



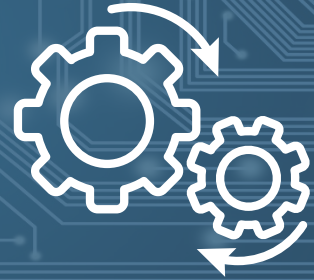
# AutoML을 활용한 금융 플랫폼 혁신 전략

SAS Korea

Advanced Analytics

# Artificial Intelligence

is the science of training systems to emulate human tasks through **learning** and **automation**.



0100  
0110  
1001  
0101  
0100  
1101

0100  
0110  
1001

1010  
0110  
1001  
0101  
0100  
1101  
0101

1010  
0110  
1001  
0101  
0100  
0011

0110  
1001

# Applications of Artificial Intelligence

## Banking



***Fraud Detection***

***Credit Analysis***

Automated  
Financial Advisers

## Government



Smart Cities

***Security  
Intelligence***

Facial Recognition

## Health and Life Sciences



***Predictive  
Diagnostics***

***Biomedical  
Imaging***

Health Monitor

## Manufacturing and Energy



Supply Chain  
Optimization

***Automated  
Defect Detection***

Energy  
Forecasting

## Communications and Retail



***Conversational  
Chat Bots***

Customer  
Intelligence

Network  
Analytics



# What customers are saying about SAS AI



**WildTrack**

AI for Good

**90%**

accuracy for ID of  
wildlife using tracks



**Rogers**

Telecom

**53%**

fewer customer  
complaints



**VUmc**

Healthcare

Improved  
liver & brain  
tumor diagnosis  
with AI & analytics



**Volvo**

Manufacturing

**70%**

reduction in  
diagnostic time

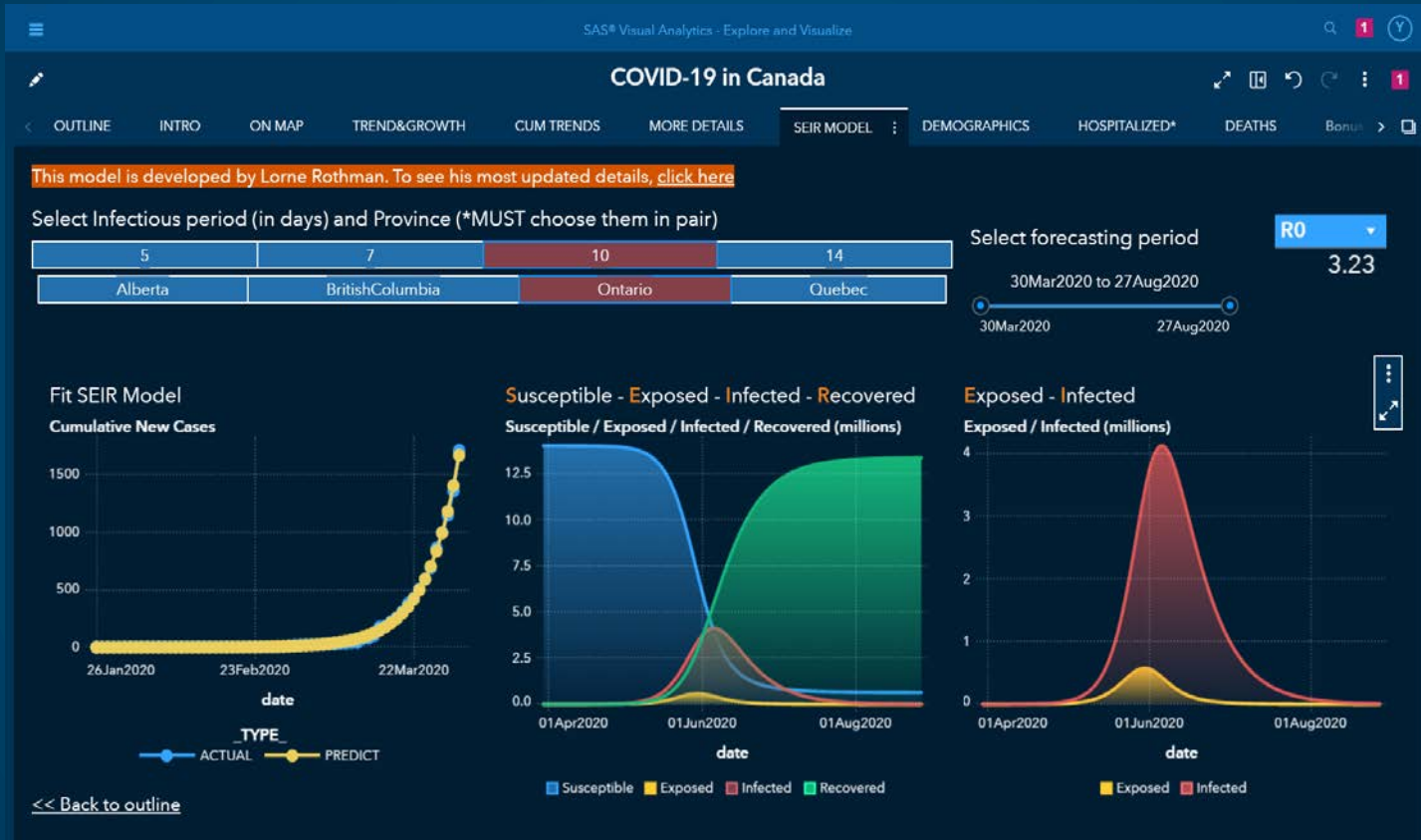


**SciSports**

Services

Improved  
performance  
analysis in real-  
time using AI &  
analytics

# COVID-19 Modeling with SAS (Cleveland Clinic)



# "한국, 2022년까지 AI 인력 1만명 부족할 것" 어떻게 해야 할까?

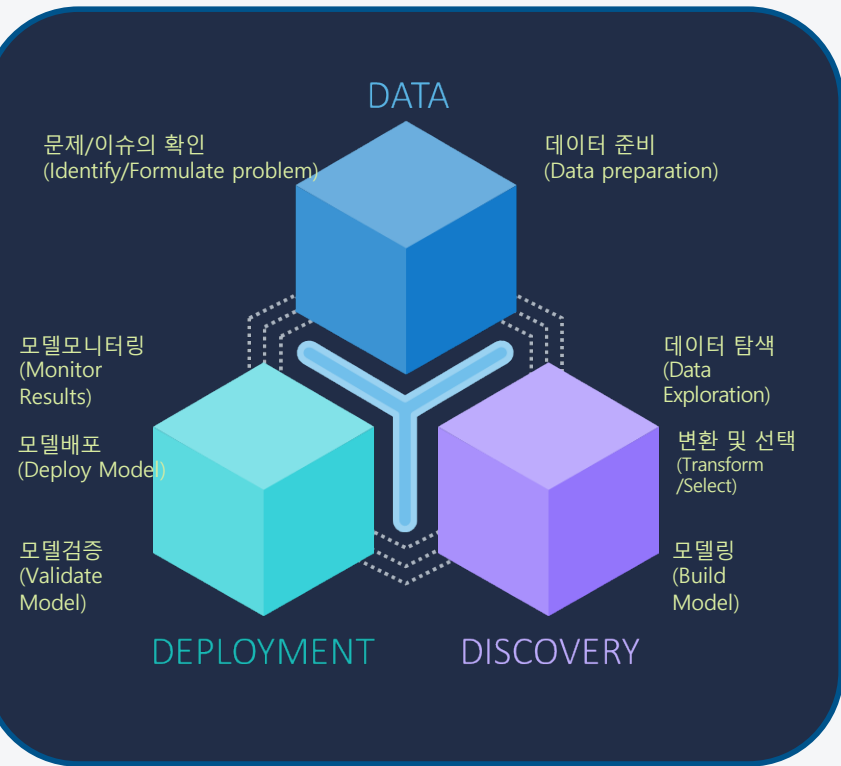


인공지능 발전에 필요한 인력도 부족하다. 소프트웨어정책연구소 분석에 따르면 2022년까지 국내 AI 개발 인력은 현장 수요보다 9986명이 부족할 전망이다. 핵심 인력으로 꼽히는 석사, 박사급 개발자는 7276명이 모자랄 것이란 분석이 나온다. <그래픽 출처=한국경제>



# Analytics Lifecycle

## SAS Analytics Lifecycle (Data-Discovery-Deployment)





# 모두를 위한 AI시대

데이터 과학자는 더 이상 필요 없을까?

## SAS ML For Everyone

The diagram illustrates the progression of SAS ML capabilities across three user roles:

- Data Scientist:** Represented by a person icon and a keyboard icon. The interface shows a code editor with SAS and Python code. Below the code, there are bullet points: "분석 노드를 Drag & Drop 하여 플로우 구성" and "분석 노드의 이해 및 프로세스 구성 필요".
- General Analyst:** Represented by a person icon and a mouse icon. The interface shows a visual pipeline builder with a flowchart of nodes. Below the pipeline, there are bullet points: "분석 노드를 Drag & Drop 하여 플로우 구성" and "분석 노드의 이해 및 프로세스 구성 필요".
- Business User:** Represented by a person icon and a brain icon. The interface shows an AI-driven automation (AutoML) interface with a flowchart. Below the automation, there are bullet points: "데이터 맞춤형 비주얼 파이프라인 (ML 모델) 자동생성" and "White box model".

At the bottom of the diagram, there is a copyright notice: "Copyright © SAS Institute Inc. All rights reserved."

This section is a reflection of the content above, showing the same text and icons mirrored across a horizontal axis.

# 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 '피처 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.

# 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



그림 4

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.

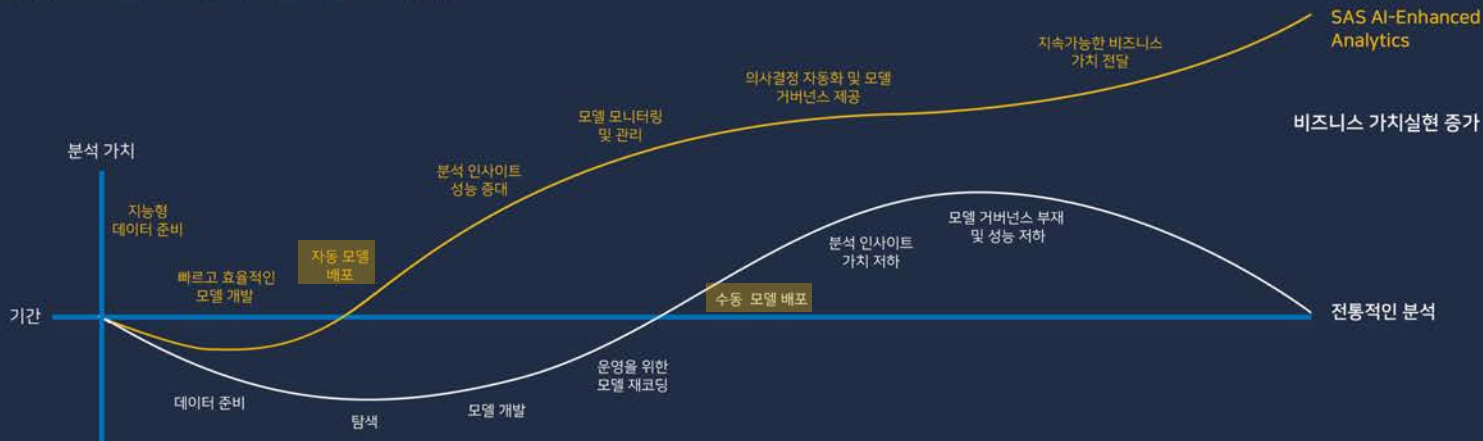
# 지속 가능한 창조와 혁신

## SAS AI-Enhanced Analytics로 비즈니스 가치실현 확대

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



# 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 매직 쿼드란트 '데이터 사이언스 및 머신 러닝 플랫폼'



COMPLETENESS OF VISION →

As of November 2019

©Gartner, Inc





# SAS AI Enhanced Analytics

SAS AutoML 주요 기능 AtoZ

## 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

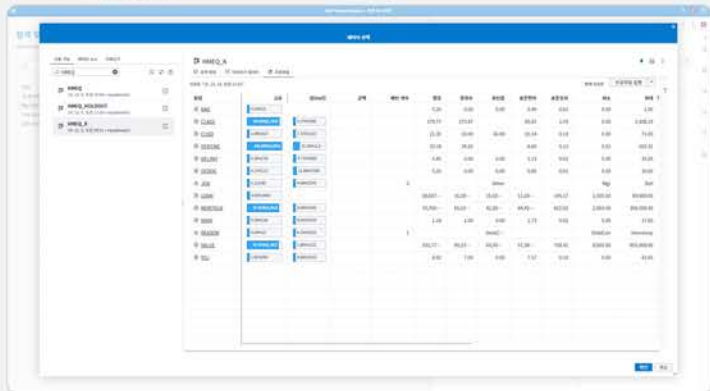


그림6

예를 들어 그림7과 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

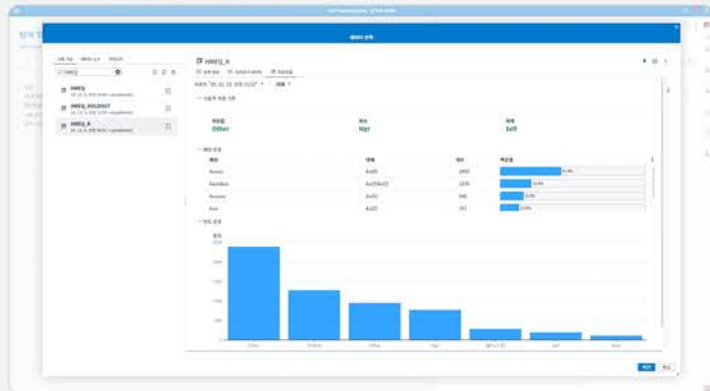


그림7

# 자동 비주얼 인사이트(Automated Explanation)

## 신속한 데이터 탐색 및 예측 인사이트 제공

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

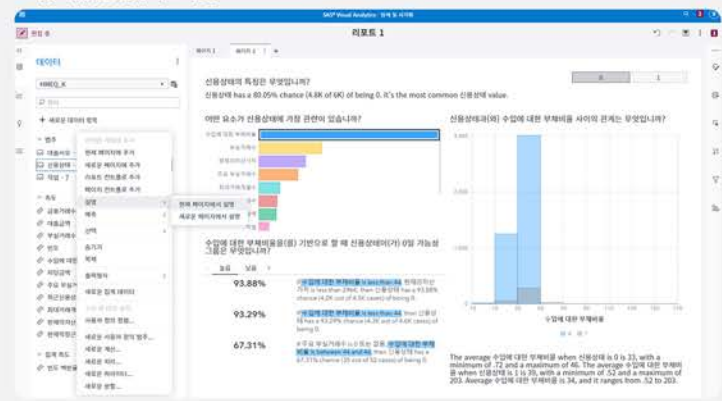
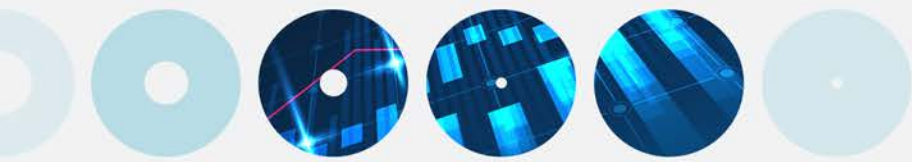


그림9





# 오토 피쳐 엔지니어링(Auto Feature Engineering)

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

피쳐 추출 노드는 저차원 공간에 입력 데이터의 중심 속성을 캡슐화 하는 새로운 피쳐(변수)를 생성하여 변수 타깃에 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

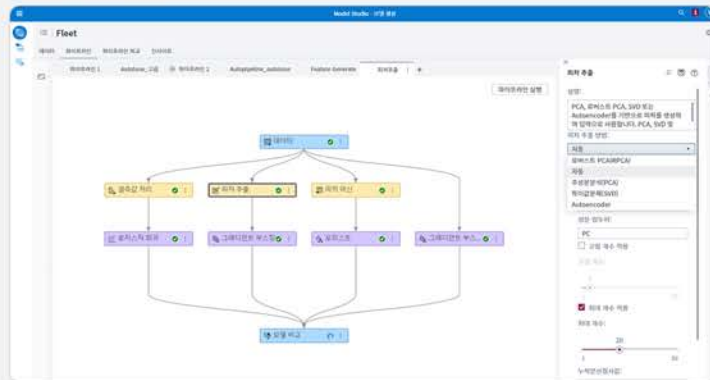


그림13

피쳐 추출 노드와 피쳐 추출 노드 옵션 화면

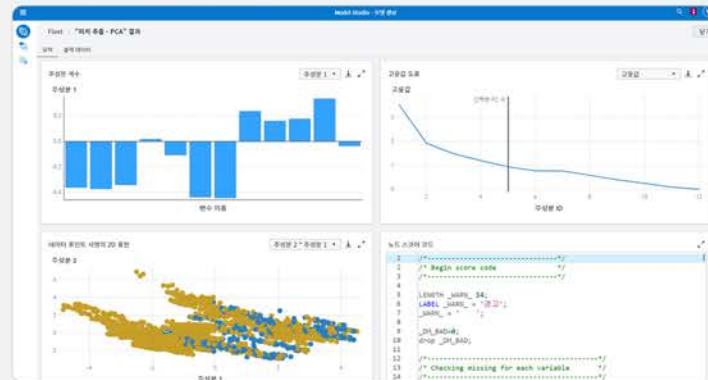


그림14



# 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 & 오토 튜닝 옵션 화면

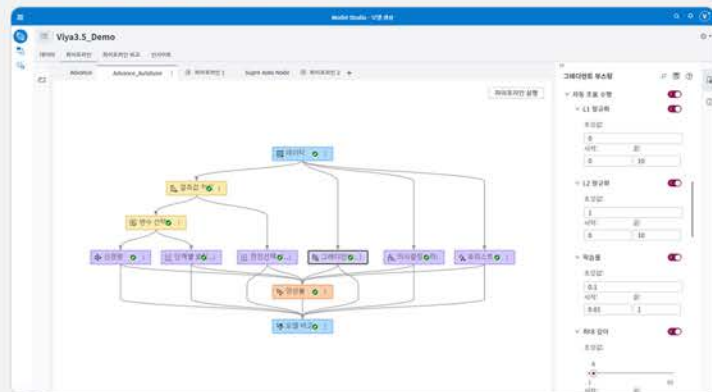


그림15

오토 튜닝 결과 화면



그림16



## 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.

### 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 수준까지 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

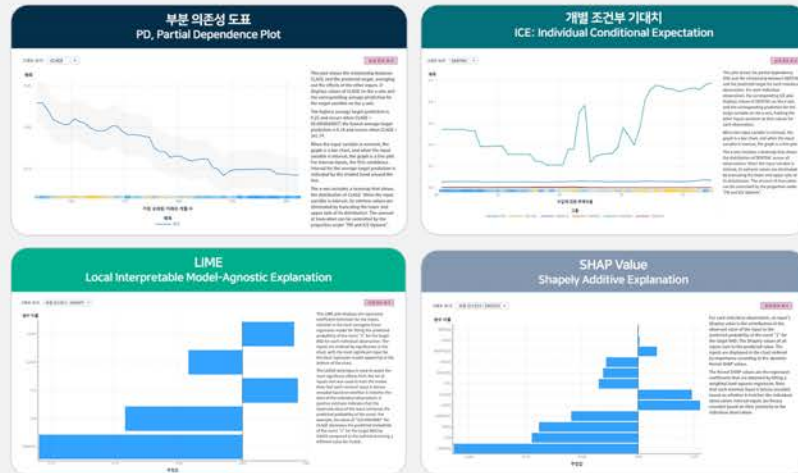


그림 18

## 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

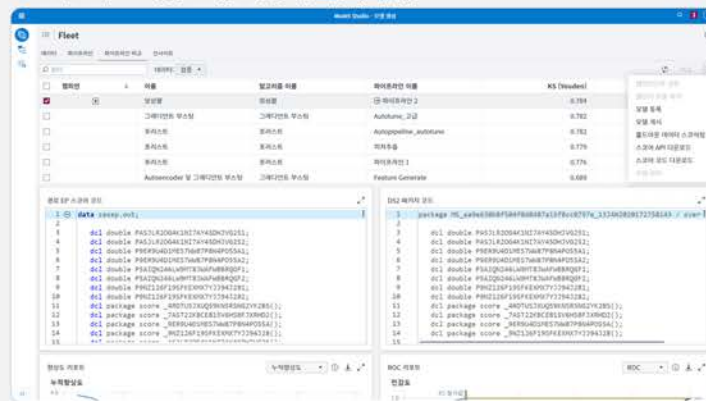


그림20



# 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

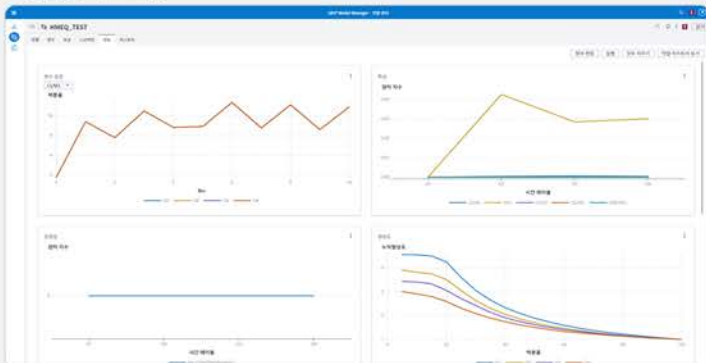


그림22

### 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 최중 모델을 자동으로 선정해주므로, 운영 업무에 곧바로 적용할 수 있습니다(그림23).

모델 셀프러닝 화면

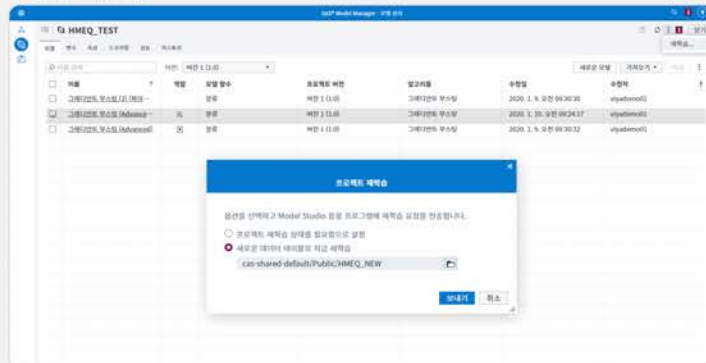


그림23



# ○ 개방성(Open)

## 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

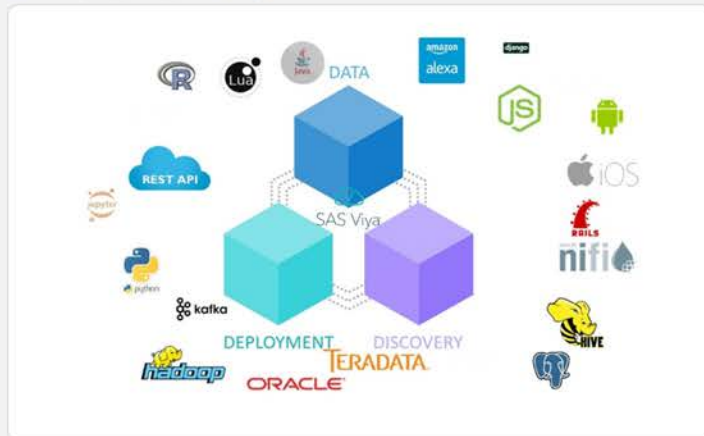


그림24



# Key Takeaway

# Key Takeaway

Analytics Accelerate

분석의 가속화

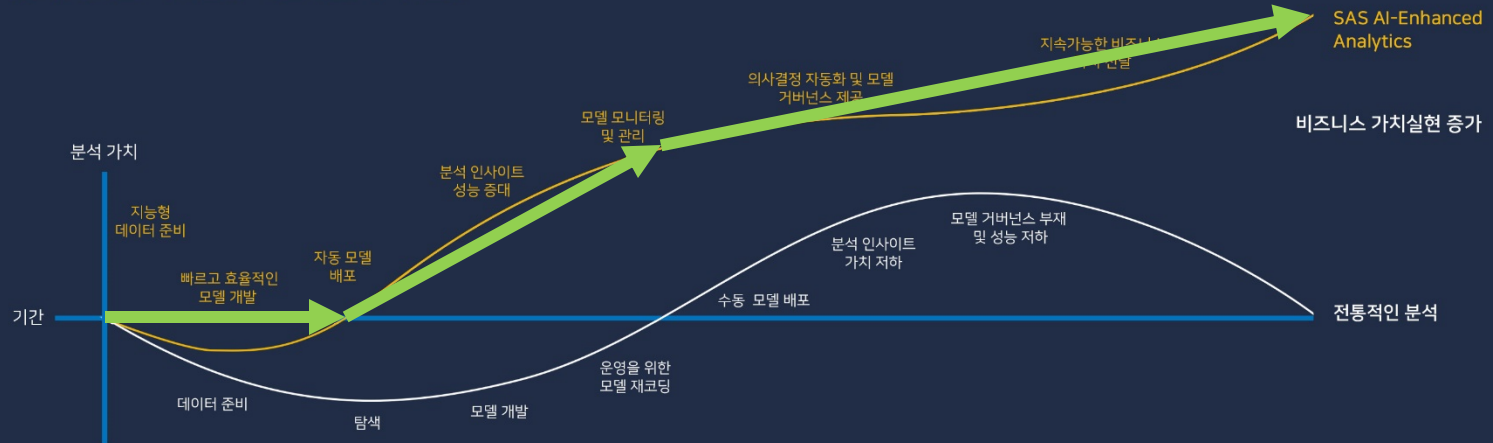
Operation Accelerate

운영의 가속화

Analytics Democratize

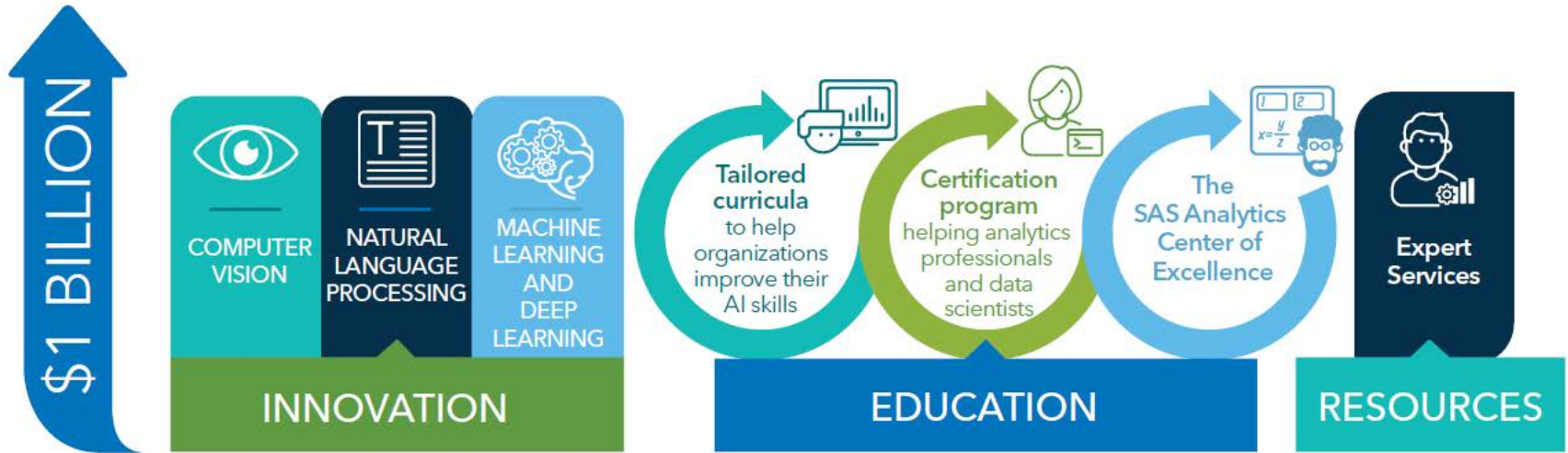
분석의 대중화

## 분석 라이프 사이클과 비즈니스 가치 실현



# SAS Institute Inc.

## SAS Invests \$1 Billion in AI



# SAS Invests \$1 Billion in AI

Empowering the next generation of AI possibilities

# SAS® Viya Trial Program (14 Days)



- SAS 클라우드 서버에 1024MB 데이터 업로드 가능
- Drag & Drop으로 빠르게 모델 결과 구축
- 클라우드 기반에서 SAS 프로그램을 사용
- 다양한 시각화 기능으로 손쉬운 모델 파이프라인 구축
- 동료와 분석결과 공유 가능
- 신청: <https://www.sas.com/korea> 무료평가판 신청







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.