 표시
- 타겟 이벤트 설정과 사전 확률 및 이익 / 손실 행렬 정의를 위한 결정 노드

회귀

- 선형 회귀 및 로지스틱 회귀
- 단계적 선택법, 전진 선택법, 후진 제거법
- 수식 조건 빌더: 다항식, 일반 교호작용, 효과적인 계층 구조 지원
- 교차 타당성 검사
- 효과적인 계층 구조 규칙
- 최적화 기술 포함 내용: Conjugate Gradient, Double Dogleg, Newton-Raphson with Line Search or Ridging, Quasi-Newton and Trust Region

- 데이터 마이닝(Dmine) 회귀 노드:
- 빠른 전진 단계적 최소 제곱 회귀 분석
- 비선형 관계를 감지하기 위한 변수 구간화 옵션
- 범주형 변수 축소 옵션
- 교호 작용 항 포함
- 테라데이터 13의 인데이터베이스(In-DB) 모델링
- PMML 스코어 코드

의사 결정 트리

- 방법론:
- CHAID, 분류 및 회귀 트리(CART), 배깅과 부스팅, 그래디언트 부스팅, 부트스트랩 포레스트.
- 이익 또는 리프트 목표를 기반으로 트리 선택, 그에 따른 가지치기
- K-fold 교차 타당성 검사
- 분할 기준: 카이 제곱 테스트, F-테스트, 지니 (Gini), 엔트로피 (Enropy) 및 분산 감소
- 다양한 목적을 가진 세분화 전략을 설계하기 위한 타게 변수 전화
- 모델링 및 그룹핑을 위한 입력변수로 활용하기 위해 리프 ID 자동 생성
- 영어 규칙 표시

- 예비 변수 선택 및 모델 해석을 위해 변수 중요도 계산
- 가지와 노드 분할에서 변수 정밀도 값 표시
- 트리 다이어그램의 독특한 통합 트리 지도 표현
- 인터랙티브 방식의 트리 기능:
- 인터랙티브 방식의 트리 성장하기 / 가치치기; 트리 노드 확장 / 축소
- 평가용 데이터를 통합하여 트리 안정성 평가
- 이지 (binary) 또는 다지 (multiway) 분류를 포함한 맞춤형 분할 지점 정의
- 모든 후보 변수 분할
- 분학 보사
- 테이블과 플롯이 동적으로 연결되어 트리 성능을 보다 효과적으로 평가
- 한 페이지 또는 여러 페이지에 걸쳐 손쉬운 인쇄가 가능한 트리 다이어그램
- 인터랙티브 방식의 하위 트리 선택
- 트리 노드에서 사용자 정의방식으로 텍스트와 통계량 표시
- 인터랙티브 방식의 트리 내에서 사용자가 제어하는 샘플 크기
- 빠른 ARBORETUM 프로시저에 기반
- PMML 스코어 코드

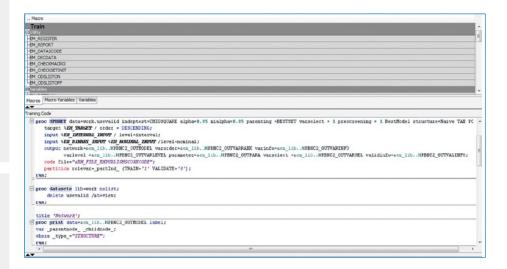


그림 5

맞춤형 SAS 코드를 통합하여 변수 변환, SAS 프로시저 통합, 새로운 노드 개발, 스코어링 로직 확대, 보고서를 맞춤화하는 등의 효과를 얻을 수 있습니다.

뉴럴 네트워크

- 뉴럴 네트워크 노드 (신경망)
- 결합함수 및 활성함수를 가지는 유연한 네트워크 아키텍처
- 10가지 트레이닝 테크닉
- 예비 최적화
- 입력변수를 자동으로 표준화
- 방향 연결 지원
- 오토뉴럴 노드 (자동신경망)
- 최적 구성을 위한 검색 위한 자동화된 다층 퍼셉트론 구축
- 아키텍처의 네 가지 유형에서 선택된 유형 및 활성함수
- PMML 스코어 코드
- DM 뉴럴 노드 (데이터 마이닝 신경망)
- 차원 축소 및 함수 선택이 가능한 모델 개발
- 빠른 트레이닝; 선형 및 비선형 추정

부분 최소 제곱 노드

- 잠재적으로 상관된 많은 변수들 중에서 팩터를 추출하는데 특히 유용함
- 주성분 회귀 및 축소된 순위 회귀 수행
- 팩터 선택에 대해 사용자 선택 또는 자동 선택
- 5가지 교차 타당성 전략에서 선택
- 변수 선택 지원

규칙 추론

- 반복적인 예측 모델링 기법
- 드문 이벤트를 모델링하는 데 특히 유용함

2단계 모델링

- 클래스 및 간격 대상 모두에 대한 순차 모델링과 동시 모델링
- 각 단계에 대한 의사 결정 트리, 회귀 또는 뉴럴 모델 선택
- 클래스 예측이 간격 예측에 적용되는 방법 제어
- 고객 가치를 정확하게 추정

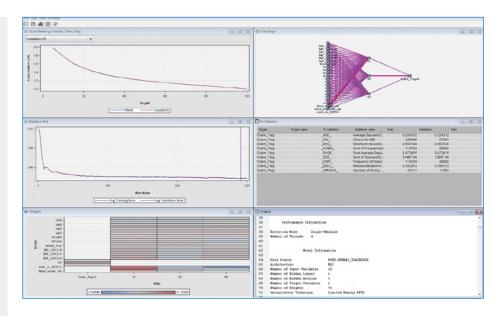


그림 6.

뉴럴 네트워크 노드를 사용하여 매우 복잡한 비선형 관계를 적합시킵니다

메모리 기반 추론

- 관찰 내용을 분류하거나 예측하는 k-근접 이웃 (Nearest Neighbor) 기법
- 특허 받은 차원 축소 트리 및 스캔

앙상블 모델

- 예측 모델을 결합하여 잠재적으로 보다 효율적인 모델 개발
- 지원되는 방법: 평균, 투표 (Voting) 및 최대

오픈소스 통합 노드

- SAS Enterprise Miner 내에서 R 언어로 코드 작성
- SAS Enterprise Miner 데이터와 메타데이터는 SAS Enterprise Miner로 반환된 R 결과와 함께 R 코드 사용이 가능함
- R 언어를 이용하여 지도 및 비지도 모델을 트레이닝하고 스코어링함. 이 노드는 데이터 변환과 탐색이 가능함
- 지원된 모델에 대해 모델 비교 및 SAS 스코어 코드 생성이 가능함

증분 응답 / 순 리프트 모델

- 순 처리군 및 대조군 모델
- 이진 타겟 및 구간 타겟
- 단계적 선택법
- 고정 또는 가변적인 수익 계산
- 순 정보 지수 (Information Value) 기반의 변수 선택
- 사용자는 처리 변수의 처리 수준을 지정할 수 있음
- 사용자는 고정비 이외에 추가적으로 비용 변수를 지정할 수 있음
- 변수 선택을 위한 PNIV (Penalized Net Information Value) 제공
- 증분 판매 모델에 사용할 수 있는 별도의 모델 선택 옵션

- 교차 상관:
- 시계열 데이터에 대한 자기 상관 및 교차 상관 분석 제공
- 시계열 상관 노드는 자기 상관이나 교차 상관 분석이 수행되면 시간 영역 통계량를 출력함
- 계절적 분해

생존 분석

- Discrete Time to Event 데이터에 대해 회귀분석과 가법 로지스틱 회귀분석 제공
- 시간 효과에 대한 이벤트 확률은 큐빅 스플라인을 사용하여 모델링

- 주요 효과뿐만 아니라 단계적 변수 선택 프로시저에서 큐빅 스플라인 기저 함수도 입력할 수 있음
- 중도절단 (Censoring)된 데이터에 대한 처리와 분석 방법을 지정하기 위한 시구간 사용자 정의
- 샘플링 옵션으로 데이터를 자동 확장
- 비시간 가변성 공변량 지원
- 홀드아웃 데이터의 생존 함수 계산으로 타당성 검사
- 경쟁 위험 또는 부위험 (subhazards) 생성
- 평균 잔여 수명 계산을 위한 스코어 코드 생성
- 사용자 지정 데이터포맷으로 (표준, 변경-시간, 전체 확장 등) 시간에 따른 공변량을 설정
- 좌측 절단 및 절단 날짜를 지정할 수 있음

시계열 데이터 마이닝

- 시계열 데이터 준비:
- 트랜잭션 및 시퀀스 데이터 집계, 변환 및 요약
- 유사성 분석, 클러스터링 및 예측 모델링 지원을 위한 시계열 자동 변환
- TimeID 변수가 있거나 없는 상태로 데이터 처리
- 유사성 분석:
- 신제품 예측, 패턴 인식 및 단기 수명 주기 예측에 유용함
- 타겟변수와 시계열 입력변수 사이, 또는 다른 시계열 입력변수들 사이의 유사성 계산
- 시계열의 모든 조합에 대한 유사성 행렬
- 유사성 행렬과 덴드로그램 결과를 사용한 계층적 클러스터링
- 클러스터를 평가하기 위한 Constrellation 플롯
- 지수평활:
- 하나 이상의 평활 파라미터를 이용하여 가중치 감소 제어
- 가장 적합한 평활 방법 자동 선택 (단순, 더블, 선형, 감소 추세, 계절별 또는 윈터스 방법)
- 차원 축소:
- 5가지 시계열 차원 축소 기법 지원: 이산 웨이블릿 변환 (DWT), 이산 푸리에 변환 (DFT), 특이값 분해 (SVD), 평균 선분 근사법, 합계 선분 근사법

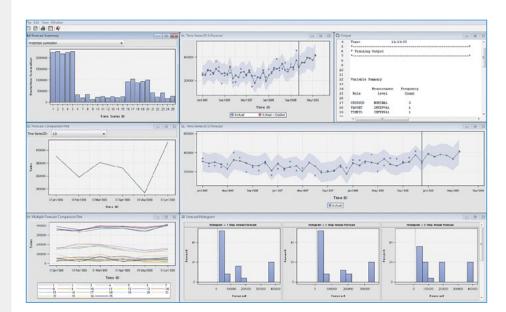


그림 7. 전통적인 계절적 분해를 사용하여 시계열 데이터를 분석합니다.

시작 및 종료 그룹 노드로 그룹 프로세싱

- 프로세스 흐름도의 세그먼트에 대한 프로세싱 반복
- 세분화 모델링, 배깅 및 부스팅, 다중 타겟 및 교차 타당성 검사에 대한 활용이 가능함

SAS® Rapid Predictive Modeler

- 다양한 비즈니스 문제를 위해 예측 모델을 자동으로 생성.
- SAS Enterprise Miner에서 모델을 열고, 확대하고 수정할 수 있음
- 손쉬운 사용과 검토를 위해 변수 중요성 리프트 차트, ROC 차트 및 모델 스코어 카드 등의 간결한 보고서 생성
- 스코어링된 데이터 세트를 저장하는 옵션과 함께 분석용 데이터를 스코어링할 수 있는 기능

하이퍼포먼스 데이터 마이닝 프로시저

- 멀티스레드로 처리되는 프로시저는 동시에 실행되며, 기존 대칭형 멀티 프로세싱 (SMP) 서버의 사용 가능한 모든 코어를 활용하여 프로세싱 속도 개선:
 - HPBIN (high-performance binning)
 - HPBNET (high-performance Bayesian networks)
 - HPCLUS (high-performance clustering)
 - HPCORR (high-performance correlation)
- HPDECIDE (high-performance decide)
- HPDMDB (high-performance data mining database)
- HPDS2 (high-performance DS2)
- HPFOREST (high-performance random forests)
- HPIMPUTE (high-performance imputation)
- HPNEURAL (high-performance neural networks)
- HPREDUCE (high-performance variable reduction)
- HPSAMPLE (high-performance sampling)
- HPSUMMARY (high-perfor—mance data summarization)
- HPSVM (high-performance Support Vector Machine)
- HP4SCORE (high-performance 4Score)



그림 8.

투자 수익률을 최대화하기 위해 마케팅 처리의 증가 효과를 모델링합니다.

- 하이퍼포먼스 지원 SAS Enterprise Miner 노드:
- HP Bayesian Network
- HP Cluster
- HP Data Partition
- HP Explore
- HP Forest
- HP GLM
- HP Impute
- HP Neural Network
- HP Principal Components
- HP Regression
- HP SVM
- HP Tree
- HP Transform
- HP Variable Selection

모델 가져오기 노드

- 다른 다이어그램과 프로젝트에 재사용할 수 있도록 SAS Enterprise Miner 모델 등록
- 외부 모델 가져오기 및 평가하기

모델 평가

- 모델 비교 노드는 단일 프레임워크에서 모든 홀드아웃 데이터에 대해 여러 모델을 비교
- 사용자 정의 모델 기준에 의거하여 최고의 모델을 자동으로 선택
- 사용자 기각의 지원
- 다양한 적합 및 진단 통계량
- 리프트 차트; ROC 곡선
- 구간 타겟변수는 랭킹 및 분포를 스코어
- 손익 차트와 의사 결정 선택; 오분류(분류) 행렬
- 클래스 확률 스코어 분포 플롯; 스코어 랭킹 행렬 플롯
- 이진형 타겟에 대한 확률 컷오프 포인트를 결정하기 위한 컷오프 노드
- 기본 선택에 대한 사용자 재정의
- 최대 KS 통계
- 최소 오분류 비용
- 최대 누적 프로파일
- 최대 True Positive Rate
- 이전 트레이닝의 최대 이벤트 정밀도
- 이벤트 정밀도 동일 재현율

리포터 노드

- SAS ODS (Output Delivery System)를 사용하여 프로세스 흐름의 PDF 또는 RTF 생성.
- 분석 프로세스를 문서화하고 결과를 쉽게 공유할 수 있도록 지원
- 문서가 저장되어 SAS Enterprise Miner 결과 패키지에 포함됨
- 프로세스 흐름도의 이미지 포함
- 사용자 정의 메모 항목

데이터 저장 노드

- 이전에 정의된 SAS 라이브러리 또는 지정된 파일 경로 중 하나에 노드의 분석용, 검증용, 평가용, 스코어용 또는 트랜잭션 데이터를 저장
- JMP[®], Excel 2010, CSV 및 탭으로 구분된 파일로 내보내기. 사용자의 입력 없이 SAS Enterprise Miner 배치 프로그램에 노드를 배포할 수 있도록 기본 옵션이 설계되어 있음
- 분석용, 검증용, 평가용, 스코어용 또는 트랜잭션 데이터를 내보내기 위해서 SAS Enterprise Miner 프로세스 흐름도의 모든 노드에 연결할 수 있음

스코어링

- SAS Enterprise Miner GUI에서 인터랙티브 방식의 스코어링을 위한 스코어 노드
- 최적화된 스코어 코드가 기본값으로 생성되며 사용되지 않는 변수를 제거함
- SAS, C, Java 및 PMML(버전 4.2)로 스코어 코드 자동 생성
- SAS, C 및 Java 스코어링 코드는 모델링, 클러스터링, 변환 및 결측값 대치 코드를 포함
- Aster Data, DB2, Greenplum, Hadoop, IBM Netezza, Oracle, SAP HANA, SAS Scalable Performance Data Server 및 Teradata 데이터베이스에서 SAS Scoring Accelerator를 사용하여 SAS Enterprise Miner 모델을 직접 스코어링

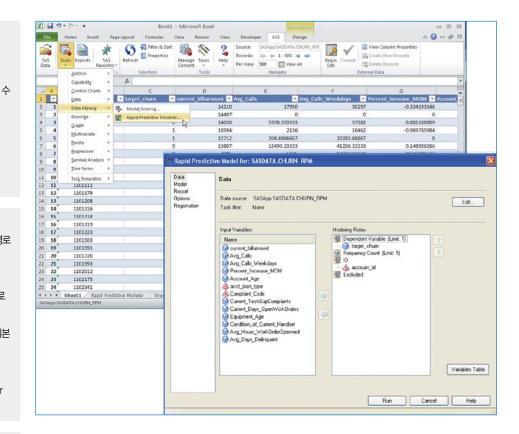


그림 9.

SAS Enterprise Guide나 Microsoft Office(Microsoft Excel 전용)의 SAS 추가 기능으로 작동하는 SAS Rapid Predictive Modeler를 사용하여 다양한 비즈니스 문제에 대한 예측 모델을 자동으로 생성합니다.

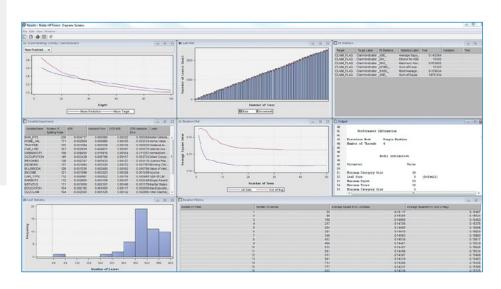


그림 10.

여러 개의 의사 결정 트리를 앙상블로 구성하여 랜덤 포레스트 모델을 구축합니다. 여러 번의 반복을 통해 샘플 선택에 대한 의존도를 줄이면서 분할을 위한 변수를 무작위로 선택합니다. 예측을 하기 위해 OOB (Out-Of-Bag) 샘플을 사용합니다.

모델 등록 및 관리

- 세분화, 분류 또는 예측 모델을 SAS 메타데이터 서버에 등록 입력 변수, 출력 변수, 타겟 변수, 마이닝 기능, 분석 데이터 및 SAS 스코어 코드 등이 메타데이터에 등록됨
- 모델 등록 노드는 등록 단계를 통합하고 SAS Enterprise Miner 배치 코드에서 실행할 수 있는 등록 메커니즘을 제공함
- SAS Metadata Server의 모델 등록으로 다음과 같은 작업을 수행할 수 있음:
- 모델의 전체 라이프 사이클 관리를 위해 SAS Model Manager와 통합
- 스코어링 모델을 위해 SAS Enterprise Guide와 SAS Data Integration Studio 통합

오픈소스 통합 노드

- SAS Enterprise Miner 내에서 R 언어로 코드 작성
- SAS Enterprise Miner 데이터와 메타데이터는 SAS Enterprise Miner로 반환된 R 결과와 함께 R 코드 사용이 가능함
- R 언어를 이용하여 지도 및 비지도 모델을 트레이닝 하고 스코어링함 이 노드는 데이터 변환과 탐색이 가능함
- 지원된 모델에 대해 모델 비교 및 SAS 스코어 코드 생성이 가능함

SAS Enterprise Miner의 자세한 내용은 sas.com/korea/enterpriseminer를 방문해주세요.

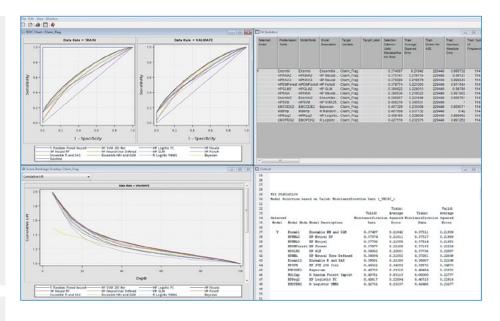


그림 11.

모델 비교 노드를 사용하여 해석하기 쉬운 하나의 프레임워크에서 여러 모델을 함께 평가합니다.

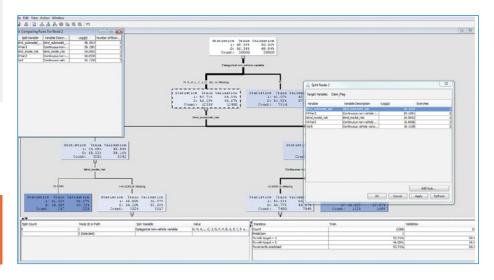


그림 12.

의사 결정 트리를 인터랙티브 방식으로 또는 배치에서 개발합니다. 수많은 평가 플롯이 전반적인 트리 안정성 측정에 도움이 됩니다.

