

AI를 쉽고 빠르게 할 수 있는 AutoML - H2O Driverless AI 소개

IBM Systems

유효종 차장, Cognitive Systems
010-4995-8466 hjrhew@kr.ibm.com



HoP



NVMe



M2P9



SDS



LinuxONE



Cognitive

기업에서의 AI 채택이 생각보다 부진한 이유

Data is a Team Sport

~100

Data science experts in the world

AI 전문 인력 부족

Months

Time for a data scientist to build a model

느린 개발 속도



Black box models

AI에 대한 신뢰 부족

“미국에서만도 분석 전문가의 인력 부족이 19만명에 달한다.”

McKinsey&Company

기업에 excel 파일이 넘쳐나는데도 그를 이용한 AI 활용이 부진한 것은 skill과 시간이 부족할 뿐만 아니라, AI의 판단에 대한 설명이 불가능하기 때문입니다.

H2O Driverless AI : 기계 학습의 자동화를 통한 쉽고 빠른 AI



Excel 표에서 예측하고자 하는 칼럼만 지정하면 예측 모델을 자동생성

시스템 자원과 상황에 따라 적절한 '정확도, 시간, 해석도'를 다이얼로 쉽게 조절

✓ H2O Driverless AI는 예측 모델을 자동으로 생성해줌으로써 "쉽고 빠른 AI 적용"을 가능하게 만들어주는 Auto ML 솔루션입니다.

✓ H2O Driverless AI는 단순한 브라우저 GUI를 통해 클릭 몇 번으로 전문지식 없이도 예측 모델을 생성하여 현업에 사용할 수 있습니다.

AI의 판단 근거를 설명해주는 사유부호 제공

✓ 현업에서는 예측 모델의 정확도 뿐만 아니라, AI가 왜 그렇게 예측했는지에 대한 설명이 필요합니다.

✓ H2O Driverless AI는 k-LIME과 decision tree 등의 기법을 통해 현업에서도 쉽게 이해할 수 있는 사유 부호를 제공합니다.

Global Reason Codes [About reason codes](#)

Global interpretable model explains 59.54% in p(Class) for the entire dataset with RMSE = 0.024.

Variable	with value/ 1 unit increase (if blank)	is associated with p(Class)
Top Positive Global Attributions		
V1		increase of 0.006
V2		increase of 0.0042
V4		increase of 0.0033

GPU를 이용한 Acceleration

- ✓ H2O Driverless AI는 GPU acceleration을 이용하여 model training의 시간을 획기적으로 단축할 수 있습니다.
- ✓ IBM의 AC922은 CPU-GPU간의 NVLink2.0 (150GB/s)기능을 탑재하고 있으며, H2O Driverless AI를 사용하기에 최고의 인프라를 제공합니다.



H2O Driverless AI의 주요 기능

1. 자동화된 파생 변수 생성
2. 설명 가능한 Machine Learning (ML)
3. TensorFlow를 이용한 자연어 분석 (NLP)
4. Time Series 데이터 분석
5. 자동화된 시각화
6. 자동화된 소스코드 생성
7. 자동화된 분석 문서 생성
8. NVIDIA GPU Acceleration
9. Bring-Your-Own Recipes (Customized ML)



H2O Driverless AI: “Expert Data Scientist in a Box”

1. Data drag & drop

Data를 클라우드, 빅데이터, 데스크탑 등에서 ingest

2. 자동화된 시각화

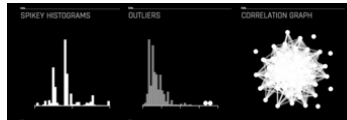
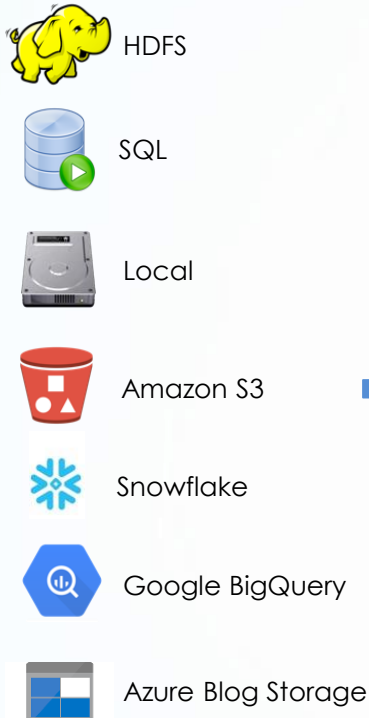
Data shape, outliers, missing values 등을 이해

3. 자동화된 기계 학습

진보된 feature engineering과 parameter tuning을 포함한 수천개의 가능성 있는 모델들에 걸쳐 반복 훈련

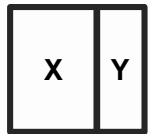
4. 자동화된 Scoring Pipelines

Low-latency Python 또는 Java로 구현된 Automatic Scoring Pipelines를 deploy



Automatic Machine Learning

Modelling

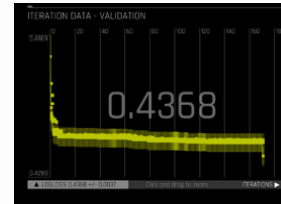


Dataset

Model Recipes:

- i.i.d. data
- Time-series
- More on the way

Survival of the Fittest



Advanced Feature Engineering

+ Algorithm + Model Tuning

Powered by GPU Acceleration

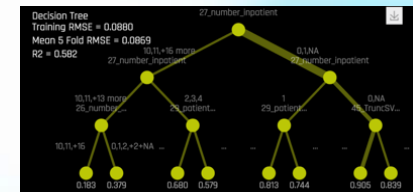


Model Documentation

Automatic Scoring Pipeline

Deploy Low-latency Scoring to Production

Machine Learning Interpretability



Use case : 금융 사기 탐지, 타겟 마케팅, 수요 예측 등



- ✓ Driverless AI는 10년 경력의 feature engineering 전문가에 필적
- ✓ 사기 행위 탐지 정확도를 0.89에서 0.947로 6% 향상
- ✓ H2O4GPU with Driverless AI 사용으로 6배 속도 향상



- ✓ 다른 부동산 디지털 마케팅 솔루션보다 2.5배 뛰어난 효과
- ✓ G5 고객사 중 연간 디지털 마케팅 비용 \$500K를 절감하면서도 웹 트래픽을 3배로 늘린 사례
- ✓ 10배 빠른 모델 생성



- ✓ 1명의 data scientist로 25% 시간 절약
- ✓ 제조 생산 라인을 위한 모델 튜닝과 훈련에 1달의 시간 단축
- ✓ Forecast 정확성 향상을 통해 미래 고객 주문을 위한 필요 부품 및 원자재를 정확히 예측

감사합니다.



HoP



NVMe



M2P9



SDS



LinuxONE



Cognitive