

# SAS AI-Enhanced Analytics

AI 기반의 의사결정 지원을 위한 SAS의 전략



# CONTENTS

|                                        |   |
|----------------------------------------|---|
| Summary                                | 1 |
| AI가 주도하는 분석 인사이트                       | 2 |
| AI 없이는 인사이트도 없다                        |   |
| 비즈니스를 움직이는 AI                          | 3 |
| AI 기반의 데이터 분석이 게임의 법칙을 바꾼다             |   |
| AI를 만드는 AI, AutoML                     | 4 |
| AutoML은 어떻게 변화해 왔을까?                   |   |
| 모두를 위한 AI 시대                           | 5 |
| 데이터 과학자는 더 이상 필요 없을까?                  |   |
| 기업 환경에 적합한 AutoML 선택 전략                | 6 |
| 서로 다른 세 가지 접근 방안                       |   |
| SAS, 데이터 사이언스 및 머신 러닝 플랫폼 리더           | 7 |
| 모두가 만족하는 최고의 분석과 운영 환경                 |   |
| SAS가 바라보는 확장된 AutoML                   | 8 |
| SAS는 어떻게 AI-Enhanced Analytics를 실현하는가? |   |
| 지속 가능한 창조와 혁신                          | 9 |
| SAS AI-Enhanced Analytics로 앞서가십시오      |   |



## SAS AI-Enhanced Analytics 주요 기능 소개

|                                                            |    |
|------------------------------------------------------------|----|
| 데이터 (Data)                                                 | 11 |
| 1. 오토 프로파일링(Auto Profiling)                                |    |
| 2. 데이터 품질 개선 추천(Transformation Suggestion)                 |    |
| 분석 (Discovery)                                             | 14 |
| 3. 자동 비주얼 인사이트(Automated Explanation)                      |    |
| 4. 시각화 추천(Visualization Suggestion)                        |    |
| 5. 오토 피쳐 엔지니어링(Auto Feature Engineering)                   |    |
| 6. 오토 튜닝(Auto Tuning)                                      |    |
| 7. 오토 파이프라인(Pipeline Automation)                           |    |
| 8. 모델 해석력(Model Interpretability)                          |    |
| 9. 문장 기반의 모델 설명(Explain Models in Plain Language)          |    |
| 적용 (Deployment)                                            | 24 |
| 10. 손쉬운 배포(Deployment)                                     |    |
| 11. 지능적 의사결정(Intelligent Decisioning)                      |    |
| 12. 모니터링 및 셀프러닝(Monitoring and Self-Learning)              |    |
| 개방성 (Open)                                                 | 28 |
| 13. SAS AI 플랫폼의 오픈 소스 개방성(SAS and Open Source Integration) |    |



# Summary

기업 조직 전체에서 데이터 기반의 의사결정을 실현하기 위해서는 AI를 특정 영역이 아닌 분석 라이프 사이클 전반으로 확대하고, 이 라이프 사이클은 의사결정 프로세스와 연결돼야 합니다. 개방된 플랫폼에서 분석 시간을 줄이고, 그 결과를 운영시스템에 빠르게 적용하여 더 큰 비즈니스 가치를 실현 하도록 돕는 것. 이것이 바로, SAS AI-Enhanced Analytics입니다.

AI가 처음 우리에게 소개됐을 때 막연한 불안감을 느꼈습니다. AI가 인간을 대신한 주도적인 의사결정으로 우리 미래를 결정하게 될 거라는 상상 때문이었습니다. 그러나 실제로는 AI가 사람을 대체하기보다는 업무에 생산성을 높이고, 보다 창의적이며 전략적인 업무에 주력하게 하며, 비즈니스를 입체적으로 지원하는 핵심 기술로서 그 역할을 해내고 있습니다.

## AI 도입 시 고려사항

AI가 이처럼 든든한 조력자로 자리하게 된 가장 큰 이유는 의사결정 프로세스의 혁신에 있습니다. 데이터 기반의 의사결정을 위해 방대한 소스 데이터를 전처리하고, 모델링 하고, 운영 환경에 적용하는 이 복잡한 프로세스를 AI와 머신 러닝 기술을 통해 자동화 함으로써 업무에 효율성을 높이고, 비즈니스 인사이트를 빠르게 얻어 최적의 의사 결정을 할 수 있게 된 것입니다. 그러나 이 혁신을 온전히 누리기 위해서는 기업 환경에 적합한 솔루션과 플랫폼을 선택해야 합니다. 어떤 문제를 해결하고자 하는지, 어느 수준의 통찰력을 얻고자 하는지, 전문 인력은 보유하고 있는지 등 여러 요소들을 면밀히 살펴 선택해야 합니다.

## SAS의 AI 전략

SAS는 기업 조직 전체에서 데이터 기반의 의사결정을 실현하기 위해서는 '데이터(Data)-분석(Discovery)-적용(Deployment)'으로 연결되는 분석 라이프 사이클 전반으로 AI가 확대돼야 한다고 생각합니다. 이 라이프 사이클은 의사결정 프로세스와도 연결돼야 합니다. SAS가 데이터, 분석, 적용을 아우르는 통합 분석 플랫폼 위에 AI 기반의 자동화된 분석 기능을 제공하는 이유입니다.

분석 라이프 사이클 전반에 적용된 AI 기능을 통해 전문 분석가는 보다 높은 수준의 분석을 수행 하고, 일반 사용자는 드래그 앤 드롭이나 몇 번의 클릭만으로 다양한 분석 결과를 확인하고, 결과를 반영하기 위한 의사결정 프로세스와도 자동으로 연결할 수 있습니다. 또한 Python, R등의 오픈소스 사용자들도 개방성 있는 SAS 플랫폼의 파워풀한 장점을 활용할 수 있습니다. 분석 시간을 줄이고, 운영시스템에 더욱 빠르게 적용하여, 더 큰 비즈니스 가치를 실현하도록 돕는 것. 이것이 SAS의 AI 전략이며, SAS는 이를 'AI-Enhanced Analytics'로 정의합니다. 본 백서에서는 AI와 머신 러닝의 트렌드, AI를 쉽게 적용하도록 돕는 AutoML의 개념, 전세계 AI 소프트웨어 플랫폼 시장을 이끌고 있는 SAS의 AI-Enhanced Analytics 전략과, SAS 솔루션의 주요 AI 기능도 함께 살펴봅니다.

# AI가 주도하는 분석 인사이트

## AI 없이는 인사이트도 없다

시장조사기관인 가트너(Gartner)는 세계 최대 규모의 IT 컨퍼런스를 매년 개최하고 있습니다. 지난 2019년 10월, 미국 올랜도에서 개최된 컨퍼런스에서는 <비즈니스를 변화시킬 10가지 데이터 및 분석 트렌드>를 발표했는데, 대부분 AI(Artificial Intelligence, 인공지능) 기술과 관련된 것입니다(그림1). 발표자로 나선 가트너의 애널리스트 리타 살람(Rita Sallam)은 이들 10가지 트렌드를 관통하는 키워드로 '인텔리전스'를 언급하며, "이는 머신 러닝과 AI 기술이 업무에 본격적으로 투입되어 인사이트(insight)를 자동으로 향상시켜주는 것을 의미한다"고 말했습니다.

66

현재 오픈 소스 플랫폼이 AI와 머신 러닝을 지배하고 있으며, 알고리즘 및 개발 환경의 주요 혁신 소스인 것만은 확실합니다. 그러나 상용 공급업체들이 오픈 소스 에코시스템을 사용할 수 있는 커넥터를 제공하고, 오픈 소스 플랫폼에는 없는 기능인 프로젝트 및 모델 관리, 재사용, 투명성, 통합 등 AI와 머신 러닝 확장에 필요한 엔터프라이즈 기능을 제공하고 있기 때문에 2022년까지 AI와 머신 러닝 기술을 활용하는 새로운 최종 사용자 솔루션의 75%가 오픈 소스 플랫폼이 아닌, 상용 공급업체의 솔루션으로 구축될 것입니다.

- 리타 살람(Rita Sallam) -

99

## 비즈니스를 변화시킬 10가지 데이터 및 분석 트렌드



출처: 가트너

그림1

# 비즈니스를 움직이는 AI

AI 기반의 데이터 분석이 게임의 법칙을 바꾼다

산업 전반에 AI 기술이 도입되면서 혁신의 속도는 더욱 빨라지고 있습니다. 의료 분야에서는 평균수명을 연장하기 위해 딥러닝 기술로 유전자 데이터와 가계도를 분석하여 난치병을 치료하는 연구를 진행하고 있습니다. 금융권에서는 자동화된 알고리즘을 통해 고객에게 자산을 추천 및 관리해주는 로보 어드바이저(robo-advisor) 시장이 빠른 속도로 확대되고, 자동차 산업에서는 도로 정체와 운송 비용, 시간을 획기적으로 줄여주는 자율주행자동차가 상용화에 속도를 내고 있습니다. 유통 분야에서는 음성 인식, 얼굴 인식, 머신 러닝, 딥러닝 등 AI 기술을 총망라한 무인 매장이 본격화되고 있으며, AI 기술을 사용해 모든 교차로의 동영상과 자동차 위치, 위성항법장치(GPS) 데이터 등 도시 전역의 정보를 수집해 교통 문제를 해결한 도시도 등장했습니다.

이처럼 비즈니스 전반에 혁신을 가속화하고 있는 AI의 가치는 '데이터 분석'에서 비롯됩니다. 데이터 분석은 일반적으로 문제 정의 → 데이터 수집 → 데이터 탐색 및 전처리 → 모델링 → 해석 및 시각화의 프로세스를 거칩니다. 어느 하나 중요하지 않은 과정은 없습니다. 문제는 데이터 양이 기하급수적으로 증가하면서 데이터 탐색과 전처리, 모델링에 소요되는 시간이 증가한다는 것입니다. 이 시간이 길어질수록 의사결정은 늦어질 수밖에 없습니다. 다행히도 발전된 AI 기술이 분석 결과물을 만들기 위한 복잡한 분석 과정에 도움을 주기 시작했습니다. 비즈니스에 활용할 수 있는 AI 분석 결과물을 AI의 도움으로 만드는 시대가 된 것입니다. 그 대표적인 예가 AutoML(자동 머신 러닝)입니다.

# AI를 만드는 AI, AutoML

## AutoML은 어떻게 변화해 왔을까?

머신 러닝은 과거부터 누적된 데이터를 기반으로 수학적 알고리즘을 만들어 컴퓨터가 스스로 생각하도록 만드는 AI 기술입니다. 데이터 과학자는 이런 기술이 자동으로 실행되는 소프트웨어를 설계하여 예측 모델을 자동화합니다. 이 과정에서 예측 가능성이 가장 높은 규칙을 찾기 위해 모델 평가 지표를 수립하고, 알고리즘 선택, 모델링, 평가, 보완에 이르는 작업을 반복적으로 수행합니다. 이와 같은 불편을 해결하기 위해 등장한 것이 바로, 'AutoML(자동 머신 러닝)'. 즉, 머신 러닝을 자동화하는 AI기술입니다.

AutoML의 시작은 학습에 필요한 학습률, 배제율 등 초매개변수(hyperparameter)를 최적화하여 더 나은 모델을 학습시키는 '하이퍼파라미터 최적화(Hyperparameter Optimization)'였습니다. 이후 학습을 통해 유의미한 변수(Feature)를 추출해서 사용하는 '피쳐 학습 (Feature Learning)'으로 발전했고, 모델링 전체 프로세스를 사람이 하나 하나 직접 설계하던 것에서 학습을 통해 최적의 아키텍처를 설계하는 '아키텍처 서치(Architecture Search)'로 발전해 왔습니다(그림2).

### AutoML 트렌드

#### Hyperparameter Optimization



학습률(learning rate), 배치 크기(mini-batch size) 등 학습에 큰 영향을 주는 hyperparameter들을 학습을 통해 추정하는 것을 의미합니다.

#### Feature Learning



Extraction 혹은 Feature Engineering을 의미하는 것으로, 학습 모델에 입력을 그대로 사용하지 않고, 학습을 통하여 유의미한 feature(특징)를 추출해서 입력으로 사용하는 방법입니다.

#### Architecture Search



모델링 전체 프로세스를 사람이 직접 하나하나 설계하는 대신에 학습을 통해 최적의 아키텍처를 설계하는 방법을 의미합니다.

그림2



# 모두를 위한 AI시대

데이터 과학자는 더 이상 필요 없을까?



AutoML 솔루션은 최근 몇 년간 가장 빠르게 진화하는 기술 중 하나로, 예측 모델 개발에 많은 시간을 소요했던 코딩, 알고리즘 선택, 튜닝 작업을 자동화합니다. 전문 지식 없이도 기업 구성원 누구나 몇 번의 마우스 클릭만으로도 손쉽게 머신 러닝을 활용하는 데 초점을 둔 솔루션들도 등장하고 있습니다. 전문 인력이 부족한 기업에게는 좋은 선택입니다. 그렇다고 해서 데이터 과학자나 분석가, 개발자의 역할이 사라지지는 않습니다. 더 짧은 시간에 더 많은 모델을 구축함으로써 모델 품질과 정확도를 개선하고, 보다 생산적인 일에 몰두하며 전문성을 강화할 수 있습니다.

머신 러닝 오픈 소스도 다양해지고 있습니다. 그러나 대부분의 오픈 소스가 키트 형태로 제공되므로 오픈 소스 파트들을 조합하여 자체 엔터프라이즈 플랫폼을 구성해야 합니다. 서로 다른 원천기술을 가진 많은 전문가들을 고용하기 힘든 일반 기업이 이를 활용하기는 쉽지 않습니다. 모델 관리, 투명성, 데이터 추적 등 대규모 조직에서 필요로 하는 엔터프라이즈급 기능도 부족합니다.

이 때문에 가트너에서는 SAS와 같은 기업에서 제공하는 '상용(Commercial) AI와 머신 러닝'에 주목할 필요가 있다고 강조합니다. SAS는 전문가를 위한 엔터프라이즈급 기능과 일반 사용자를 위한 편의 기능을 함께 제공합니다. 또한 전문가와 일반 사용자 간의 협업을 지원하는 다양한 기능을 통해 보다 폭 넓고 깊이 있는 인사이트를 얻을 수 있습니다.

# 기업환경에적합한 AutoML 선택 전략

## 서로 다른 세가지 접근 방안

01

### 오픈 소스를 활용한 직접 개발

다양한 오픈 소스 기술을 접목하여 필요한 플랫폼을 직접 개발하는 방식입니다. 분석 프로세스 단계들을 지원하기 위해 여러 원천 기술들을 직접 연계 및 통합해야 하며, 운영에 필요한 기술에 대응하기 위해 장기적으로 많은 리소스를 투자해야 합니다. 전문 개발자를 많이 보유할 수 없는 일반 기업은 도입하기 어렵습니다.

02

### 특정 분석 목적의 자동화 플랫폼

전문 지식 없이도 모델을 개발할 수 있는 자동 모델 개발 기능을 기반으로 특정한 비즈니스 문제를 빠르게 해결하려는 목적으로 개발됐습니다. 제공하는 분석의 종류가 정해져 있고, 결과에 대한 커스터마이징에 한계가 있어 해결 가능한 비즈니스 범위에 제약이 있을 수 있습니다.

03

### 자동화를 포함하는 범용적 상용 플랫폼

폭넓은 분석 기능을 제공하는 플랫폼 위에 AI 기반의 자동화된 분석 기능을 제공합니다. 자동화 플랫폼 개념뿐만 아니라 폭 넓고 깊이 있는 기능을 제공하므로, 전문 분석가에서 일반 사용자에게 이르기까지 넓은 사용자와 비즈니스 문제 해결에 대응할 수 있습니다. 일부 상용 플랫폼은 오픈 소스 기술과 오픈 소스 사용자를 지원하는 아키텍처를 제시하고 있습니다.

# SAS, 데이터 사이언스 및 머신 러닝 플랫폼 리더

모두가 만족하는 최고의 분석과 운영 환경

시장조사분석기관인 가트너는 <2020 매직 쿼드란트 : 데이터 사이언스 및 머신러닝 플랫폼(Gartner 2020 Magic Quadrant for Data Science and Machine-Learning Platforms) > 보고서에서 SAS를 리더로 선정(그림3)하고, 다음과 같은 점을 강점으로 언급했습니다.

## 모든 사용자의 니즈를 충족시키는 편리한 기능

'SAS 비주얼 데이터 마이닝 앤드 머신러닝(SAS VDMML)'은 드래그 앤 드롭(Drag-and-Drop) 인터페이스 방식을 원하는 시티즌 데이터 사이언티스트(Citizen Data Scientists)는 물론, 코딩 작업을 선호하는 데이터 사이언티스트(Data Scientists)를 동시에 지원한다. SAS VDMML은 데이터 품질 개선 및 전처리에 대한 자동 추천 기능을 제공하며, 변수의 분포 및 다양한 측정치를 기반으로 적절한 데이터 시각화도 추천해 준다.

## 최고의 모델 운영 및 관리 플랫폼

SAS는 최고의 모델 운영 및 관리 플랫폼을 제공한다. 구체적으로 모델 성능 모니터링은 물론 모델 성능 저하가 임계 값을 초과할 경우 자동으로 모델을 재학습하는 기능을 제공한다. 이 외에도 템플릿 및 버전 제어 기능, 통합된 모델 리포지토리를 통한 거버넌스 제공, SAS 및 오픈 소스 모델에 대한 영향도 분석 기능 등을 지원한다.

2020 가트너 매직쿼드란트 '데이터 사이언스 및 러닝 플랫폼'



그림3

# SAS가 바라보는 확장된 AutoML

## SAS는 어떻게 AI-Enhanced Analytics를 실현하는가?

일반적으로 분석 라이프 사이클은 '데이터(Data)-분석(Discovery)-적용(Deployment)'으로 구성됩니다. 시장에서의 AutoML이 분석(Discovery)에 집중된 반면, SAS는 특정 영역뿐만 아니라, 분석 라이프 사이클 전반에 걸쳐 AI의 도움이 필요하다고 생각합니다. SAS가 데이터, 분석, 적용을 아우르는 통합 분석 플랫폼 위에 AI 기반의 자동화된 분석 기능을 제공하는 이유입니다(그림4).

### AI-Enhanced Analytics

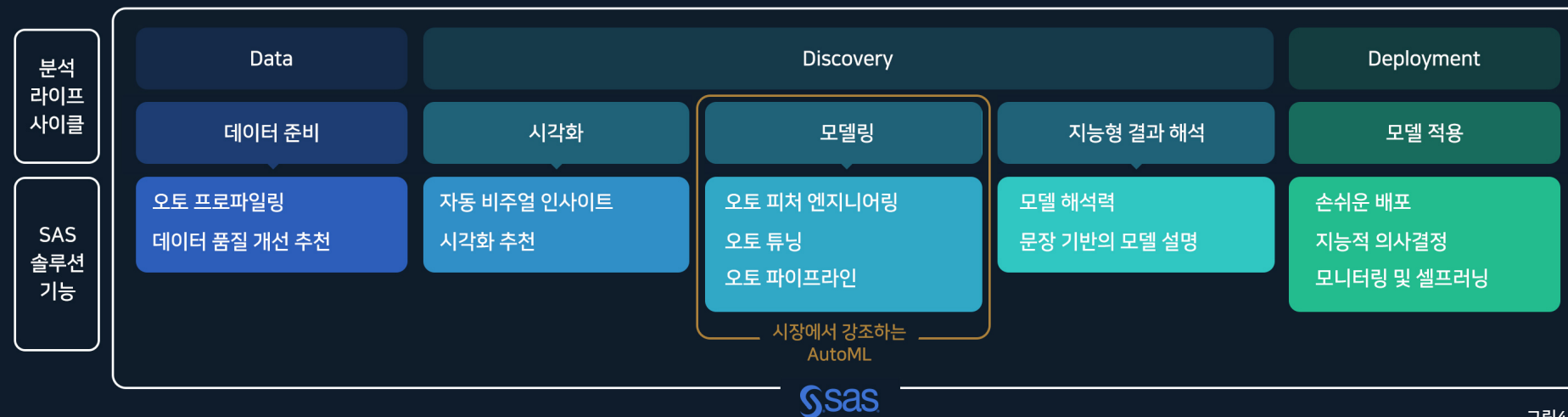


그림4

기업은 SAS의 AI-Enhanced Analytics를 통해 다음과 같은 효과를 얻을 수 있습니다.

- 신속한 분석 결과 생성 및 비즈니스 적용(빠른 모델 프로토타이핑 및 개발)
- 조직의 분석 역량 강화(분석 장벽 해소, 시민 데이터 과학자 확대)
- 모델의 품질, 생산성 향상(정확성 향상, 오류 방지)
- 분석의 적용 범위 확대(더 많은 업무와 시나리오에 분석 적용)

이러한 AI-Enhanced Analytics 환경은 비주얼 인터페이스와 오픈 소스를 포함한 다양한 코딩 언어에 대한 선택권을 제공합니다. 효율적인 모델 개발과 모델 적용에 어려움을 겪어 왔던 Python, R, Java와 같은 오픈 소스 분석가들도 자신에게 익숙한 언어를 기반으로 SAS의 다양하고 강력한 AI-Enhanced Analytics 기능을 활용할 수 있습니다.

# 지속 가능한 창조와 혁신

## SAS AI-Enhanced Analytics로 앞서가십시오

데이터 준비에서 모델 배포까지 시간이 많이 소요될수록 의사결정은 그만큼 늦어지고, 비즈니스 기회 또한 잃게 됩니다. 모델을 배포하는 것으로 업무가 끝나서도 안 됩니다. 수시로 변화하는 데이터를 실시간으로 적용하지 않으면 분석의 가치가 떨어지기 때문입니다. SAS 솔루션은 기업이 더 많은 데이터를 준비하여 더 빠르게 모델을 구축하고, 신속히 업무에 적용할 수 있게 함으로써 비즈니스 기회 손실을 최소화 합니다.

한발 더 나아가 분석 라이프 사이클을 의사결정 프로세스에 통합하고, 지속적인 모델 모니터링과 관리, 비즈니스 영향 점검 등 일련의 과정을 통해 비즈니스 가치 실현에 속도를 내게 합니다(그림5). 개방된 플랫폼에서 분석 시간을 줄이고, 그 결과를 운영시스템에 빠르게 적용하여, 더 큰 비즈니스 가치를 얻도록 돕는 것. 이것이 SAS AI-Enhanced Analytics 입니다. SAS AI-Enhanced Analytics로 지속 가능한 비즈니스 혁신을 이뤄보십시오.

### 분석 라이프 사이클과 비즈니스 가치 실현

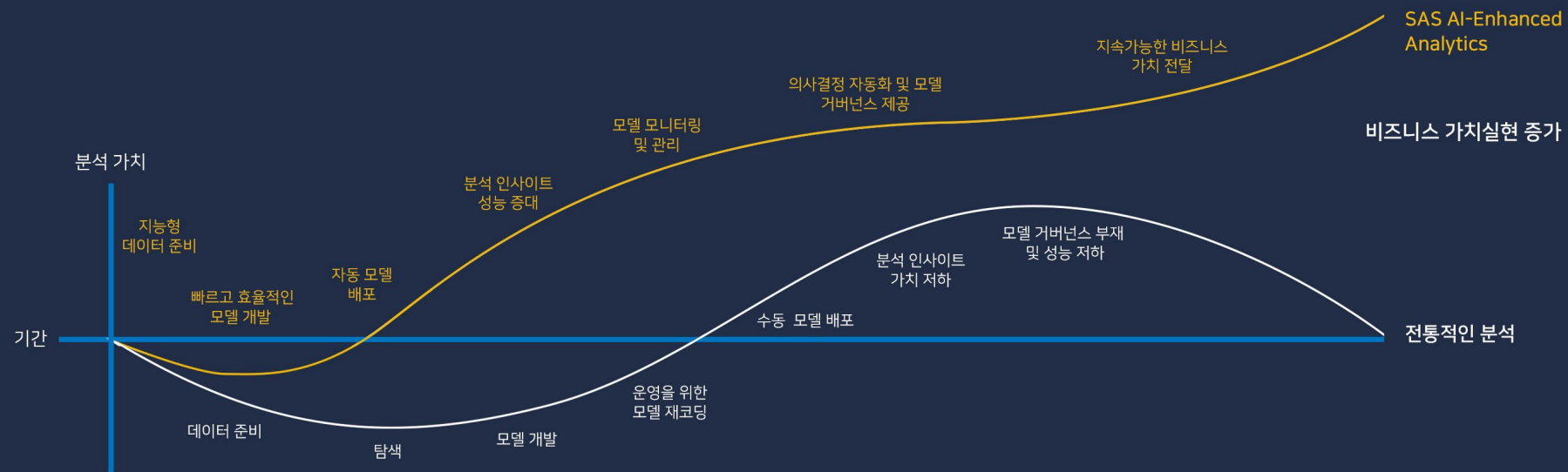


그림5

# SAS AI-Enhanced Analytics 주요 기능 소개

## 데이터 (Data)

### 데이터 준비(Data Preparation)

#### 1. 오토 프로파일링(Auto Profiling)

빠르고 명확한 속성 파악으로 데이터 이해도 증대

#### 2. 데이터 품질 개선 추천(Transformation Suggestion)

품질 최적화를 위한 데이터 개선 추천

#### 6. 오토 튜닝(Auto Tuning)

하이퍼파라미터 자동 튜닝으로 최적의 모델 생성

#### 7. 오토 파이프라인(Pipeline Automation)

최적의 모델링 프로세스 자동 디자인

#### 11. 지능적 의사결정(Intelligent Decisioning)

비즈니스 룰과 머신 러닝 접목으로, 지능적 의사결정 지원

#### 12. 모니터링 및 셀프러닝(Monitoring and Self-Learning)

최적의 모델링 프로세스 자동 디자인

## 분석 (Discovery)

### 시각화(Visualization)

#### 3. 자동 비주얼 인사이트(Automated Explanation)

신속한 데이터 탐색 및 예측 인사이트 제공

#### 4. 시각화 추천(Visualization Suggestion)

분석 인사이트 향상을 위한 자동 차트 생성

### 모델링(Modeling)

#### 5. 오토 피처 엔지니어링(Auto Feature Engineering)

최적의 피처 자동 추출로 모델 정확도 향상

### 지능형 결과 해석(Model Interpretability & Explainability)

#### 8. 모델 해석력(Model Interpretability)

블랙박스를 설명할 수 있는 모델 해석력

#### 9. 문장 기반의 모델 설명(Explain Models in Plain Language)

누구나 쉽게 이해할 수 있는 자연어 기반의 설명

## 적용 (Deployment)

### 모델 적용(Model Deployment)

#### 10. 손쉬운 배포(Deployment)

추가 변환 없이, 다양한 운영 환경에 모델 적용

## 개방성 (Open)

#### 13. SAS AI 플랫폼의 오픈 소스 개방성

(SAS and Open Source Integration)

AI-Enhanced Analytics의 인터페이스와 코딩 언어  
선택권 제공

## ○ 데이터 (Data)

SAS AI-Enhanced Analytics는 데이터 구조, 크기, 속도에 관계없이 모든 데이터에 빠르게 접근하여, AI 기반의 통합 데이터 품질 개선 및 데이터 프로파일링 도구를 이용해 데이터 준비를 간소화하고, 데이터 품질을 높입니다. 이를 통해 사용자는 IT 도움을 최소화하면서 분석 데이터 전처리 작업의 생산성과 속도를 획기적으로 개선합니다.

### 데이터 준비(Data Preparation)

효율적인 분석을 위해서는 데이터 연동 및 가공, 품질 개선 등을 거쳐 분석에 적합한 데이터를 생성해야 합니다. 데이터 준비 과정에서는 자동 프로파일링 기능으로 데이터의 패턴 및 이슈를 쉽게 파악합니다. 또한, 데이터 품질 개선 추천 기능을 통해 자동화된 방법으로 데이터 품질 이슈를 빠르게 파악하고 품질을 개선합니다.

## 1. 오토 프로파일링(Auto Profiling)

**빠르고 명확한 속성 파악으로 데이터 이해도 증대**

오토 프로파일링은 마우스 클릭만으로 분석하고자 하는 데이터의 전반적인 특성은 물론, 각 변수의 데이터 값에 대한 세부적인 패턴 정보를 알 수 있는 기능입니다(그림6). 자동으로 생성되는 데이터 프로파일 보고서를 통해 선택한 테이블의 데이터에서 이상 또는 불일치를 쉽게 식별할 수 있습니다.

## 오토 프로파일링 화면

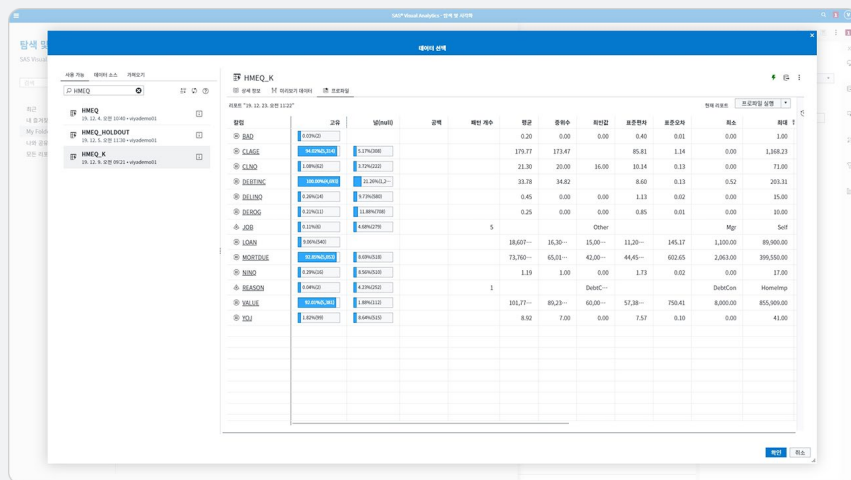


그림6

예를 들어 그림 7과 같은 프로파일링 결과를 통해 일부 열에 null 값이 있고, 한 열의 값이 얼마나 고유한지에 대한 정보를 파악할 수 있습니다. 또한, 특정 변수를 선택했을 때 통계적 수치, 데이터 패턴, 빈도 분포 등 상세 정보를 제공하므로 데이터 패턴을 보다 깊이 이해할 수 있습니다.

오토 프로파일링 결과로 제공되는 변수값의 상세 속성 화면

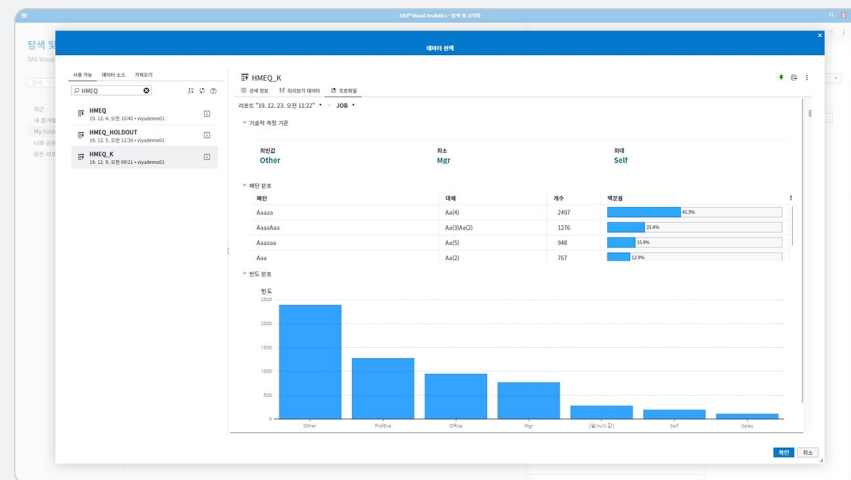


그림7

## 2. 데이터 품질 개선 추천(Transformation Suggestion)

### 품질 최적화를 위한 데이터 개선 추천

데이터 품질 개선 추천은 분석을 위한 데이터 전처리 기능 중 하나입니다. 다양한 소스에서 취합된 데이터는 분석을 위해 취합되고 가공되어야 하는데, 데이터 전처리는 분석에 필요한 데이터 접근 및 가공, 파생 변수 생성, 데이터 품질 개선 등 다양한 작업을 포함합니다.

데이터 품질 개선을 위한 작업은 데이터 전처리 과정 중에서 가장 중요한 반면, 반복적이고 많은 시간을 소모합니다. 지능적인 데이터 변환 추천과 자동 변환 기능은 사용자의 관여없이 자동으로 데이터 품질 이슈들을 찾아내고, 사용자가 개선을 원하는 문제를 해결해 줌으로써 데이터 전처리 프로세스의 생산성을 높이고 분석 결과의 질을 향상 시킬 수 있습니다(그림8).

데이터 품질 개선 추천 화면

The screenshot displays the SAS AI-Enhanced Analytics interface. On the left, a sidebar lists various data quality issues such as '고유 ID', '선택된 및 스케일링', 'BAD', '선택된 및 스케일링', 'LOAN', '선택된 및 스케일링', '결측값 처리', '선택된 및 스케일링', 'VALUE', '완-결 인코딩(One-hot encoding)', '완-결 인코딩(One-hot encoding)', '선택된 및 스케일링', '결측값 처리', '선택된 및 스케일링', and '결측값 처리'. The main area shows a table of suggestions for the 'HMEQ\_K (session)' dataset. The table has columns for 'BAD', 'LOAN', 'VALUE', 'REA...', 'JOB', 'YOJ', and 'DE'. The right-hand panel shows a summary of the suggestions and a code editor with the following code:

```

4 dataPreprocess.transform status=rc /
5
6 table(name=dp_inputTable, caslib=dp_inputCaslib);
7 casout(name=dp_outputTable, caslib=dp_outputCaslib, promote=0, replace=1);
8 requestPackages={functions:method="STANDARDIZE", arguments={location="mean", scale="std"}},
9 outVariableGlobalPrefix="
10 copyVar=copyvars_columns;
11
12

```

그림8

## ● 분석(Discovery)

분석(Discovery)은 데이터를 분석하고 모델을 생성하는 과정으로 탐색과 발견, 확률과 예측, 최적의 의사 결정 등을 포함합니다. SAS AI-Enhanced Analytics는 분석 과정에서 다양한 자동화, 추천 및 AI 기반의 지능형 모델 생성 기능을 제공합니다. 이를 통해 기업은 빠르게 조직의 분석 역량을 높이고, 분석의 품질과 생산성을 높여 다양하고 광범위한 비즈니스 문제에 민첩하게 대응할 수 있게 됩니다.

## 시각화(Visualization)

시각화에서는 시각적 기법을 통한 직관적인 데이터 탐색, 분석 결과의 시각적 표현 및 결과 공유 등을 제공합니다. 원클릭만으로 데이터를 의미 있게 분석해주는 자동 설명 기능과 시각화 그래프 추천 기능 등으로 분석의 효율성을 크게 향상시켜 줍니다.

### 3. 자동 비주얼 인사이트(Automated Explanation)

#### 품질 최적화를 위한 데이터 개선 추천

자동 비주얼 인사이트는 알고 싶은 현상에 대한 변수(수익, 제품 불량/등급 등)를 선택하고 클릭하면 그림9와 같이 데이터를 자동으로 분석하여 그 결과를 시각적으로 보여줍니다. 복잡한 데이터 탐색을 통해 수시간에 걸쳐 알아내야 했던 예측 관점의 분석 인사이트를 단 수초 만에 얻을 수 있습니다.

#### 변수 선택에 따라 제공되는 다양한 인사이트

- 주요 영향 인자 : 관심 이벤트에 영향을 주는 주요 인자는 무엇인가?
- 영향 인자와 관계 : 영향을 주는 인자와 관심 이벤트 간의 관계는?
- 관심 현상 예측 조건 : 영향 인자를 중심으로 관심 이벤트가 발생할 확률이 높은/낮은 조건은?
- 새로운 관점 추가 : 관심 이벤트 예측에 새로운 관점을 더한다면?

자동 비주얼 인사이트 화면

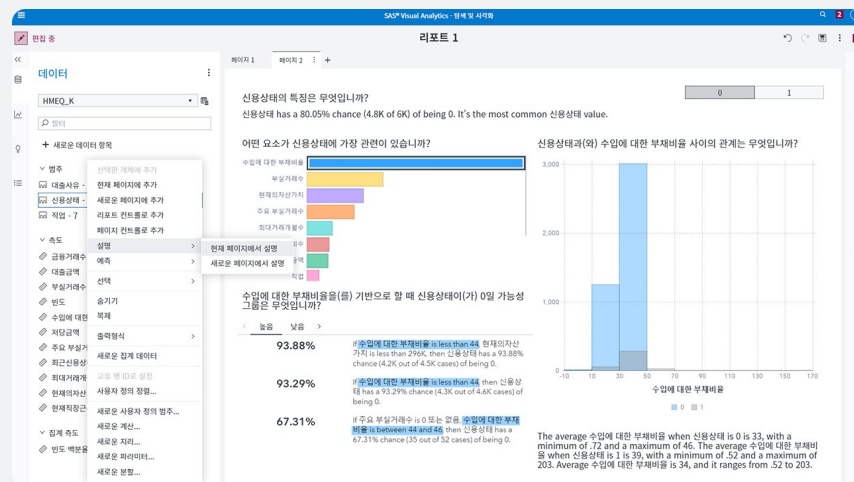


그림9

## 4. 시각화 추천(Visualization Suggestion)

### 분석 인사이트 향상을 위한 자동 차트 생성

시각화 추천은 유의미한 시각화 차트를 자동으로 생성하여 추천하는 기능입니다. 사용자가 데이터 항목과 알맞은 차트를 선택하기 위해 고민하는 대신, 시각화 추천 버튼을 누르면 분석 툴이 지능적으로 데이터를 분석하여 관계성이 높은 변수들을 조합하여 다양한 시각화 그래프를 추천합니다(그림10). 사용자는 추천 그래프 중 마음에 드는 결과를 드래그 앤 드롭하여 리포트에 포함시킬 수 있습니다.

### 시각화 차트는 어떻게 추천될까?

시각화 추천은 소스 데이터로부터 유의미한 차트를 추천하기 위해 다양한 개념들을 활용합니다. 많은 변수 중 상관성이 높은 변수들끼리 조합되고, 각 변수가 가진 고유한 값의 개수 등이 고려됩니다. 또한, 데이터의 패턴을 표현하는데 가장 적당한 차트의 종류가 자동으로 선택됩니다.

시각화 추천 화면

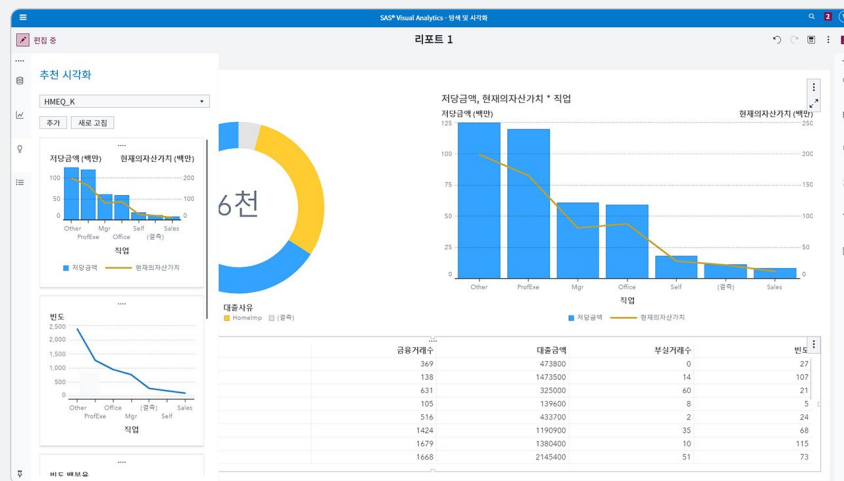


그림10



## 모델링(Modeling)

모델링은 피처(feature, 변수)를 생성하고 최적의 모델을 생성하는 과정입니다. 최근 시장에서 AutoML이 모델링의 진화된 방법으로 주목받고 있습니다. AutoML은 AI의 도움을 받아 자동으로 모델을 만드는 것을 의미합니다.

## 5. 오토 피처 엔지니어링(Auto Feature Engineering)

### 최적의 피처 자동 추출로 모델 정확도 향상

오토 피처 엔지니어링\*은 사람이 직접 피처 추출 방법을 정하는 대신, 학습을 통해 최적의 피처 추출 방법을 찾는 것을 의미합니다. 기존에는 각 모델마다 최적의 피처 추출 방식을 찾는 데 많은 시간이 소요됐지만, 이제는 오토 피처 엔지니어링의 GUI를 통해 모델의 피처를 자동으로 정리, 변환, 선택할 수 있습니다. 아울러 데이터의 품질 문제를 식별하여 올바른 유형의 데이터로 변환할 수 있습니다.

\* 피처 엔지니어링(Feature Engineering)이란 주어진 데이터로부터 모델링 성능을 높이는 새로운 변수를 만드는 과정입니다.



## 5. 오토 피쳐 엔지니어링(Auto Feature Engineering)

### 1) 피쳐 머신 노드(Feature Machine Node)

피쳐 머신 노드는 데이터 프로파일링을 통해 개별 변수의 특성을 반영한 변수 변환을 수행하여, 데이터의 질과 모델 정확도를 향상시키는 새로운 피쳐를 자동으로 생성합니다(그림11).

- 피쳐 머신은 데이터 품질을 개선하고 모델의 정확도를 향상할 수 있도록 정제 및 변환하여 새로운 피쳐를 생성합니다.

피쳐 머신 노드와 피쳐 노드 옵션 화면

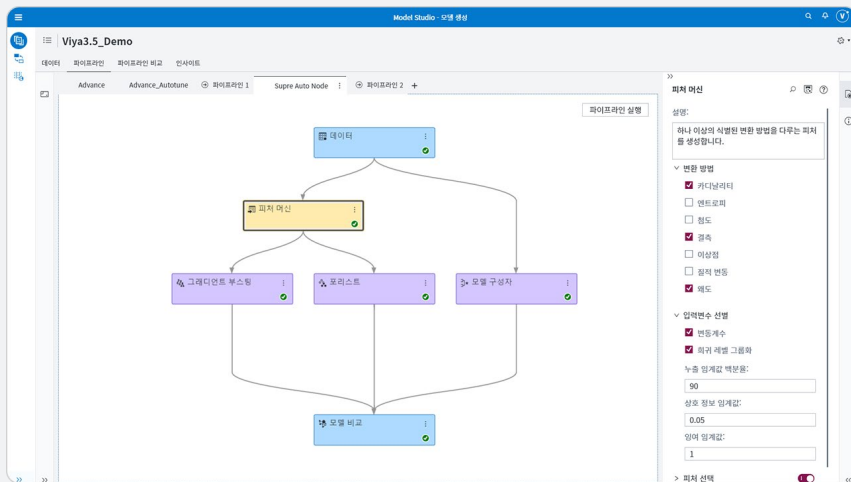


그림11

- 피쳐 머신은 높은 카디널리티(Cardinality), 높은 첨도, 높은 왜도, 낮은 엔트로피, 낮은 질적 분산도(IQV), 결측값, 이상점 등을 고려하여 피쳐를 생성합니다.
- 피쳐 선택을 사용하여 입력 변수당 최상위 순위의 피쳐만 내보낼 수 있습니다.
- 자동 생성된 피쳐 중에서 판단 기준(Ranking Criterion)을 생성하여 분석에 반영할 최종 피쳐를 자동으로 선택합니다(그림12).

피쳐 머신 결과 화면

생성된 피쳐

| ORIS | 피쳐                         | 설명                                                                             | 레벨       | 입력변수    | 입력 레이블          | 순위화 기준  | 피쳐 순위 |
|------|----------------------------|--------------------------------------------------------------------------------|----------|---------|-----------------|---------|-------|
| 1    | mkoks_nkoks_dtree_5_CLAGE  | CLAGE: Not high (outlier, kurtosis, skewness) - five bin decision tree binning | NOMINAL  | CLAGE   | 가장 오래된 거래의 개월 수 | 0.01743 | 1     |
| 2    | mkoks_nkoks_pow_r0_5_CLAGE | CLAGE: Not high (outlier, kurtosis, skewness) - power(-0.5) * impute(median)   | INTERVAL | CLAGE   | 가장 오래된 거래의 개월 수 | 0.01669 | 2     |
| 3    | mkoks_nkoks_pow_r1_1_CLAGE | CLAGE: Not high (outlier, kurtosis, skewness) - power(-1) * impute(median)     | INTERVAL | CLAGE   | 가장 오래된 거래의 개월 수 | 0.01416 | 3     |
| 4    | mkoks_nkoks_log_CLAGE      | CLAGE: Not high (outlier, kurtosis, skewness) - log * impute(median)           | INTERVAL | CLAGE   | 가장 오래된 거래의 개월 수 | 0.01386 | 4     |
| 5    | mkoks_nkoks_dtree_10_CLAGE | CLAGE: Not high (outlier, kurtosis, skewness) - ten bin decision tree binning  | NOMINAL  | CLAGE   | 가장 오래된 거래의 개월 수 | 0.01316 | 5     |
| 6    | mkoks_nkoks_pow_r2_2_CLAGE | CLAGE: Not high (outlier, kurtosis, skewness) - power(2) * impute(median)      | INTERVAL | CLAGE   | 가장 오래된 거래의 개월 수 | 0.01219 | 6     |
| 7    | mkoks_nkoks_pow_r0_5_CLAGE | CLAGE: Not high (outlier, kurtosis, skewness) - power(0.5) * impute(median)    | INTERVAL | CLAGE   | 가장 오래된 거래의 개월 수 | 0.01214 | 7     |
| 8    | cpy_inr_med_imp_CLAGE      | CLAGE: Low missing rate - median imputation                                    | INTERVAL | CLAGE   | 가장 오래된 거래의 개월 수 | 0.01136 | 8     |
| 9    | mkoks_nkoks_pow_r1_1_CLAGE | CLAGE: Not high (outlier, kurtosis, skewness) - power(1) * impute(median)      | INTERVAL | CLAGE   | 가장 오래된 거래의 개월 수 | 0.01138 | 8     |
| 10   | mkoks_nkoks_pow_r2_2_CLAGE | CLAGE: Not high (outlier, kurtosis, skewness) - power(2) * impute(median)      | INTERVAL | CLAGE   | 가장 오래된 거래의 개월 수 | 0.00906 | 10    |
| 11   | cpy_inr_med_imp_CLINO      | CLINO: Low missing rate - median imputation                                    | INTERVAL | CLINO   | 금융거래의 수         | 0.01008 | 1     |
| 12   | mis_jhd_DELTINC            | DELTINC: Significant missing - missing indicator                               | NOMINAL  | DELTINC | 수입에 대한 부채비율     | 0.25079 | 1     |
| 13   | cpy_inr_med_imp_DELTINC    | DELTINC: Low missing rate - median imputation                                  | INTERVAL | DELTINC | 수입에 대한 부채비율     | 0.09591 | 2     |
| 14   | hs_btc_r2_DELINO           | DELINO: High skewness - Box-Cox(lambda=2) * impute(median)                     | INTERVAL | DELINO  | 부실거래의 수         | 0.09005 | 1     |
| 15   | hs_btc_r2_DELINO           | DELINO: High skewness - Box-Cox(lambda=2) * impute(median)                     | INTERVAL | DELINO  | 부실거래의 수         | 0.08996 | 2     |
| 16   | hs_btc_r1_DELINO           | DELINO: High skewness - Box-Cox(lambda=1) * impute(median)                     | INTERVAL | DELINO  | 부실거래의 수         | 0.08923 | 3     |
| 17   | cpy_inr_med_imp_DELINO     | DELINO: Low missing rate - median imputation                                   | INTERVAL | DELINO  | 부실거래의 수         | 0.08921 | 4     |
| 18   | hs_btc_r0_DELINO           | DELINO: High skewness - Box-Cox(lambda=0) * impute(median)                     | INTERVAL | DELINO  | 부실거래의 수         | 0.08921 | 4     |
| 19   | hs_btc_r1_DELINO           | DELINO: High skewness - Box-Cox(lambda=1) * impute(median)                     | INTERVAL | DELINO  | 부실거래의 수         | 0.08921 | 4     |
| 20   | hs_dtree_disc10_DELINO     | DELINO: High skewness - ten bin decision tree binning                          | NOMINAL  | DELINO  | 부실거래의 수         | 0.07149 | 7     |
| 21   | hs_dtree_disc15_DELINO     | DELINO: High skewness - five bin decision tree binning                         | NOMINAL  | DELINO  | 부실거래의 수         | 0.07149 | 7     |
| 22   | mis_jhd_DELINO             | DELINO: Significant missing - missing indicator                                | NOMINAL  | DELINO  | 부실거래의 수         | 0.01026 | 9     |
| 23   | hs_btc_r1_DEFOG            | DEFOG: High skewness - Box-Cox(lambda=1) * impute(median)                      | INTERVAL | DEFOG   | 주요 부실거래의 수      | 0.06539 | 1     |
| 24   | cpy_inr_med_imp_DEFOG      | DEFOG: Low missing rate - median imputation                                    | INTERVAL | DEFOG   | 주요 부실거래의 수      | 0.06536 | 2     |
| 25   | hs_btc_r0_DEFOG            | DEFOG: High skewness - Box-Cox(lambda=0) * impute(median)                      | INTERVAL | DEFOG   | 주요 부실거래의 수      | 0.06536 | 2     |
| 26   | hs_btc_r1_DEFOG            | DEFOG: High skewness - Box-Cox(lambda=1) * impute(median)                      | INTERVAL | DEFOG   | 주요 부실거래의 수      | 0.06536 | 2     |
| 27   | hs_btc_r2_DEFOG            | DEFOG: High skewness - Box-Cox(lambda=2) * impute(median)                      | INTERVAL | DEFOG   | 주요 부실거래의 수      | 0.06536 | 2     |
| 28   | hs_btc_r2_DEFOG            | DEFOG: High skewness - Box-Cox(lambda=2) * impute(median)                      | INTERVAL | DEFOG   | 주요 부실거래의 수      | 0.06424 | 6     |

그림12

## 5. 오토 피쳐 엔지니어링(Auto Feature Engineering)

### 2) 피쳐 추출 노드(Feature Extraction Node)

피쳐 추출 노드는 저자원 공간에 입력 데이터의 중심 속성을 캡슐화 하는 새로운 피쳐(변수)를 생성하여 변수 타겟에 대한 설명력을 높이는 다양한 피쳐를 생성합니다(그림13).

- 주성분 분석(PCA), 특이값 분해(SVD), 로버스트 PCA(RPCA) 및 Autoencoder의 방법론을 사용합니다.
- 각 피쳐 추출 방법별로 다양한 옵션을 선택 및 조정할 수 있습니다.
- 피쳐 추출 노드가 Interval 입력 변수의 수에 따라 자동으로 방법을 선택하도록 할 수 있습니다. Interval 입력 변수의 수가 500보다 작거나 같으면 PCA가 자동으로 선택되며, 그 외에 SVD 방법을 자동으로 적용합니다. PCA, SVD, RPCA 방법은 Interval 입력 변수를 사용하여 선형 주성분을 생성하고, Autoencoder 방법은 Interval 입력 변수와 Class 입력 변수를 모두 사용하여 비선형 피쳐를 생성합니다(그림14).

피쳐 추출 결과 화면

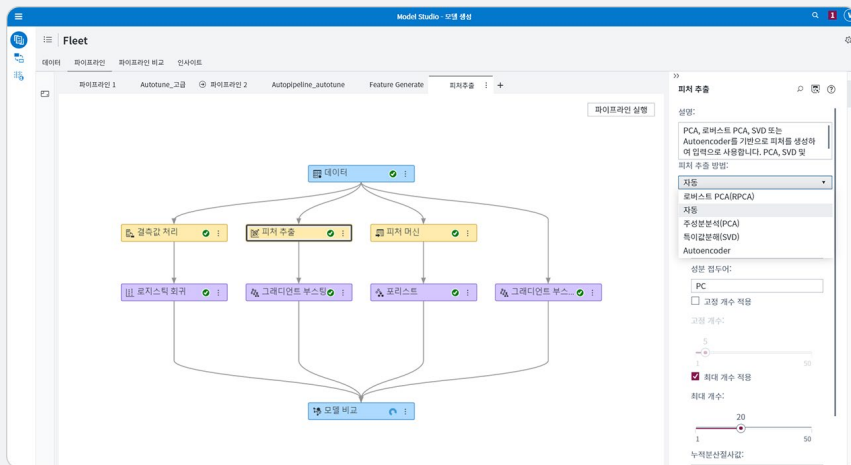


그림13

피쳐 추출 노드와 피쳐 추출 노드 옵션 화면

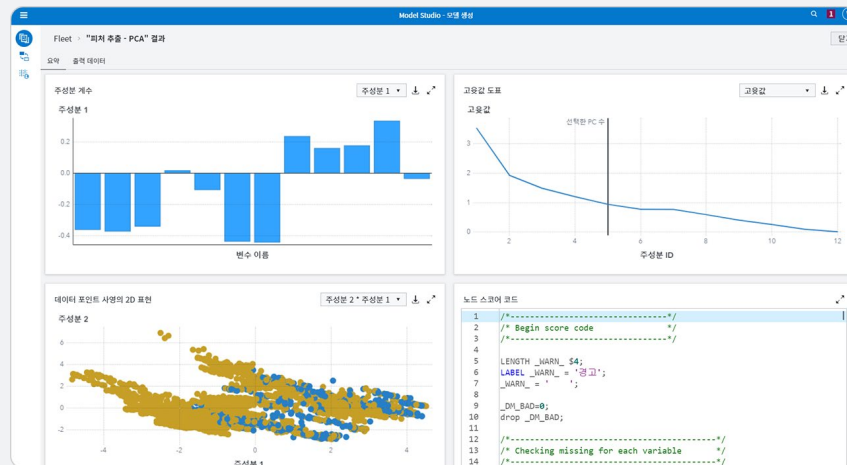


그림14

## 6. 오토 튜닝(Auto Tuning)

### 하이퍼파라미터 자동 튜닝으로 최적의 모델 생성

오토 튜닝은 학습률(learning rate), 배치 크기(mini-batch size) 등 학습에 큰 영향을 주는 하이퍼파라미터(hyperparameter)들을 Trial & Error 방식이 아닌, 학습을 통해 자동으로 추정하는 것을 의미합니다.

SAS에서 제공하는 오토 튜닝은 그리드, Random 서치뿐만 아니라, 라틴 하이퍼큐브, 베이지안, Genetic 알고리즘 등의 방법을 통해 최소의 시간으로 하이퍼파라미터에 대한 최적의 조합을 찾아줍니다(그림15, 16).

오토 튜닝이 포함된 고급 템플릿 및 오토 튜닝 옵션 화면

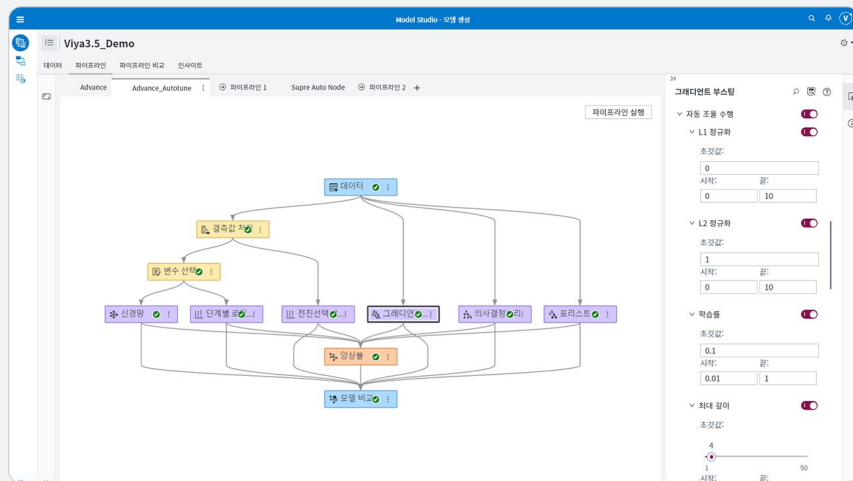


그림15

오토 튜닝 결과 화면

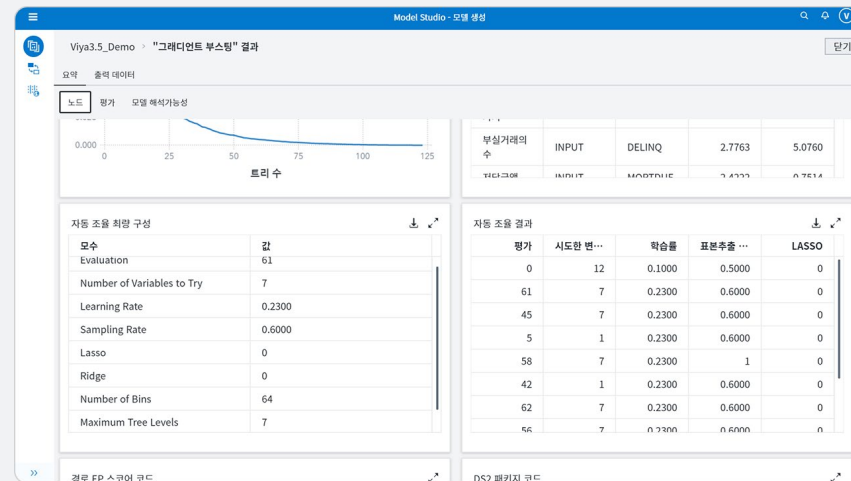


그림16

## 7. 오토 파이프라인(Pipeline Automation)

### 최적의 모델링 프로세스 자동 디자인

오토 파이프라인은 AI의 도움을 받아 모델 개발 프로세스를 자동으로 제공하는 기능을 의미합니다. AI가 데이터의 속성을 자동으로 분석하여 최적의 구조로 분석하기 위한 파이프라인을 생성합니다. 일반적인 AutoML이 생성하는 분석 결과물은 수정할 수 없는 블랙박스(black-box)의 형태인 반면, SAS의 결과물은 사용자 수준에 따라 편집 및 확장이 가능한 화이트박스(white-box)의 형태로 제공됩니다.

파이프라인 자동 생성 기능과 자동화 시간을 선택하면, AI 알고리즘이 데이터를 프로파일링한 후 Input Table의 특성을 반영한 피쳐 엔지니어링 및 알고리즘을 제시하여 최적의 모델을 생성하도록 지원합니다(그림17).

SAS는 AutoML 방법론 이외에도 타깃의 유형 및 분석 수준, 자동 조율 포함 여부 등에 따라 미리 정의된 템플릿을 제공하여 분석가의 분석 생산성을 향상시킵니다. 타깃 변수의 유형(Class/Interval)에 따라 사전에 정의되어 있는 분석 템플릿을 제시하므로, 분석가의 목적에 따라 기본/중급/고급 등 난이도별로 선택할 수 있습니다. 또한 자동 튜닝 여부도 선택할 수 있으므로 분석 경험이 부족한 사람도 쉽고 빠르게 우수한 성능의 모델을 구축할 수 있습니다.

파이프라인 자동 생성 선택, 실행, 결과 화면

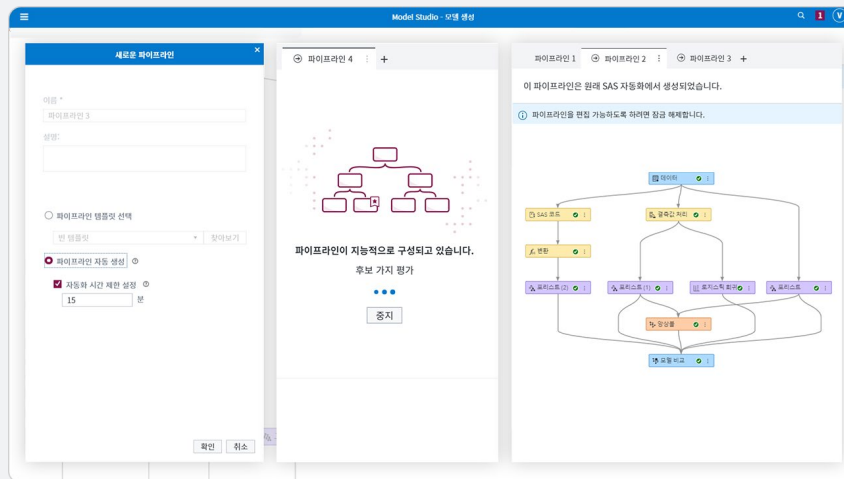


그림17

## 지능형 결과 해석(Model Interpretability & Explainability)

머신 러닝과 AI 기술이 발전하면서 복잡한 DNN(Deep Neural Network, 심층 신경망)의 형태나 다양한 모델들을 조합하는 등 모델이 매우 복잡해졌습니다. 이처럼 복잡한 머신 러닝 모델을 '블랙박스 모델'이라고 부릅니다. 이 모델을 쉽게 이해하고 해석할 수 있도록 돕는 기능이 중요합니다.

## 8. 모델 해석력(Model Interpretability)

### 블랙박스를 설명할 수 있는 모델 해석력

모델 해석력이란 최종 사용자에게 특정한 의사결정이나 프로세스에 대해 설명할 수 있는 능력을 의미합니다.

SAS는 Explainable ML(설명 가능한 머신 러닝) 트렌드에 맞춰 부분 의존성 도표(PDP), 조건부 기대치(ICE), LIME, SHAP Value 등 다양한 모델 해석 기능을 제공합니다(그림 18).

#### ■ 부분 의존성 도표(PDP)

다른 변수들의 영향력은 평균화(average out)하고, 입력 변수와 모델 예측값 사이의 관계를 표현(선형함수, 계단함수, 곡선함수 등의 관계) 합니다.

#### ■ 개별 조건부 기대치(ICE)

부분 의존성 플롯을 몇개의 하위 그룹 또는 관측값으로 나누어 표현(개별 관측의 수준 까지 상세하게 탐색하고, 하위 그룹과 모델 입력 사이의 상호 작용 식별)합니다.

#### ■ LIME

적용된 알고리즘과 상관없이 설명하려는 데이터 근처에 있는 데이터 샘플에 대해 설명 가능한 모델을 구축할 수 있습니다.

#### ■ SHAP Value

각 개별 고객의 예측에 대해 설명합니다.

모델 해석력 화면

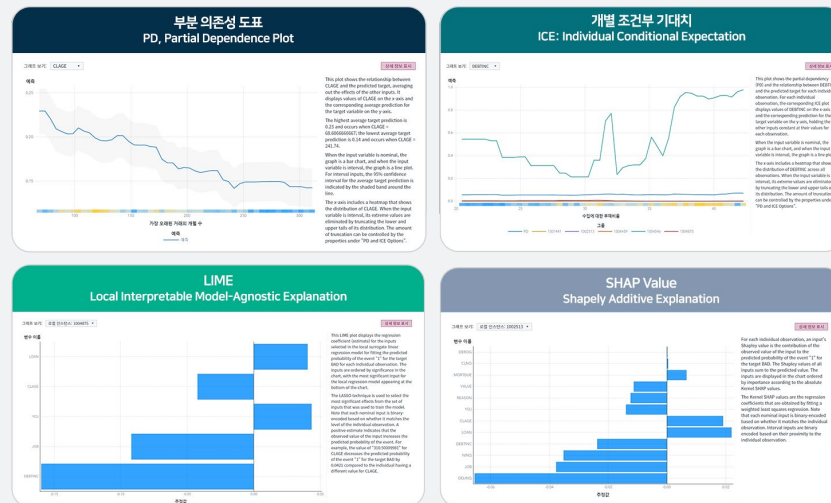


그림 18

## 9. 문장 기반의 모델 설명(Explain Models in Plain Language)

### 누구나 쉽게 이해할 수 있는 자연어 기반의 설명

문장 기반의 모델 설명은 각 알고리즘의 결과, 파이프라인 비교, 인사이트 화면 등에서 모델의 주요 결과 및 통계량의 특징 등을 자연어 기반 문장의 형태로 설명하는 기능입니다. 이를 통해 일반 사용자도 복잡한 표나 차트를 해석하는 대신, 서술되어 있는 문장을 읽어가면서 모델을 손쉽게 파악할 수 있습니다.

특히, 인사이트 화면에서는 프로젝트에 대한 요약 정보, 사용된 변수, 변수 중요도, 모델에 대한 평가 등을 확인할 수 있고, '프로젝트 노트'의 주석을 이용해 모델에 대한 상세 설명 등을 추가할 수 있습니다(그림19). 이를 통해 모델링 프로젝트에 대한 전반적인 개요를 관리하고 분석 결과에 대한 종합적 이해도를 높일 수 있습니다.

지도학습 노드 결과 및 인사이트 화면

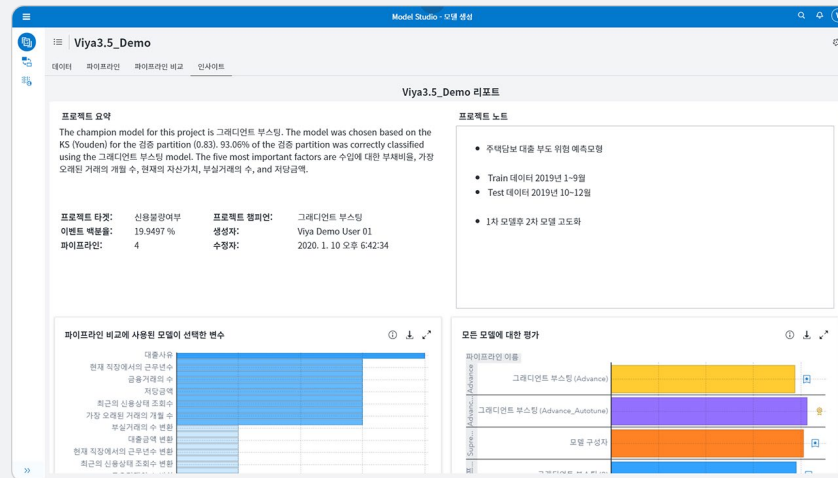


그림19

## ● 적용 (Deployment)

개발된 모델은 운영 환경에 적시에 적용돼야 합니다. SAS의 강력한 모델 관리 기능을 이용해 모델 생성, 최적의 모델 선정, 배포까지 쉽고 빠르게 진행할 수 있습니다. 특히 한번의 모델 구축을 통해 재코딩 작업 없이 어디에나 모델을 배포하고, 통합 코드 베이스로 시간과 업무를 획기적으로 줄일 수 있습니다.

모델은 운영 환경에 적용된 후 시간이 지나면서 그 효용을 상실하거나 잘못된 결과를 산출하기 시작합니다. 따라서 모델의 시작 단계에서부터 사용, 폐기에 이르는 전 과정을 모니터링하고, 필요한 경우 재학습하여 모델을 개선해야 합니다.

### 모델 적용(Model Deployment)

SAS AI-Enhanced Analytics는 모델 운영에 필요한 확장성과 거버넌스를 제공합니다. SAS의 시각적인 웹 기반의 인터페이스를 이용해 분석 모델을 손쉽게 등록, 수정, 적용 및 모니터링할 수 있습니다. 또한 후보 모델을 개발 및 검증하여 평가한 후 최종 모델을 선정하고, 최종 선택된 모델을 게시하고 모니터링하여 최상의 성능을 보장할 수 있습니다.



## 10. 손쉬운 배포(Deployment)

### 추가 변환 없이, 다양한 운영 환경에 모델 적용

분석 전처리 단계까지 모두 포함된 스코어 코드를 이용하여 변환없이 하나의 코드파일로 모든 과정을 처리할 수 있습니다. 또한 SAS, Python, REST API를 제공하므로 다양한 환경에서 모델을 적용할 수 있습니다. 파이프라인 비교를 통해 여러 모델을 비교한 후 자동으로 최적의 모델을 선정하고, '모델 게시' 버튼을 눌러 손쉽게 배포할 수 있습니다(그림20). 생성된 모델은 DB/Hadoop, 배치, 온라인, 실시간 등 다양한 비즈니스 운영 시나리오의 형태로 적용될 수 있습니다.

### SAS Astore, 빠르고 손쉬운 모델 배포의 비결

머신 러닝의 모델 결과는 전통적인 분석 모델과 달리 매우 복잡하므로, 코드의 길이가 수만~수십만 라인에 달하기도 합니다. 이는 개발한 모델을 다른 운영 환경에 이식하기 어렵게 만들고, 실행 성능에도 영향을 주게 됩니다. 따라서 SAS는 복잡한 머신러닝 결과를 Astore라는 압축된 바이너리 파일 형태로 생성하여 빠르고 손쉽게 운영환경에 배포하는 방법을 제공합니다.

스코어 코드, API 다운로드 및 모델 등록, 모델 게시 화면

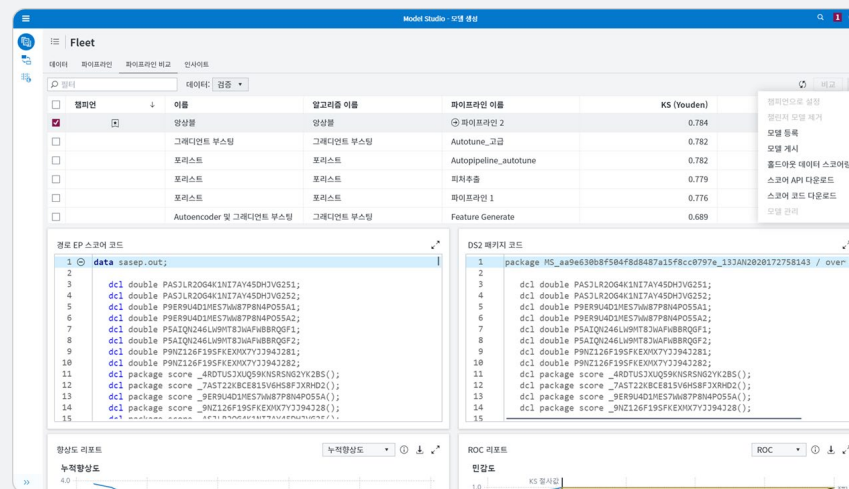
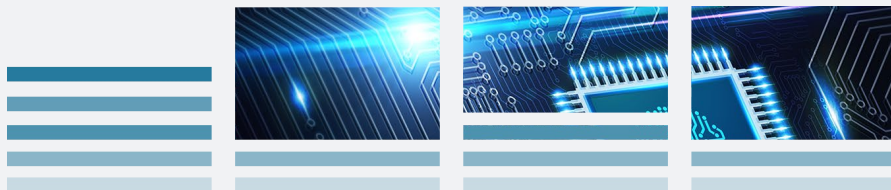


그림20



## 11. 지능적 의사결정(Intelligent Decisioning)

### 비즈니스 룰과 머신 러닝 접목으로, 지능적 의사결정 지원

SAS AI 플랫폼은 의사결정의 관점에서 필요한 비즈니스 룰과 프로세스를 설계하고, 여기에 머신 러닝 모델을 손쉽게 연동하는 지능적 의사결정 기능을 제공합니다. 비즈니스 사용자가 직관적인 비주얼 환경에서 직접 다양한 이벤트를 정의하고 그에 따른 실시간 의사결정 프로세스를 설계합니다. 이를 통해 단순한 룰이 아닌 머신러닝 모델이 함께 결합된 지능적인 의사결정 프로세스를 구성할 수 있습니다(그림21).

의사결정 프로세스에는 SAS뿐만 아니라 Python 및 DB SQL이 함께 사용될 수 있으며, 이를 통해 정의된 의사결정 프로세스는 운영환경 이식을 위한 별도의 추가 개발 없이, 배치 및 실시간 비즈니스 환경에 빠르게 적용됩니다.

의사결정 프로세스 설계 화면

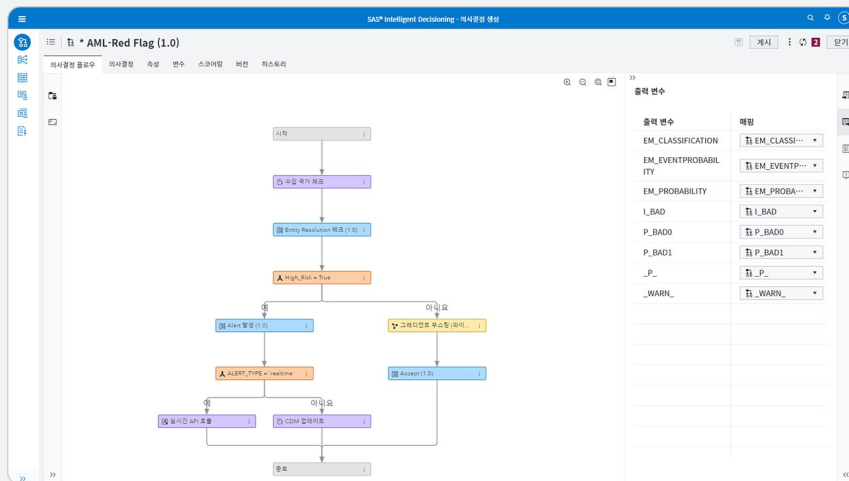


그림21

## 12. 모니터링 및 셀프러닝(Monitoring and Self-Learning)

### 모델 모니터링/셀프러닝으로 모델 성능 최적화

#### 모델 모니터링

모델 모니터링은 보다 효율적인 모델 처리 및 웹 기반 인터페이스를 통해 분석 모델을 손쉽게 테스트하고 비교할 수 있는 기능입니다. 성능 벤치마킹 리포트와 Alert을 생성하여 모델의 성능 저하를 쉽게 추적할 수 있으므로, 지속적인 모니터링을 통해 모델의 정제 또는 폐기가 필요한 때를 식별할 수 있습니다(그림22).

모델 성능 리포트 화면

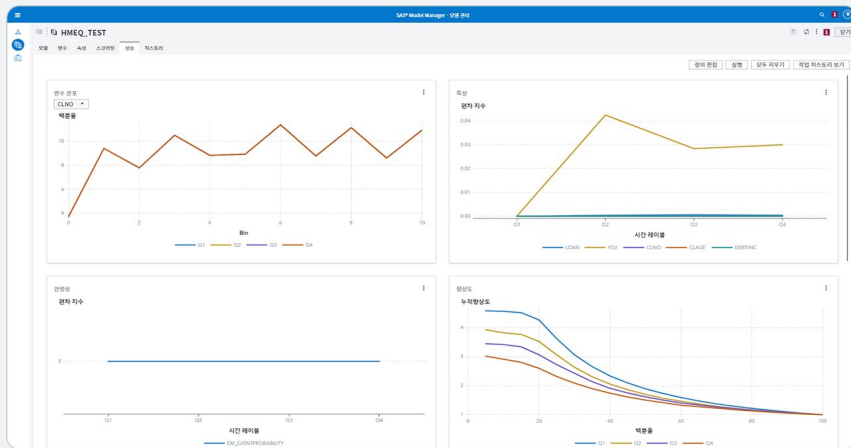


그림22

#### 모델 셀프러닝

모델 셀프러닝은 모델을 재학습시키는 기능입니다. 모델 성능을 평가한 후 성능이 저하된 모델에 대해 '재학습'을 선택하면, 모델 파이프라인과 연동하여 새로운 데이터가 반영된 모델 파이프라인이 수행됩니다. 사용된 변수 재선택과 알고리즘 적용 과정에 대한 재학습을 통해 새로운 최종 모델을 자동으로 선정해주므로, 운영 업무에 곧바로 적용할 수 있습니다(그림23).

모델 셀프러닝 화면

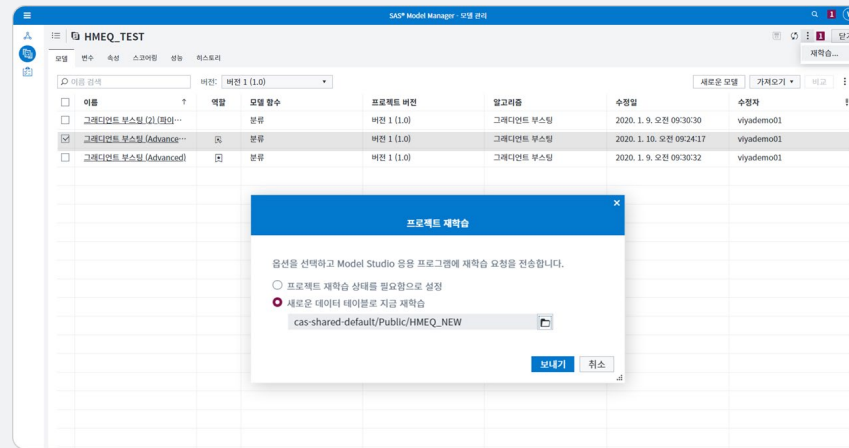


그림23

## ● 개방성 (Open)

### 13. SAS AI 플랫폼의 오픈 소스 개방성(SAS and Open Source Integration)

#### AI-Enhanced Analytics의 인터페이스와 코딩 언어 선택권 제공

SAS AI 플랫폼은 사용자가 인터페이스와 코딩 언어를 선택할 수 있는 개방형 분석 플랫폼입니다(그림24). 효율적인 모델 개발 및 모델 적용에 어려움을 겪어왔던 Python, R 또는 Java와 같은 오픈 소스 사용자들도 자신에게 익숙한 언어로, 앞에서 다루었던 SAS의 매우 다양하고 강력한 AI-Enhanced Analytics 기능을 활용할 수 있습니다.

기업 내 서로 다른 부서들이 각자의 분석 툴과 언어를 사용하면서 부서 간 분석의 장벽(Silo)이 발생하고 있습니다. 이 장벽을 없애는 것이 바로 개방형 분석 플랫폼입니다. 개방형 분석 플랫폼은 모든 분석가들이 익숙한 도구를 가지고 통합 플랫폼이 제공하는 AI-Enhanced Analytics의 장점을 활용하여 빠르게 분석 결과를 도출할 수 있도록 지원합니다. 또한, 분석 결과를 업무에 적용하는데 반드시 필요한 표준 모델 배포 방식을 제공합니다.

SAS AI 플랫폼의 이점과 오픈 소스 에코시스템을 하나로 결합함으로써 편리한 협업 환경에서 서로 다른 분석 도구와 분석 자산이 통합되며, 이는 분석에 대한 높은 생산성과 빠른 비즈니스 적용으로 이어질 수 있습니다.

SAS AI 플랫폼의 오픈 소스 개방성

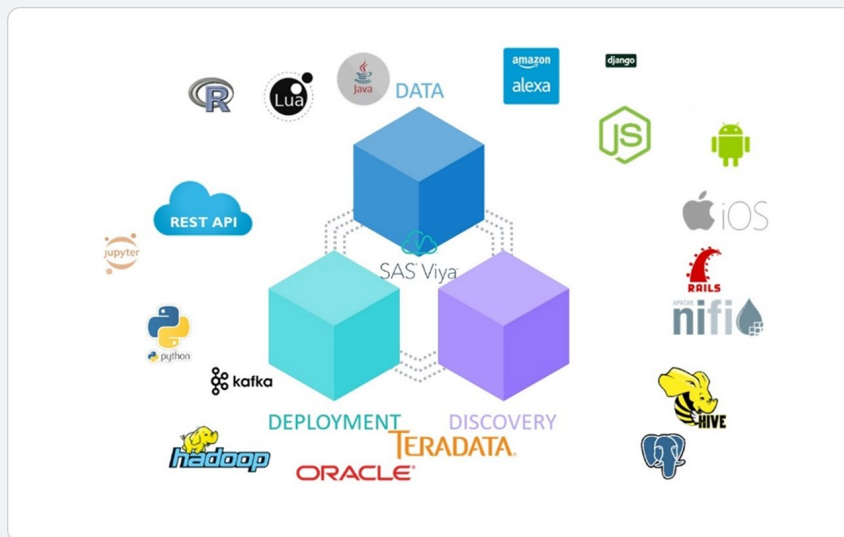
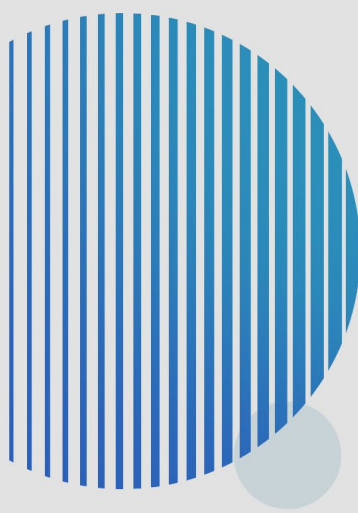


그림24



## AI 가치 실현을 위한 SAS의 도전

# 언제, 어디서나, 누구나, 분석이 가능한 세상

AI는 지난 수년간 SAS 솔루션의 기반이자 핵심이었으며, SAS가 열고자 하는 데이터 세상의 근간입니다.

SAS는 최근, AI 혁신을 지속하기 위해 향후 3년간 AI 분야에 총 10억 달러(약 1조 1천억 원)의 투자 계획을 발표했습니다. 대규모 투자를 통해 AI 연구개발(R&D) 혁신, 교육 프로그램, AI 전문가 컨설팅 강화에 집중할 계획입니다. 특히 고급 분석, 머신 러닝, 딥러닝, 자연어처리(NLP), 컴퓨터 비전 등 모든 AI 핵심 분야에서 지속적인 투자를 통해 기업이 AI 도입 혜택을 보다 잘 이해하고 수익을 최적화하며 비즈니스 과제를 해결하도록 적극 지원하겠습니다.

전체 매출의 26%를 연구개발에 재투자하며 전세계 AI 소프트웨어 플랫폼 시장을 리드하는 SAS는 고급 분석 분야에서 검증된 전문 역량을 바탕으로 '언제, 어디서나, 누구나, 분석이 가능한 세상'을 열고, 전세계 기업 고객이 AI의 가치를 실현할 수 있도록 더욱 노력하겠습니다.





관련하여 자세한 내용은 [sas.com](https://sas.com) 에서 확인하실 수 있습니다.

SAS 및 기타 SAS Institute Inc. 제품 또는 서비스 명칭은 미국을 비롯해 기타 국가에 소재한 SAS Institute Inc.의 등록 상표 또는 상표입니다.  
은 미국 등록을 의미합니다. 그 밖의 브랜드 및 제품 명칭은 각 기업의 상표입니다. Copyright©2020, SAS Institute Inc. All rights reserved.