

 데이터솔루션



---

## MS Azure Machine Learning(Automated ML)

데이터솔루션 AI사업팀

# Azure Machine Learning

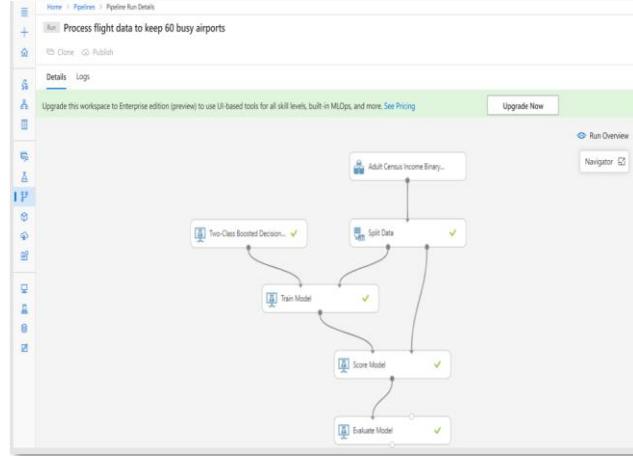
# 01 | Azure Machine Learning

## Machine Learning for all skills

The screenshot shows the Azure Machine Learning UI. At the top, there's a navigation bar with 'Refresh' and 'Create experiment'. Below it is a 'Run Status History' chart with four categories: 'Running' (2), 'Completed' (0), 'Failed' (0), and 'Others' (10). The 'Completed' section has a note: '2018-09-29 2018-09-29'. Below the chart is a table of experiments:

| Experiment                           | Run Id                                       | Status   | Created               | Start Time            | End Time       |
|--------------------------------------|--|----------|-----------------------|-----------------------|----------------|
| adult-lb-user-remote-dnn-v1-pipeline | AdultLB_3f0361d4-f419-4f05-9e6c-0f6c22e1a1b2 | Running  | 9/30/2019 11:58:45 AM | 9/30/2019 12:10:48 PM |                |
| ui-Jason-1                           | AdultLB_fabf151b-4760-4f95-9c26-5fa4a1320495 | Running  | 9/30/2019 11:45:57 AM | 9/30/2019 12:02:04 PM |                |
| adult-lb-user-remote-dnn-v1-spark    | AdultLB_b3b8aef5-ecae-45af-9bae-0f5ca0490055 | Canceled | 9/30/2019 1:54:18 AM  | 9/30/2019 2:05:41 AM  | 9/30/2019 1... |
| adult-lb-user-remote-dnn-v1-spark    | AdultLB_30373109-e53b-47b1-9f9c-0f6c22e1a1b2 | Canceled | 9/30/2019 1:48:20 PM  | 9/30/2019 8:23:14 PM  | 9/30/2019 1... |
| adult-lb-user-remote-dnn-v1-spark    | AdultLB_440494ca-9bb4-4f62-9f9c-0f6c22e1a1b2 | Canceled | 9/30/2019 7:40:14 PM  | 9/30/2019 8:11:09 PM  | 9/30/2019 1... |
| adult-lb-user-remote-dnn-v1-spark    | AdultLB_fbf4794c-4f51-44d1-b6de-0f5ca0490055 | Canceled | 9/30/2019 5:56:53 PM  | 9/30/2019 5:58:21 PM  | 9/30/2019 1... |
| adult-lb-user-remote-dnn-v1-spark    | AdultLB_aed9105e-3432-42d3-92cc-0f6c22e1a1b2 | Canceled | 9/30/2019 5:15:32 PM  | 9/30/2019 5:22:38 PM  | 9/30/2019 1... |
| adult-lb-user-remote-dnn-v1-spark    | AdultLB_29f1b6cc-955c-452c-9ed6-0f6c22e1a1b2 | Canceled | 9/30/2019 4:57:31 PM  | 9/30/2019 4:58:48 PM  | 9/30/2019 1... |

### Automated Machine Learning UI



### Azure Machine Learning Designer

The screenshot shows the Machine Learning Notebooks interface. At the top, it says 'Compute Instance' and 'Search or search by name'. Below that is a 'Train the automatic regression model' section with a note: 'Create an experiment object in your workspace. An experiment acts as a container for your machine learning tasks. You can either upload, add, or drag objects to the experiment, and set the output to "Train" to run the process during the run.' There are several code snippets and explanations for training, testing, and evaluating models.

### Machine Learning Notebooks

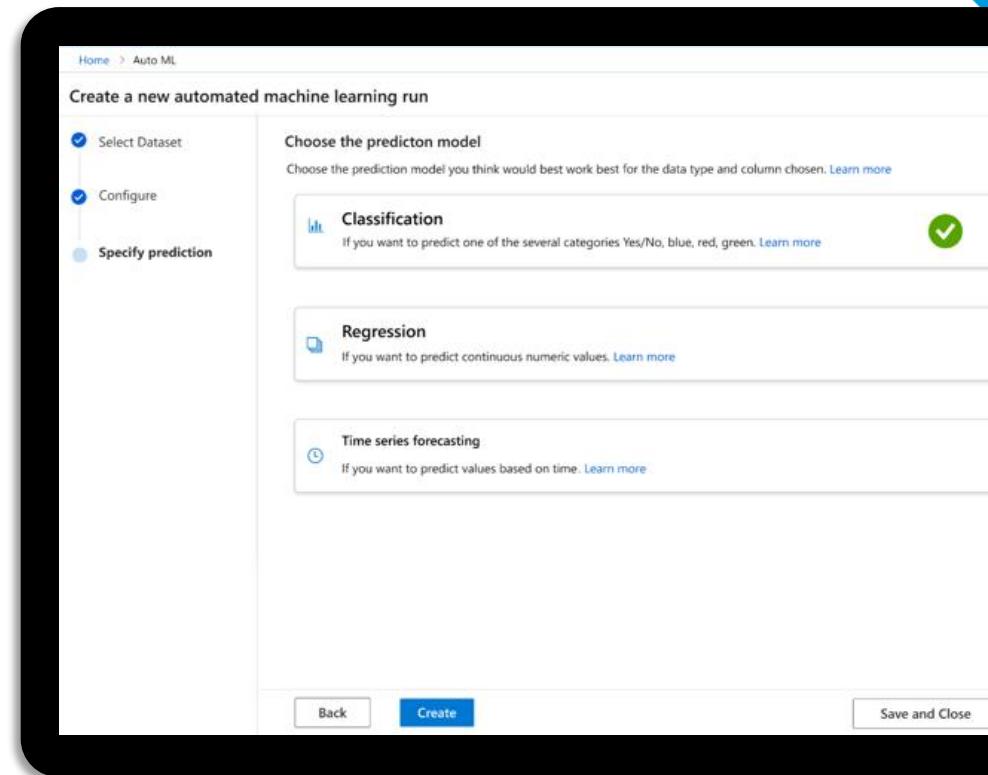
## 02 | Automated Machine Learning

머신 러닝 모델 개발의 시간이 많이 걸리는 반복적인 작업을 자동화 하는 프로세스

코드 없는 UI를 통해 머신 러닝 모델을 자동 구축 및 배포

모델 품질을 유지하면서 확장성, 효율성 및 생산성이 높은 ML 모델을 구축할 수 있음

데이터에 맞게 다양한 알고리즘 및 하이퍼 파라미터 튜닝으로 정확한 모델을 쉽게 생성할 수 있음

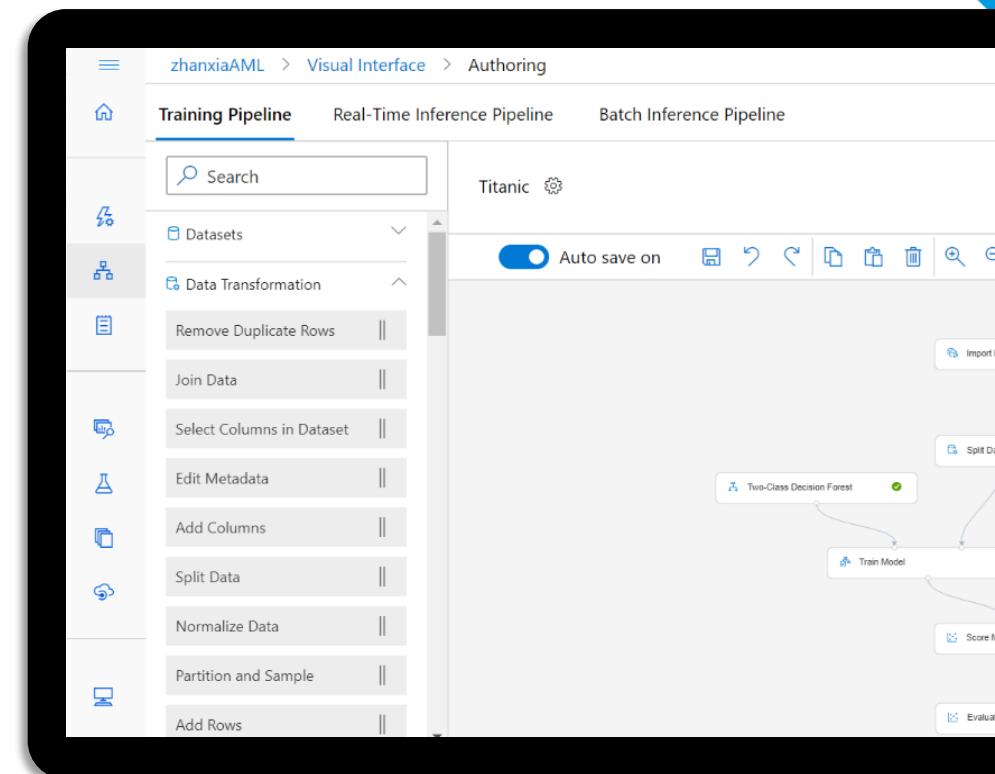


# 03 | Designer

Drag-n-drop 워크플로우 기능

머신러닝 모델 구축, 테스트 및 운영하는 프로세스 단순화

Built-in 모듈, 데이터 시각화, 모델 평가 등 pipeline으로 모델 구축

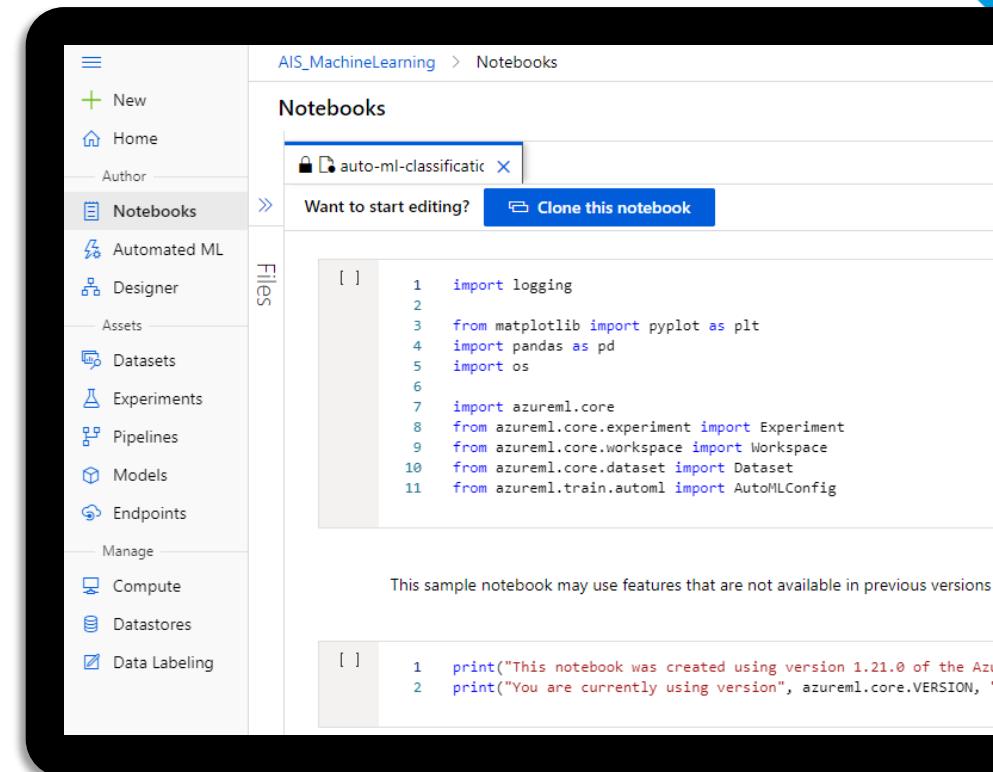


# 04 | Machine Learning Notebooks

데이터 과학자가 머신 러닝을 시작할 수 있는 클라우드 기반 솔루션

Azure Machine Learning Python 및 R SDK를 사용하는 통합 Notebooks  
를 통해 모델을 쉽게 구축 및 배포

공동 작업 공간을 통해 모델 디버그 및 Notebooks 공유



```
import logging
from matplotlib import pyplot as plt
import pandas as pd
import os
import azurerm.core
from azurerm.core.experiment import Experiment
from azurerm.core.workspace import Workspace
from azurerm.core.dataset import Dataset
from azurerm.train.automl import AutoMLConfig
```

This sample notebook may use features that are not available in previous versions

```
print("This notebook was created using version 1.21.0 of the Azurerm Python library")
print("You are currently using version", azurerm.core.VERSION, "of the Azurerm Python library")
```

# 05 | Automated ML Overview

## Input

101010  
010101  
101010

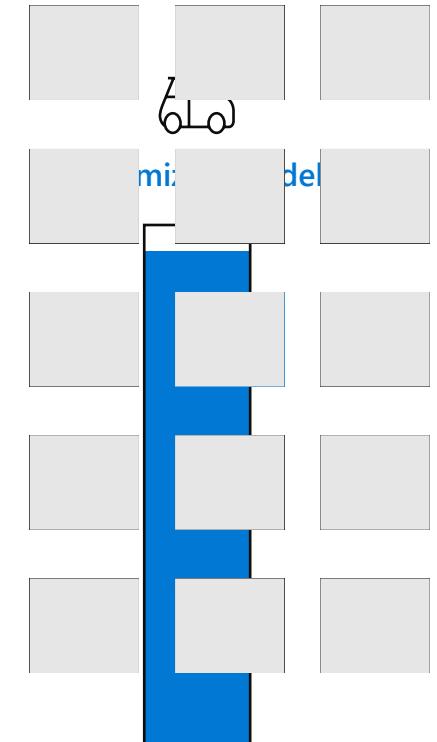
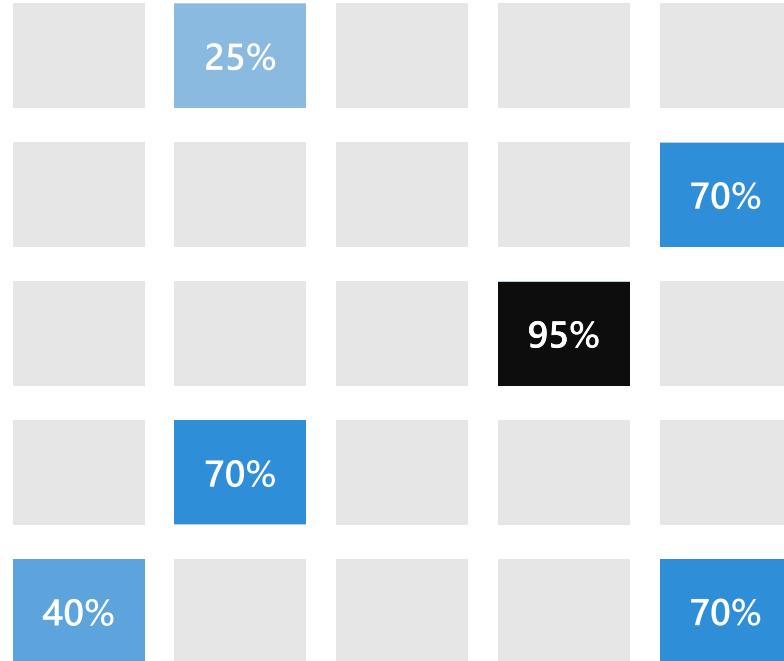
Enter data



Define goals



Intelligently test multiple models in parallel



# 06 | Algorithms

| Classification                    | Regression                        | Time Series Forecasting           |
|-----------------------------------|-----------------------------------|-----------------------------------|
| Logistic Regression               | Elastic Net                       | Elastic Net                       |
| Stochastic Gradient Descent (SGD) | Gradient Boosting                 | Gradient Boosting                 |
| Multinomial Naive Bayes           | Decision Tree                     | Decision Tree                     |
| Bernoulli Naive Bayes             | K-Nearest Neighbors               | K-Nearest Neighbors               |
| SVM                               | LARS Lasso                        | LARS Lasso                        |
| K-Nearest Neighbors               | Stochastic Gradient Descent (SGD) | Stochastic Gradient Descent (SGD) |
| Decision Tree                     | Random Forest                     | Random Forest                     |
| Random Forest                     | Extremely Randomized Trees        | Extremely Randomized Trees        |
| Extreme Random Trees              | Light GBM                         | Light GBM                         |
| Light GBM                         | Xgboost Regressor                 | Xgboost Regressor                 |
| Gradient Boosting                 | Fast Linear Regressor             | Auto-ARIMA                        |
| XGBoost Classifier                | Online Gradient Descent Regressor | Prophet                           |
| Linear SVM                        |                                   | TCN Forecaster                    |
| Averaged Perceptron Classifier    |                                   |                                   |

## 알고리즘 자동화

- AutoML은 분류, 회귀 및 시계열 예측을 포함하여 세 가지 유형의 작업을 지원
- AutoML 작업에서 자동화 및 튜닝 프로세스 중에 다음 알고리즘을 지원하고 있으며 사용자는 알고리즘을 지정할 필요가 없음

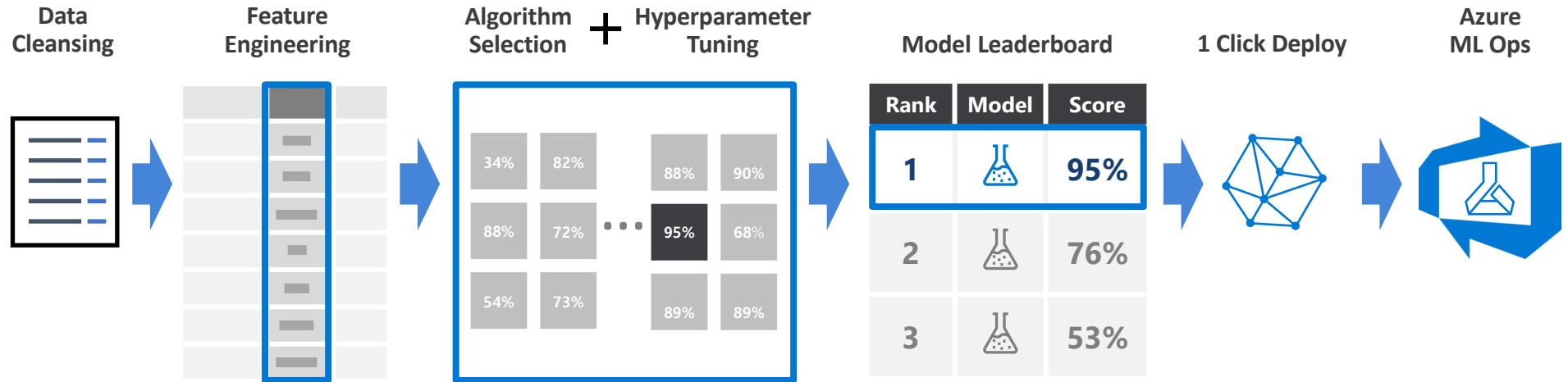
# 07 | Features

| Standard Featurizers | Advanced Featurizers                    |
|----------------------|---|
| StandardScaleWrapper | 높은 카디널리티 제외 (Drop High Cardinality)     |
| MinMaxScaler         | 결측값 대체/처리 (Missing Value Imputation)    |
| MaxAbsScaler         | 추가 변수 생성 (Generate Additional Features) |
| RobustScaler         | 변환 및 인코딩 (Transform and Encode)         |
| PCA                  | 워드임베딩 (Word Embedding)                  |
| TruncatedSVDWrapper  | 타켓 인코딩 (Target Encoding)                |
| SparseNormalizer     | 텍스트 타켓 인코딩 (Text Target Encoding)       |
|                      | Weight of Evidence (WoE)                |
|                      | 클러스터 거리 (Cluster Distance for K-means)  |

## Feature Engineering

- AutoML에 데이터를 적용하기 전에 데이터가 알맞게 표현되었는지 확인하고 기존 변수를 기반으로 추가 변수를 생성

# 08 | Azure AutoML Process



# Data Science with Azure Automated ML

DATA SCIENCE  
DEPARTMENT

# 01 | 실습데이터

| Columns                   | Description  |
|---------------------------|--|
| transactionID             | Unique transaction id  |
| accountID                 | Unique account id  |
| transactionAmountUSD      | Transaction amount in USD e.g., 12345.00   |
| transactionAmount         | Transaction amount in currency expressed in transactionCurrencyCode e.g., 12345.00   |
| transactionCurrencyCode   | Currency code of the transaction. 3 alphabet letters, e.g., USD  |
| transactionDate           | Date when transaction occurred. Typically in the time zone of the processor, Format: yyyyymmdd, e.g., 20000101   |
| transactionTime           | Time when transaction occurred. Typically in the time zone of processing end. Format: hhmmss, eg. 153059   |
| localHour                 | The hour in local time. Value of 0-23  |
| transactionIPaddress      | Full IP Address for IPv4: 000.000.000.000  |
| ipState                   | State of IP address originated from 2 alphabet letters   |
| ipPostcode                | Postal Code of IP address originated from  |
| ipCountryCode             | Country code of IP address originated from   |
| isProxyIP                 | Whether the IP address is a proxy or not   |
| browserLanguage           | Similar to country code  |
| paymentInstrumentType     | Type of payments:<br>C -- Credit Card<br>D -- Debit Card<br>P -- Paypal<br>K -- Check<br>H -- Cash<br>O -- Other   |
| cardType                  | Type of cards M -- Magnetic C -- Chip  |
| paymentBillingPostalCode  | payment billing postal code  |
| paymentBillingState       | payment billing state  |
| paymentBillingCountryCode | payment billing country code   |
| cvvVerifyResult           | M-- CVV2 Match<br>N-- CVV2 No Match<br>P--Not Processed<br>S--Issuer indicates that CVV2 data should be present on the card, but the merchant has indicated data is not present on the card<br>U--Issuer has not certified for CVV2 or Issuer has not provided Visa with the CVV2 encryption keys<br>Empty--Transaction failed because wrong CVV2 number was entered or no CVV2 number was entered |
| digitalItemCount          | Number of digital items purchased. (e.g. music, ebook, software, etc, that can be directly downloaded online)  |
| physicalItemCount         | Number of physical items purchased (that needs to be shipped)  |
| label                     | True/False for fraud transaction   |

## 온라인 거래 사기 데이터

- Woodgrove 은행에서 제공하는 온라인 거래 내역
- 사기 데이터: 1,151건
- 비사기 데이터: 198,849건
- Undersampling을 통해 비사기 데이터 1,151건으로 축소
- 총 2,302건의 온라인 거래 데이터를 활용

## 실습 목적

- 온라인 거래 데이터 기반으로 사기 데이터를 식별하기 위한 최적 분류 모델 구축

## 02 | MS Azure 로그인

The screenshot shows the Microsoft Azure homepage. At the top, there is a navigation bar with links for '개요', '솔루션', '제품', '설명서', '가격', '교육', 'Marketplace', '파트너', '지원', '블로그', and '기타'. On the far right of the navigation bar, there are buttons for '영업팀에 문의', '검색', '내 계정', '포털' (which is highlighted with a red box), and '로그인'. Below the navigation bar, there is a large banner with the text 'Azure 목표를 기반으로 개발하세요.' and a subtext about building, deploying, and managing applications. To the right of the banner is a graphic featuring a laptop, code snippets, and a cloud. A call-to-action button labeled '체험 계정' is visible. In the center, the URL 'https://azure.microsoft.com/' is displayed. Below the URL, there is a section titled 'Azure를 통해 획기적인 아이디어를 솔루션으로 구현하기' with four cards: '미래를 위한 준비' (with a rocket icon), '원하는 조건에 따른 빌드' (with a server icon), '원활하게 하이브리드 운영' (with a building icon), and '신뢰할 수 있는 클라우드' (with a cloud icon). Each card has a brief description below it.

<https://azure.microsoft.com/>

Azure를 통해 획기적인 아이디어를 솔루션으로 구현하기

미래를 위한 준비  
원하는 조건에 따른 빌드  
원활하게 하이브리드 운영  
신뢰할 수 있는 클라우드

Microsoft는 지속적인 혁신을 통해 오늘 날의 개발과 미래의 제품 비전을 지원합니다.  
오픈 소스를 보장하고 모든 언어와 프레임워크를 지원하므로 원하는 방법으로 빌드하고 원하는 곳에 배포할 수 있습니다.  
온-프레미스, 클라우드, 에지 등 모든 환경을 지원합니다. 하이브리드 클라우드용으로 디자인된 서비스를 사용하여 환경을 통합하고 관리하세요.  
여러 엔터프라이즈, 정부 기관, 스타트업이 신뢰하는 능동적 규정 준수와 보안 전문가 팀의 지원을 통해 철저한 보안성을 제공합니다.

### MS Azure 로그인 화면

- 계정 로그인 진행
- 로그인 진행한 후에 <포털>탭을 클릭

# 03 | 리소스 그룹 접속

**Azure 서비스**

리소스 만들기 1 리소스 그룹

리소스 그룹

가상 머신 App Services SQL 데이터베이스 Azure Database for PostgreSQL... Azure Cosmos DB → 추가 서비스

**최신 리소스**

| 이름                       | 형식                       | 마지막으로 본 날짜 |
|--------------------------|--------------------------|------------|
| AIS_ML                   | 기계 학습                    | 2시간 전      |
| AIS_TEST                 | 리소스 그룹                   | 2시간 전      |
| asomachinelear0069975310 | 스토리지 계정                  | 2시간 전      |
| ASO_ML                   | 리소스 그룹                   | 2시간 전      |
| ASO_MachineLearning      | 기계 학습                    | 3시간 전      |
| asogen2storage           | 스토리지 계정                  | 2개월 전      |
| ds_Databricks            | Azure Databricks Service | 2개월 전      |
| ASO_Demo                 | 리소스 그룹                   | 2개월 전      |
| ASODataFactory           | Data Factory(V2)         | 2개월 전      |
| ASO_Databricks           | 리소스 그룹                   | 2개월 전      |

## 리소스 그룹 접속

1. Azure 포털 화면 상단에 있는 Azure 서비스 중에 <리소스 그룹>을 선택

# 04 | 리소스 그룹 생성

**리소스 그룹**

1. 새로 만들기

2. 구독 == 모두

3. 위치 == 모두

4. AIS\_BOT

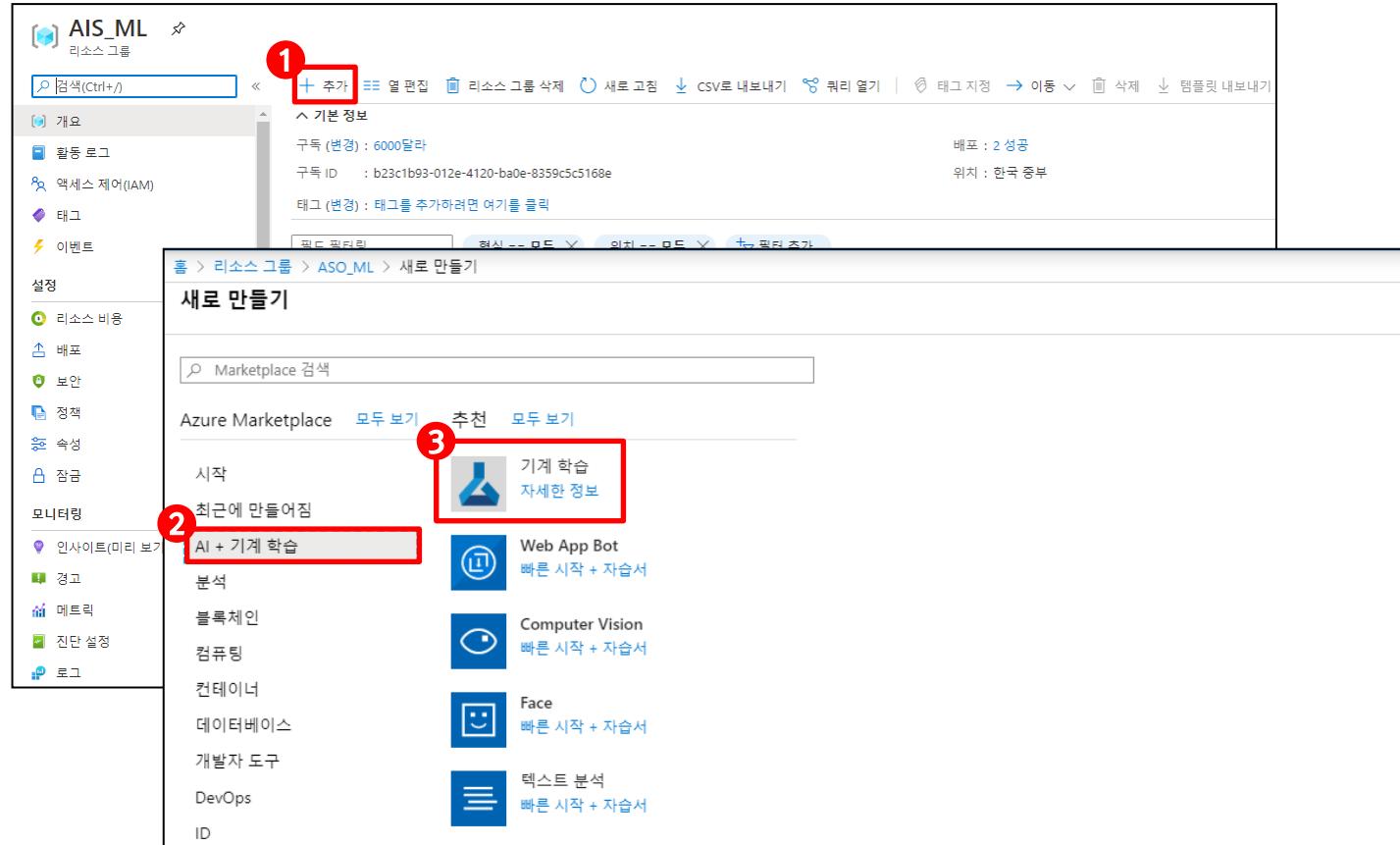
5. (Asia Pacific) 한국 중부

6. 검토 + 만들기

## 리소스 그룹 생성

- 리소스 그룹 화면 상단에 있는 <새로 만들기>을 선택
- 구독 ID를 선택하고 리소스 그룹 이름을 "AIS\_ML"으로 입력
- 리소스 영역에서 "한국 중부"를 선택
- <검토 + 만들기>버튼을 클릭하여 리소스 그룹 생성

# 05 | Azure Machine Learning 추가



## Azure Machine Learning 추가

1. "AIS\_ML" 리소스 그룹 화면에서 상단에 있는 <추가> 버튼을 클릭하여 <새로 만들기> 화면을 접속
2. <Azure Marketplace>목록에서 <AI+ 기계 학습>을 클릭
3. 추천 리스트에서 <기계 학습>을 선택

# 06 | Azure Machine Learning 추가

홈 > 리소스 그룹 > AIS\_ML > 새로 만들기 >

**Machine learning**  
기계 학습 작업 영역 만들기

기본 네트워킹 고급 태그 검토 + 만들기

**프로젝트 정보**

배포된 리소스와 비용을 관리할 구독을 선택합니다. 둘더 같은 리소스 그룹을 사용하여 모든 리소스를 정리 및 관리합니다.

구독 \* ⓘ 6000달라

리소스 그룹 \* ⓘ AIS\_ML  
새로 만들기

**작업 영역 세부 정보**

작업 영역의 이름 및 지역을 지정하세요.

작업 영역 이름 \* ⓘ 4 AIS\_MachineLearning1

지역 \* ⓘ 한국 중부

스토리지 계정 \* ⓘ (신규) aismachinelear7771574490  
새로 만들기

키 자격 증명 모음 \* ⓘ (신규) aismachinelear4832606669  
새로 만들기

Application Insights \* ⓘ (신규) aismachinelear5295402772  
새로 만들기

컨테이너 레지스트리 \* ⓘ 5 Sade405451584f0c87b917019e5aff77  
새로 만들기

6 검토 + 만들기

## Azure Machine Learning 추가

4. <기계 학습> 화면에서 생성할 작업 영역 이름을 "AIS\_MachineLearning"으로 입력
5. 컨테이너 레지스트리에서 기존 생성된 레지스트리를 선택
6. 모두 기본 정보를 입력 완료 후 <검토 + 만들기>버튼을 클릭

# 07 | 리소스 그룹 재접속

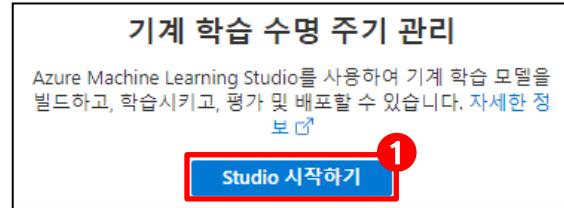
The screenshot shows the Microsoft Azure portal interface. At the top left, there's a '리소스 그룹' icon. In the main content area, there are two windows. The top window is titled '리소스, 서비스 및 문서 검색(G+)' and shows a list of services like Azure Storage, App Services, and Azure Cosmos DB. The bottom window is titled '리소스 그룹' and shows a list of resources under the 'AIS\_ML' resource group, including 'AIS\_MachineLearning'. A red box labeled '1' is around the '리소스 그룹' button in the main bar. A red box labeled '2' is around the 'AIS\_ML' resource group in the list. A red box labeled '3' is around the 'AIS\_MachineLearning' workspace within the resource group.

리소스 그룹 → AIS\_ML → AIS\_MachineLearning

## 리소스 그룹 재접속

- <리소스 그룹> 클릭
- 리소스 그룹 목록 확인
  - AIS\_ML 선택
- AIS\_ML 목록 확인
  - 추가된  
AIS\_MachineLearning 선택

# 08 | Azure ML Studio 접속



## AIS-MachineLearning 화면

1. <Studio 시작하기>를 클릭

## Azure Machine Learning Studio 화면

2. 자동화된 ML (Automated ML) 선택

# 09 | Automated ML 접속

**Automated ML 접속**

1. 새 자동화된 ML 실행

2. 최근 자동화된 ML 실행

3. 설명서

| 실행   | 실행 ID                                 | 실험                | 상태  |
|------|---------------------------------------|-------------------|-----|
| 실행 1 | AutoML_3cb8a9b3-e0a5-4322-aba1-ee...  | fraudtest         | 완료됨 |
| 실행 1 | AutoML_783e0db0-b6d5-4575-ad15-78...  | biketest          | 완료됨 |
| 실행 1 | AutoML_eb8fe350-1ce7-4c1c-8668-437... | automl_regression | 완료됨 |

## Automated ML 접속

1. Automated ML 새로 만들기
2. 수행했던 실험 확인할 수 있음
3. 설명서
  - Automated ML관련 컨셉 이해, 자습서, 블로그 등을 확인 할 수 있음

# 10 | 데이터 세트 만들기

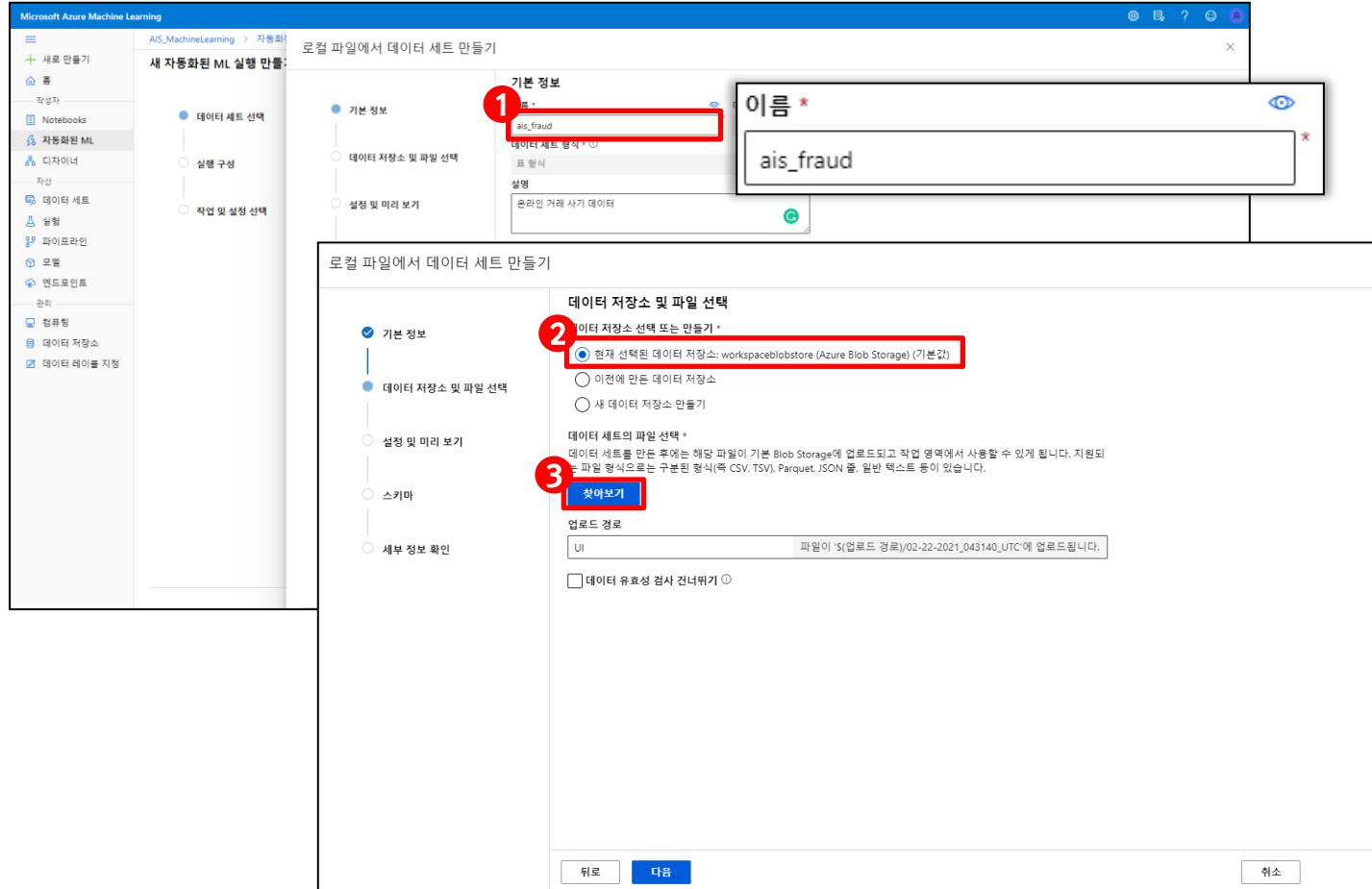
The screenshot shows the Azure Machine Learning Studio interface. On the left, there's a sidebar with various options like '새로 만들기', '자동화된 ML', etc. The main area is titled '새 자동화된 ML 실행 만들기'. It shows a '데이터 세트 선택' (Data Set Selection) step. A dropdown menu is open, with a red box highlighting the '데이터 세트 만들기' (Create Data Set) option. Below it, another red box highlights the '로컬 파일에서' (From Local File) option. To the right, there's a table of existing data sets and a search bar.

| 데이터 세트 형식 | 만든 날짜:               | 수정한 시간               |
|-----------|----------------------|----------------------|
| 표 형식      | 2021년 2월 16일 오후 5:13 | 2021년 2월 16일 오후 5:13 |
| 표 형식      | 2021년 2월 16일 오전 9:34 | 2021년 2월 16일 오전 9:34 |
| 표 형식      | 2021년 2월 8일 오후 4:16  | 2021년 2월 8일 오후 4:16  |

## 데이터 세트 만들기

- **로컬 파일**
    - 로컬환경에서 데이터 선택
  - **데이터 저장소**
    - 기존 구축된 저장소 활용
  - **웹 파일**
    - Directory URL 연결
    - 온라인 데이터를 가져오기
  - **열린 데이터 세트**
    - 무료 제공 데이터
- 이번 실습에 로컬 파일에서 데이터 불러오기

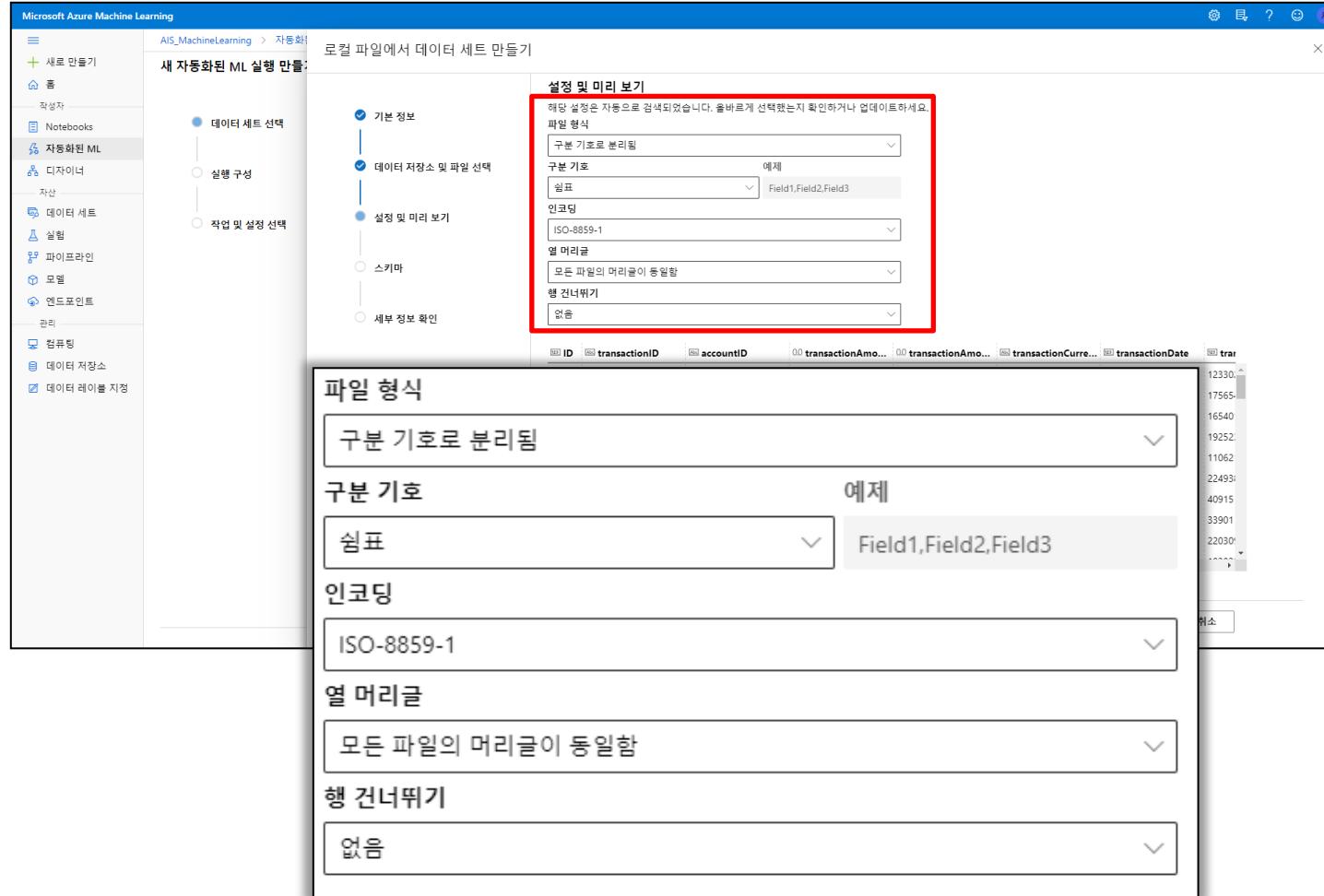
# 11 | 데이터 저장소 및 파일 선택



## 데이터 저장소 및 파일 선택

- 데이터 기본 정보 작성**
  - 데이터 세트 이름  
➤ **ais\_fraud**
  - 데이터 형식  
➤ **표 형식**
  - 데이터 설명
- 데이터 저장소 선택**
  - 현재 선택된 데이터 저장소: workspaceblobstore (Azure Blob Storage) (기본값)**
- 데이터 세트 파일 선택**
  - <찾아보기> 버튼을 클릭하여 **"Fraud.csv"** 파일을 불러오기

# 12 | 데이터 세트 설정



## 데이터 세트 설정

- 파일 형식 선택
  - 구분 기호로 분리됨
- 구분 기호 선택
  - 쉼표
- 인코딩
  - ISO-8859-1
- 열 머리글
  - 모든 파일의 머리글이 동일함
- 행 건너뛰기
  - 없음
- 미리보기
  - 데이터 맞게 들어오는지 확인

# 13 | 스키마

| 포함                                  | 열 이름          | 속성                    | 유형  | 형식 설정 및 예제  |
|-------------------------------------|---------------|-----------------------|-----|---|
| <input checked="" type="checkbox"/> | Path          | 선택한 유형에 적용할 수 없습니다... | 문자열 | dbo.balanced_transactionstest.csv/part-00000-tid-27123... |
| <input checked="" type="checkbox"/> | transactionID | 선택한 유형에 적용할 수 없습니다... | 문자열 | 206EC255-09E1-4601-9B2D-D2C0B668ED3F, E37A10BB-6...       |
| <input checked="" type="checkbox"/> | accountID     | 선택한 유형에 적용할 수 없습니다... | 문자열 | A914800878614266, A985157013380235, A84442812012...       |

| 포함                                  | 열 이름          | 속성                    | 유형  | 형식 설정 및 예제  |
|-------------------------------------|---------------|-----------------------|-----|---|
| <input checked="" type="checkbox"/> | Path          | 선택한 유형에 적용할 수 없습니다... | 문자열 | dbo.balanced_transactionstest.csv/part-00000-tid-27123... |
| <input checked="" type="checkbox"/> | transactionID | 선택한 유형에 적용할 수 없습니다... | 문자열 | 206EC255-09E1-4601-9B2D-D2C0B668ED3F, E37A10BB-6...       |
| <input checked="" type="checkbox"/> | accountID     | 선택한 유형에 적용할 수 없습니다... | 문자열 | A914800878614266, A985157013380235, A84442812012...       |

## 스키마

- 각 열의 데이터 유형을 확인 및 변경
- 분석에 필요 없는 데이터가 있으면 제외할 수 있음
- 데이터 유형은 사용자 지정 없이 자동으로 인식하고 설정됨

# 14 | 세부 정보 확인

로컬 파일에서 데이터 세트 만들기

세부 정보 확인

|                     |  |
|---------------------|--|
| 기본 정보               | 데이터 저장소 및 파일 선택                          |
| 이름<br>ais_fraud     | 데이터 저장소<br>workspaceblobstore            |
| 데이터 세트 버전<br>1      | 선택한 파일(1)<br>Fraud.csv                   |
| 데이터 세트 형식<br>표 형식   | 경로<br>UI/02-22-2021_050237_UTC/Fraud.csv |
| 설명<br>온라인 거래 사기 데이터 |  |

파일 설정

|                          |
|--------------------------|
| 파일 형식<br>구분 기호로 분리됨      |
| 구분 기호<br>쉼표              |
| 인코딩<br>ISO-8859-1        |
| 열 머리글<br>모든 파일의 머리글이 동일함 |
| 행 건너뛰기<br>없음             |

만든 후 이 데이터 세트 프로파일링

[뒤로](#) [만들기](#) [취소](#)

## 데이터 세부 정보 확인

- 생성된 데이터 세트 정보 확인 후 <만들기> 버튼을 클릭하여 데이터 세트를 생성

# 15 | 데이터 정보 및 통계 확인

The screenshot shows the 'Dataset Selection' step in the 'Automated ML' workflow. The left sidebar shows the navigation path: AI\_MachineLearning > 자동화된 ML > 실행 시작. The main area displays a list of datasets:

| 데이터 세트 이름   | 데이터 세트 형식 | 만든 날짜:               | 수정한 시간               |
|---|-----------|----------------------|----------------------|
| <input checked="" type="radio"/> <b>ais_fraud</b>       | 표 형식      | 2021년 2월 22일 오후 2:39 | 2021년 2월 22일 오후 2:39 |
| <input type="radio"/> <b>fraud</b>                      | 표 형식      | 2021년 2월 16일 오후 5:13 | 2021년 2월 16일 오후 5:13 |
| <input type="radio"/> <b>bikeshare</b>                  | 표 형식      | 2021년 2월 16일 오전 9:34 | 2021년 2월 16일 오전 9:34 |
| <input type="radio"/> <b>training-formatted-dataset</b> | 표 형식      | 2021년 2월 8일 오후 4:16  | 2021년 2월 8일 오후 4:16  |

A large callout box highlights the 'ais\_fraud' dataset entry.

## 데이터 세트 정보 확인

- 생성된 데이터 세트 클릭

# 16 | 데이터 정보 및 통계 확인

**데이터 미리 보기**

| transactionID      | accountID    | transactionAmountUSD | transactionAmount | transactionCurrencyCode | transactionDate | transactionTime | localHour | transactionIPaddress | ipState  | ipP... |
|--------------------|--------------|----------------------|-------------------|-------------------------|-----------------|-----------------|-----------|----------------------|----------|--------|
| 206EC255-09E1-4... | A91480087... | 139.99               | 139.99            | USD                     | 20130830        | 123302          | 7         | 129.20               | maryland | 207... |
| E37A10B8-6B38-...  | A98515701... | 84.99                | 84.99             | USD                     | 20130523        | 175654          | -99       | 216.15               | new york | 145... |
| 5C40AE8-98BC-...   | A84442812... | 85.79                | 85.79             | USD                     | 20130903        | 165401          | 11        | 71.72                | ohio     | 452... |

**데이터 통계**

참고: 데이터 샘플을 기반으로 통계가 계산됨

| 열                       | 프로필            | 유형   | 최소             | 최대            | 개수   | 누락... | 빈 개수 | 오류 수 | 평균값         | 표준 편차  | 분산        |
|-------------------------|----------------|------|----------------|---------------|------|-------|------|------|-------------|--------|-----------|
| transactionID           | 그래프 데이터가 없습니다. | 문자열  | 0031F5FA-94... | FF74DFD9-5... | 2302 | 0     | 0    | 0    | 0.00        | 0.00   | 0.00      |
| accountID               | 그래프 데이터가 없습니다. | 문자열  | A105552042...  | A985157127... | 2302 | 0     | 0    | 0    | 0.00        | 0.00   | 0.00      |
| transactionAmountUSD    |                | 10진수 | 0.00           | 10066.59      | 2302 | 0     | 0    | 0    | 281.92      | 446.66 | 199507.51 |
| transactionCurrencyCode |                | 문자열  | USD            | CAD           | ZAR  | 2302  | 0    | 0    | 0.00        | 0.00   | 0.00      |
| transactionDate         |                | 정수   | 20130401       | 20130930      | 2302 | 0     | 0    | 0    | 20130650.13 | 167.17 | 27944.35  |

## 데이터 세트 정보 확인

1. 데이터 미리 보기 확인
  - 사용할 데이터 정보 확인
2. 데이터 통계 확인
  - 데이터 세트관련 통계 확인 가능

# 17 | 데이터 세트 선택

The screenshot shows the 'Dataset Selection' step in the 'Automated ML' workflow. The 'ais\_fraud' dataset is selected and highlighted with a red box. The table below lists four datasets:

| 데이터 세트 이름  | 데이터 세트 형식 | 만든 날짜:               | 수정한 시간               |
|--|-----------|----------------------|----------------------|
| <input checked="" type="radio"/> ais_fraud       | 표 형식      | 2021년 2월 22일 오후 2:39 | 2021년 2월 22일 오후 2:39 |
| <input type="radio"/> fraud                      | 표 형식      | 2021년 2월 16일 오후 5:13 | 2021년 2월 16일 오후 5:13 |
| <input type="radio"/> bikeshare                  | 표 형식      | 2021년 2월 16일 오전 9:34 | 2021년 2월 16일 오전 9:34 |
| <input type="radio"/> training-formatted-dataset | 표 형식      | 2021년 2월 8일 오후 4:16  | 2021년 2월 8일 오후 4:16  |

## 데이터 세트 선택

- 데이터 목록에서 생성된 데이터 세트를 선택

# 18 | 실험 구성

Microsoft Azure Machine Learning

AIS\_MachineLearning > 자동화된 ML > 실행 시작

새 자동화된 ML 실행 만들기

데이터 세트 선택     실행 구성     작업 및 설정 선택

실행 구성  
기존 실험 중에서 선택하거나 새 실험을 만든 후 대상 열 및 학습 컴퓨팅을 선택합니다. [실험을 구현하세요.](#)

데이터 세트  
ais fraud (데이터 세트 보기)

실험 이름 \*  
 기존 항목 선택     새로 만들기  
 새 실험 이름  
fraudtest

대상 열 \*  
label

컴퓨팅 클러스터 선택 \*  
 컴퓨팅을 선택하세요.     새 컴퓨팅 만들기     컴퓨팅 새로 고침

다음

## 실험 구성

- 새로 만들기를 클릭하여 실험 이름 입력  
➤ **fraudtest**
- 데이터 분석 대상 선택  
➤ **label**
- 학습 컴퓨팅 생성 및 선택
  - 분석 과제에 따라 적합한 학습 컴퓨팅 선택
- "새 컴퓨팅 만들기" 클릭하여 새 학습 컴퓨팅을 생성

# 19 | 새 컴퓨팅 만들기 (1/2)

컴퓨팅 클러스터 만들기 ①

가상 머신  설정

가상 머신 선택  
컴퓨팅 클러스터에 사용할 가상 머신 크기를 선택합니다.

위치 ①  
koreacentral

가상 머신 우선 순위 ①  
 전용  낮은 우선 순위

가상 머신 유형 ①  
 CPU  GPU

가상 머신 크기 ①  
 원장 옵션에서 선택  모든 옵션에서 선택

사용 가능한 총 할당량: 코어 20개 ①

| 이름   | 범주      | 워크로드 유형                                      | 사용 가능... ① | 비용 ①       |
|--|---------|--|------------|------------|
| <input type="radio"/> Standard_DS2_v2<br>코어 2개, 7GB RAM, 14GB 스토리지             | 범용      | Notebook(또는 다른 IDE)에서 개발 및 경량 테스트            | 코어 20개     | \$0.16/... |
| <input checked="" type="radio"/> Standard_DS3_v2<br>코어 4개, 14GB RAM, 28GB 스토리지 | 범용      | 클래식 ML 모델 학습, AutoML 실행, 파이프라인 실행(기본 컴퓨팅)    | 코어 20개     | \$0.33/... |
| <input type="radio"/> Standard_DS12_v2<br>코어 4개, 28GB RAM, 56GB 스토리지           | 메모리 최적화 | 대규모 데이터 세트(1GB 초과) 병렬 실행 단계, 일괄 처리 추론에 대한 학습 | 코어 20개     | \$0.40/... |
| <input type="radio"/> Standard_F4s_v2<br>코어 4개, 8GB RAM, 32GB 스토리지             | 컴퓨팅 최적화 | 실시간 추론 및 기타 대기 시간이 중요한 작업                    | 코어 20개     | \$0.19/... |

뒤로  자동화에 대한 템플릿 다운로드 취소

## 가상 머신 선택

1. 위치  
➤ 한국중부 koreacentral
2. 가상 머신 우선 순위  
➤ 전용
3. 가상 머신 유형  
➤ CPU
4. 가상 머신 크기  
➤ 권장 옵션에서 선택 (체크)  
➤ Standard\_DS3\_v2  
  
➤ 최소 노드 수: 0  
➤ 최대 노드 수: 6
5. 규모 축소 전 유회 시간(초)  
➤ 120

# 20 | 새 컴퓨팅 만들기 (2/2)

컴퓨팅 클러스터 만들기 ⓘ

설정 구성

선택한 가상 머신 크기의 컴퓨팅 클러스터 설정을 구성합니다.

| 이름              | 병주 | 코어 | 사용 가능한 할당량 | RAM   | 스토리지  | 비용/노드     |
|-----------------|----|----|------------|-------|-------|-----------|
| Standard_DS3_v2 | 병용 | 4  | 코어 20개     | 14 GB | 28 GB | \$0.33/시간 |

컴퓨팅 이름 \* ⓘ

최소 노드 수 \* ⓘ

최대 노드 수 \*

규모 축소 전 유휴 시간(초) \* ⓘ

SSH 액세스 사용 ⓘ

[고급 설정](#)

[뒤로](#) [만들기](#) [자동화에 대한 템플릿 다운로드](#) [취소](#)

## 설정 구성

1. 컴퓨팅 이름  
➤ aml-compute
2. 노드 수  
➤ 최소 노드 수: 0  
➤ 최대 노드 수: 4
3. 규모 축소 전 유휴 시간(초)  
➤ 120

# 21 | 작업 형식 선택

1. [추가 구성 설정 보기](#)

2. [기능화 설정 보기](#)

**작업 형식 선택**

실험의 기계 학습 작업 유형을 선택합니다. 실험을 미세 조정하려면 추가 구성 또는 기능화 설정을 선택합니다.

**분류**  
대상 열에서 예/아니요, 파랑, 빨강, 녹색 등의 여러 범주 중 하나를 예측하려는 경우

딥 러닝 사용

**회귀**  
연속 숫자 값을 예측하려는 경우

**시계열 예측**  
시간을 기준으로 값을 예측하려는 경우

**추가 구성**

주 메트릭: 정확도  
 최적 모델 설명

차단된 알고리즘: 자동화된 ML에서 학습 중 사용하지 않을 알고리즘 목록입니다.

종료 기준: 학습 작업 시간(시간): 3  
메트릭 정수 임계값: 메트릭 정수 임계값

유효성 검사: 유효성 검사 유형: 자동

동시성: 최대 동시 반복 횟수: 4

## 작업 형식 선택

- 분류 ✓
- 회귀
- 시계열 예측

## 작업 추가 설정

1. “추가 구성 설정 보기” 클릭
2. 추가 설정 입력 화면

# 22 | 작업 추가 설정

**주 메트릭** ⓘ  
AUC 가중치

**최적 모델 설명** ⓘ

**차단된 알고리즘** ⓘ  
자동화된 ML에서 학습 중 사용하지 않을 알고리즘 목록입니다.

**종료 기준**  
학습 작업 시간(시간) ⓘ 1  
메트릭 점수 임계값 ⓘ

**유효성 검사**  
유효성 검사 유형 ⓘ 학습-유효성 검사 분할  
데이터 유효성 검사 백분율 \* ⓘ 30  
자동화된 ML은 유효성 검사를 위해 10~30%의 데이터를 유지하는 것이 좋습니다.

**동시성**  
최대 동시 반복 횟수 ⓘ 4

**차단된 알고리즘** ⓘ  
자동화된 ML에서 학습 중 사용하지 않을 알고리즘 목록입니다.

- LogisticRegression
- SGD
- MultinomialNaiveBayes
- BernoulliNaiveBayes
- SVM
- LinearSVM
- KNN
- DecisionTree
- RandomForest

## 분류관련 추가 설정

1. 주 메트릭 선택
  - 정확도
    - AUC 가중치
    - 일반 매크로 회수
    - 평균 정밀도 점수 가중치
    - 정밀도 점수 가중치
2. 최적 모델 설명 체크
3. 차단된 알고리즘 선택 (없음)
4. 종료 기준 설정
  - 학습 작업 시간 (1)
  - 메트릭 점수 임계값 (없음)
5. 유효성 검사 유형 선택
  - Train-Validation Split
  - 데이터 유효성 검사 백분율 (30)
6. 동시성 설정
  - 최대 동시 반복 횟수 (6)

- ❖ 학습 작업 시간(분) – X
- ❖ 학습 작업 시간(시) – O

# 23 | 실험 결과 확인 (1/13)

Microsoft Azure Machine Learning

AI\_MachineLearning > 자동화된 ML

**자동화된 ML**

자동화된 ML을 학습시켜 코드를 작성 없이도 데이터를 기반으로 최상의 모델을 찾을 수 있습니다. 자동화된 ML에 대해 자세히 알아보기

| 실행   | 실행 ID                                 | 실험                | 상태  | 제출된 시간               | 기간      | 제출한 사람 | 컴퓨팅 대상         | 태그 |
|------|---------------------------------------|-------------------|-----|----------------------|---------|--------|----------------|----|
| 실행 1 | AutoML_3cb8a9b3-e0a5-4322-aba1-ee...  | fraudtest         | 완료됨 | 2021년 2월 16일 오후 5:15 | 27분 51초 | 윌리엄    | aml-compute... |    |
| 실행 1 | AutoML_783e0db0-b6d5-4575-ad15-78...  | biketest          | 완료됨 | 2021년 2월 16일 오전 9:40 | 22분 34초 | 윌리엄    | aml-compute... |    |
| 실행 1 | AutoML_eb8fe350-1ce7-4c1c-8668-437... | automl_regression |     |                      |         |        |                |    |

**설명서**

- 개념: 자동화된 ML이란 무엇인가요?
- 자습서: 자동화된 ML을 사용하여 첫 번째 분류 모델 만들기
- 블로그: 자동화된 ML의 새로운 기능을 사용하여 더욱 정확한 예측 결과 만들기

## 실험 결과 확인

- <자동화된 ML>탭을 클릭
- 실행 이름에 따라 실행 ID를 클릭하여 실험 결과 확인 가능

# 24 | 실험 결과 확인 (2/13)

**최적의 모델 요약**

알고리즘 이름  
VotingEnsemble

AUC 가중치  
0.84435 [다른 모든 메트릭 보기](#) 5

샘플링  
100.00 % ⓘ

등록된 모델  
등록한 모델 없음

배포 상태  
아직 배포 없음

**실행 요약**

작업 유형  
분류 [모든 실행 설정 보기](#)

주 메트릭  
AUC 가중치

실행 이름  
fraudtest

**설명**

( ⓘ 설명을 추가하려면 편집 아이콘을 클릭하세요.)

## 실험 결과 확인

3. <세부 정보> 탭을 클릭
4. 최적 모델 정보 확인
  - 알고리즘 이름
  - 선택한 메트릭 값
5. <다른 모든 메트릭 보기>를 클릭하여 다른 측정 메트릭을 확인 가능

## 결과 해석

- 측정 메트릭은 AUC 가중치이며 84.4%의 예측 정확도를 보였음
- 이에 따라 다른 모델을 비해서 Voting Ensemble은 성능 좋은 모델(챔피언 모델)임을 확인할 수 있음
- AUC 평가 기준
  - 0.90-1 = excellent
  - 0.80-0.90 = good
  - 0.70-0.80 = fair
  - 0.60-0.70 = poor
  - 0.50-0.60 = fail

# 25 | 실험 결과 확인 (3/13)

The screenshot shows the Microsoft Azure Machine Learning interface. On the left, there's a sidebar with various options like '새로 만들기', '노트북', '자동화된 ML', etc. The main area shows an experiment named 'fraudtest' with a status of '완료됨'. The '세부 정보' tab is selected, displaying details such as '속성' (Status: 완료됨), '만들어짐' (Created: 2021년 2월 16일 오후 5:15), '시작됨' (Started: 2021년 2월 16일 오후 5:15), and '기간' (Duration: 27분 51.49초). Below this, there are sections for '데이터 세트' (Data Sets) and '설명' (Description). The '자동화된 ML' tab is also visible.

**Metrics**

| 메트릭            | 값       |
|----------------|---------|
| 정확도            | 0.75942 |
| AUC 매크로        | 0.84435 |
| AUC 마이크로       | 0.84610 |
| AUC 가중치        | 0.84435 |
| 평균 정밀도 점수 매크로  | 0.84645 |
| 평균 정밀도 점수 마이크로 | 0.84935 |
| 평균 정밀도 점수 가중치  | 0.84662 |
| 균형 정확도         | 0.75950 |
| F1 점수 매크로      | 0.75942 |
| F1 점수 마이크로     | 0.75942 |
| F1 점수 가중치      | 0.75941 |
| 로그 손실          | 0.50011 |
| Matthews 상관 관계 | 0.51904 |

**Metrics 실행**

| 메트릭            | 값       |
|----------------|---------|
| 정확도            | 0.75942 |
| AUC 매크로        | 0.84435 |
| AUC 마이크로       | 0.84610 |
| AUC 가중치        | 0.84435 |
| 평균 정밀도 점수 매크로  | 0.84645 |
| 평균 정밀도 점수 마이크로 | 0.84935 |
| 평균 정밀도 점수 가중치  | 0.84662 |
| 균형 정확도         | 0.75950 |
| F1 점수 매크로      | 0.75942 |
| F1 점수 마이크로     | 0.75942 |
| F1 점수 가중치      | 0.75941 |
| 로그 손실          | 0.50011 |
| Matthews 상관 관계 | 0.51904 |
| 일반 매크로 회수      | 0.51901 |
| 정밀도 점수 매크로     |         |

## 실험 결과 확인

- AUC 가중치 외 다른 측정 메트릭을 확인 가능

# 26 | 실험 결과 확인 (4/13)

## 매크로 평균 (Macro Average) 및 마이크로 평균 (Micro Average)

| 반     | A  | B  | C  | D  |
|-------|----|----|----|----|
| 학생 수  | 50 | 5  | 2  | 3  |
| 평균 점수 | 50 | 60 | 80 | 85 |

$$\text{매크로 평균} = \frac{50 + 60 + 80 + 85}{4} = 68.75$$

$$\text{마이크로 평균} = \frac{(50 \times 50) + (5 \times 60) + (2 \times 80) + (3 \times 85)}{50 + 5 + 2 + 3} = 53.58$$

### 관련 설명

- 매크로 평균과 마이크로 평균의 차이가 크다
  - 매크로 평균은 A반에 학생이 몰려있어서 그 정보를 반영하지 못했기 때문이다.
- 여러 개의 카테고리로 이루어진 데이터에서 한 카테고리에 쓸려있는 데이터에서는 균형 잡힌 평균을 구하기 위하여 매크로 평균보다 마이크로 평균을 이용하는 것이 좋다.

# 27 | 실험 결과 확인 (5/13)

**최적의 모델 요약**

**알고리즘 이름** VotingEnsemble 7

**AUC 가중치** 0.84435 다른 모든 모델 보기

**샘플링** 100.00 % i

**등록된 모델**  
등록한 모델 없음

**배포 상태**  
아직 배포 없음

**세부 정보** 모델 설명(미리 보기) 메트릭 출력 + 로그 이미지 자식 실행 스냅샷

**모델 요약**

알고리즘 이름  
**VotingEnsemble**

AUC 가중치  
0.84435 다른 모든 메트릭 보기

샘플링  
100.00 % i

등록된 모델  
등록한 모델 없음

배포 상태  
아직 배포 없음

## 실험 결과 확인

- 모델 이름(VotingEnsemble)을 클릭하여 최적 모델의 자세 설명 및 시각화 정보를 확인 가능

# 28 | 실험 결과 확인 (6/13)

Microsoft Azure Machine Learning

실행 64 환료됨

8. 메트릭

9.

Confusion Matrix

|            |       | Predicted Label |    |
|------------|-------|-----------------|----|
| True Label | False | True            |    |
|            | False | 263             | 79 |
| True       | 87    | 261             |    |

ROC

True Positive Rate

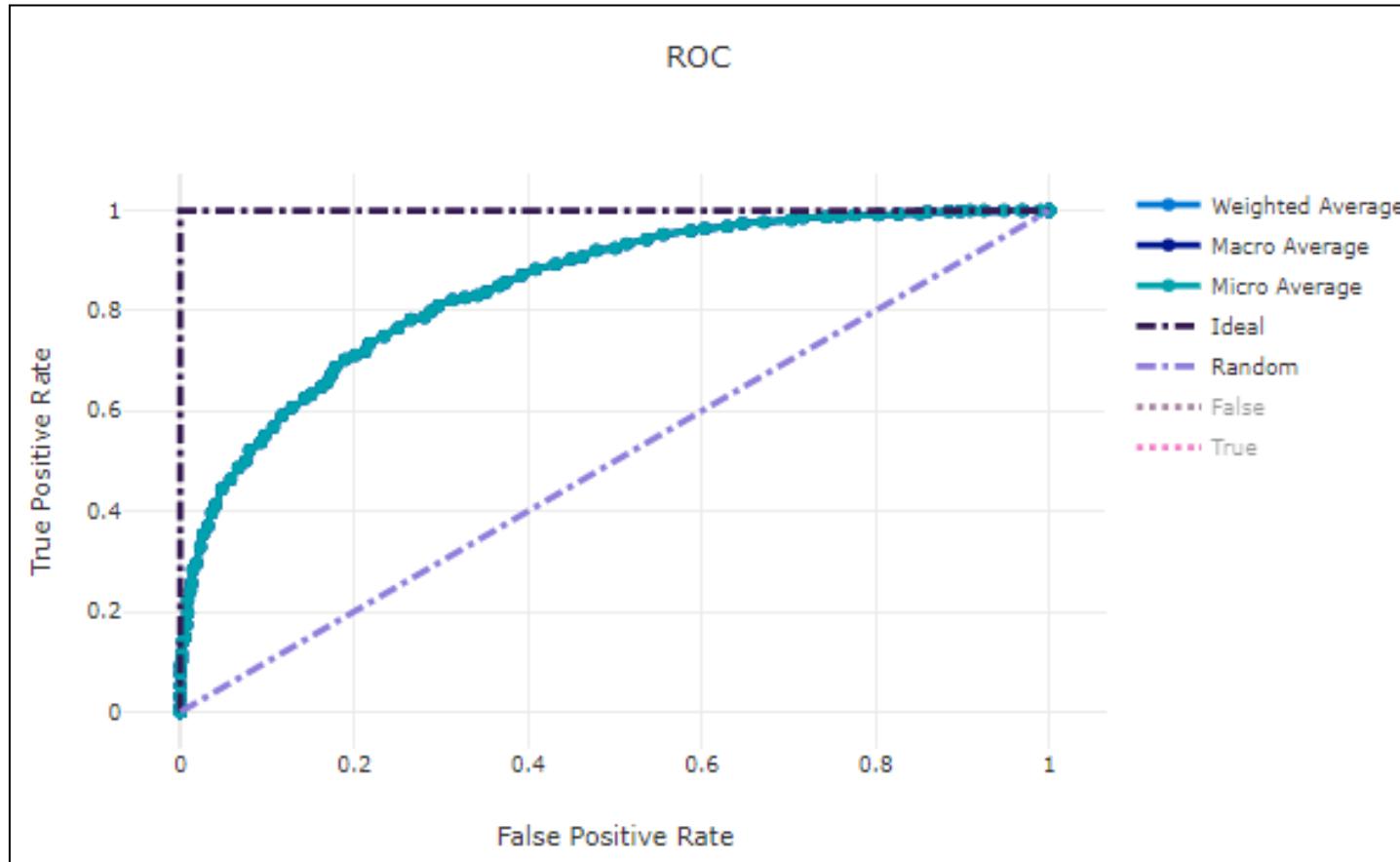
False Positive Rate

## 실험 결과 확인

8. <메트릭> 탭을 클릭

9. 다양한 시각화 형식을 확인할 수 있음

## 29 | 실험 결과 확인 (7/13)

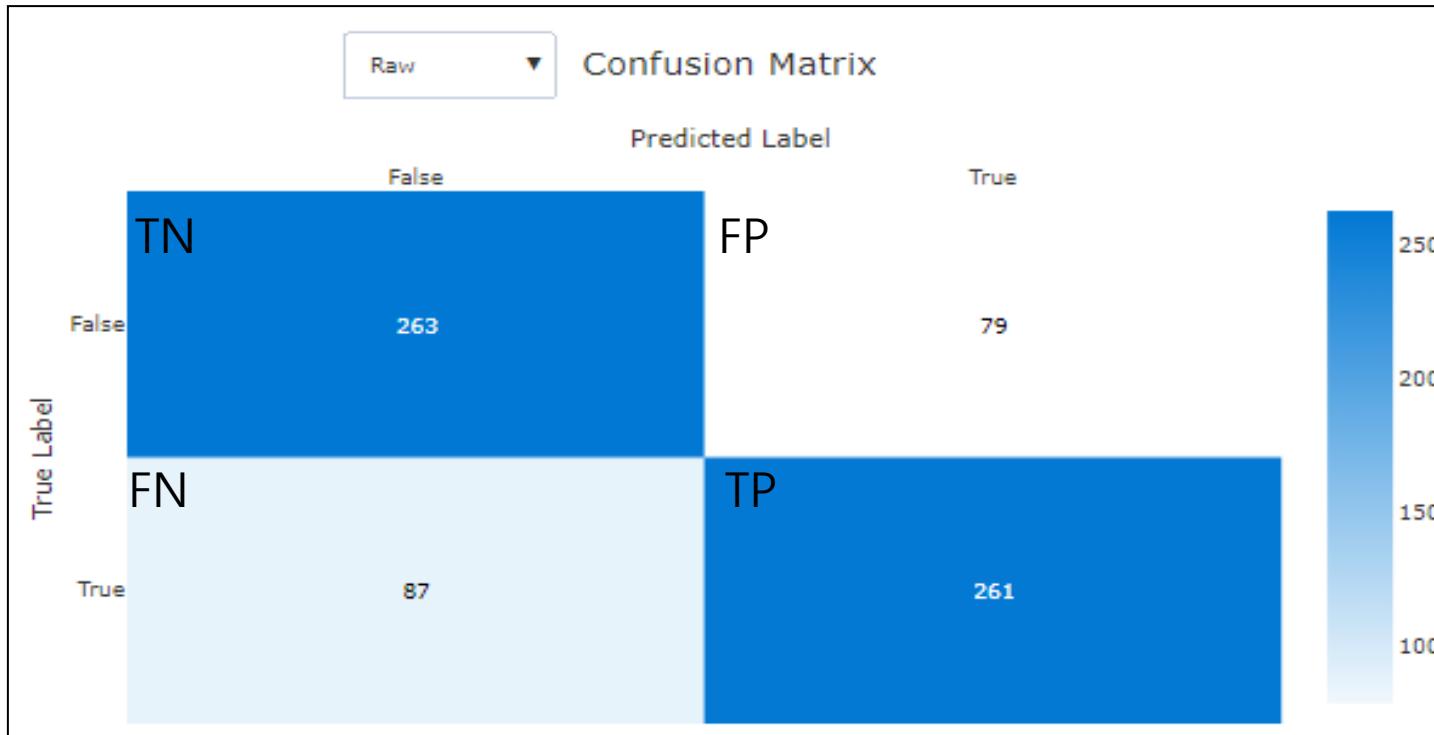


- ❖ TPR: True Positive Rate
- ❖ FPR: False Positive Rate

## ROC 곡선

- ROC Curve (Receiver Operating Characteristic Curve)
- ROC 곡선은 False Positive Rate (FPR) / True Positive Rate (TPR)의 곡선
- 값이 1에 가까울수록 모델의 성능이 좋음
- AUC (Area Under Curve)는 ROC 곡선 아래부분의 넓이로 확인할 수 있음
- AUC 면적이 넓을수록 모델의 성능이 좋음

# 30 | 실험 결과 확인 (8/13)



- Precision  

$$= \frac{TP}{TP+FP}$$
  

$$= \frac{261}{261+79}$$
  

$$= 76.8\%$$

- Recall  

$$= \frac{TP}{TP+FN}$$
  

$$= \frac{261}{261+87}$$
  

$$= 75.0\%$$

- Accuracy  

$$= \frac{(TP+TN)}{(TP+TN+FP+FN)}$$
  

$$= \frac{(261+263)}{690}$$
  

$$= 75.9\%$$

- F1-Score  

$$= \frac{2(Precision*Recall)}{(Precision+Recall)}$$
  

$$= \frac{2(0.768*0.75)}{(0.768+0.75)}$$
  

$$= 0.759$$

## Confusion Matrix

- 모델 예측 결과 10건을 예측 했을 때 7.5건의 정확도를 가짐
- 이에 기반으로 Precision(정밀도), Recall(재현율), Accuracy(정확성), F1-Score를 구할 수 있음
- 계산 결과에 따라 이 모델의 정확성은 75.9%로 보였음
- 이 모델에서 어떤 데이터가 사기일 것이라고 평가했을 때, 이 평가가 정확할 확률은 76.8%로 확인됨
- 이 모델에서는 모든 사기 데이터 중 75.0%가 정확하게 식별됨
- F1-Score가 높아야 성능이 좋다고 보면 되며, 이 모델의 F1-Score는 0.759이것으로 보였음

# 31 | 실험 결과 확인 (9/13)

설명(미리 보기)

설명자: mimic

+ 코호트 추가

설명 ID

ceffe5d6

600fc93f

설명(미리 보기)

설명(미리 보기) 탭을 클릭하면 모델의 예측 결과를 이해하는 데 도움이 되는 설명을 제공합니다.

| 설명 ID    | 데이터     | 설명자            | 기능      | 모델              | 만든 시간         |
|----------|---------|----------------|---------|-----------------|---------------|
| ceffe5d6 | tabular | mimic.lightgbm | 원시      | classificati... | Feb 16, 20... |
| 600fc93f | tabular | mimic.lightgbm | 엔지니어... | classificati... | Feb 16, 20... |

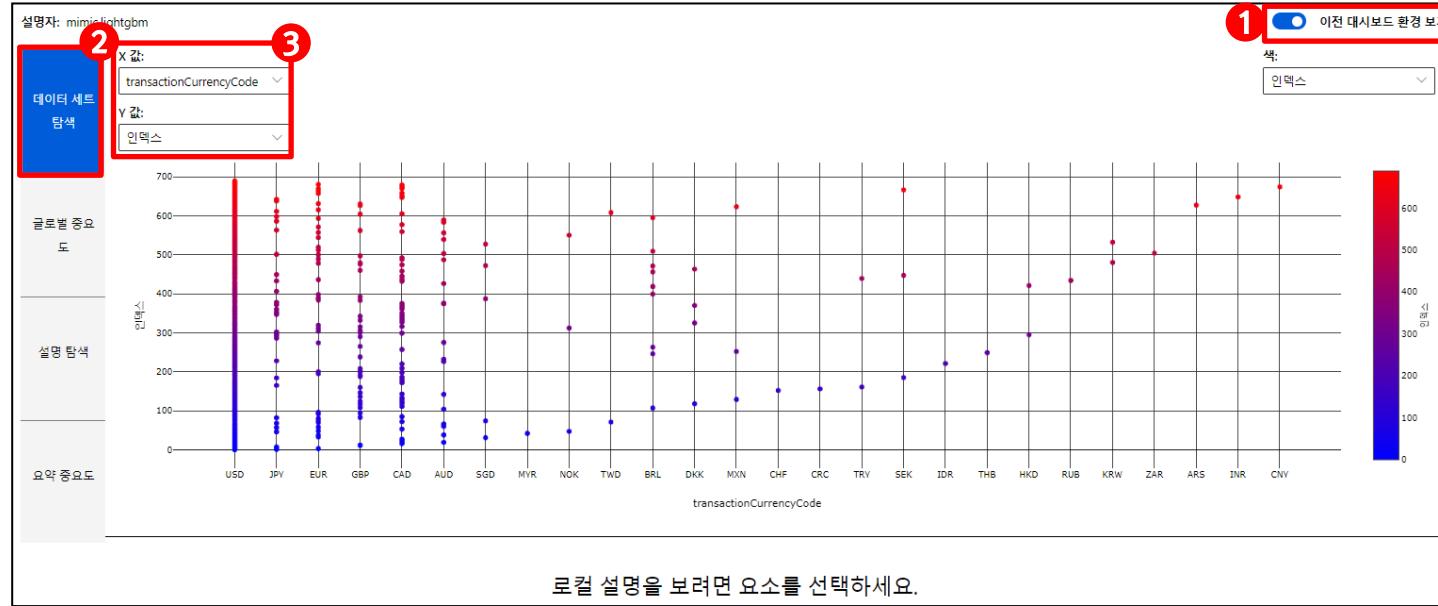
## 실험 결과 확인

- <설명(미리 보기)> 탭을 클릭
- 설명 ID의 확장 버튼을 클릭
- 확인하고 싶은 결과 기능을 선택 (원시/엔지니어링됨)

## 참고 링크:

- <https://docs.microsoft.com/en-us/azure/machine-learning/how-to-machine-learning-interpretability-aml>
- <https://docs.microsoft.com/en-us/azure/machine-learning/how-to-machine-learning-interpretability-automl>

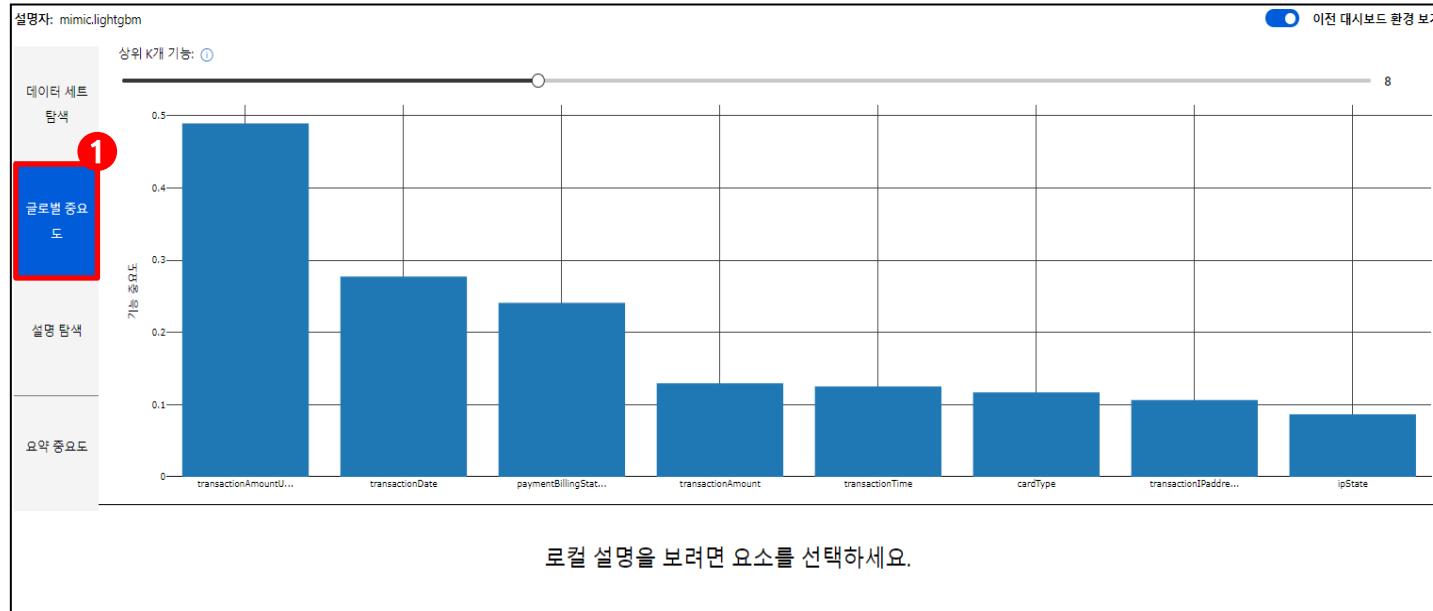
# 32 | 실험 결과 확인 (10/13)



## 실험 결과 확인

- <이전 대시보드 환경 보기> 옵션을 체크
- <데이터 세트 탐색> 버튼을 클릭
- X 값과 Y 값을 선택

# 33 | 실험 결과 확인 (11/13)



## 실험 결과 확인

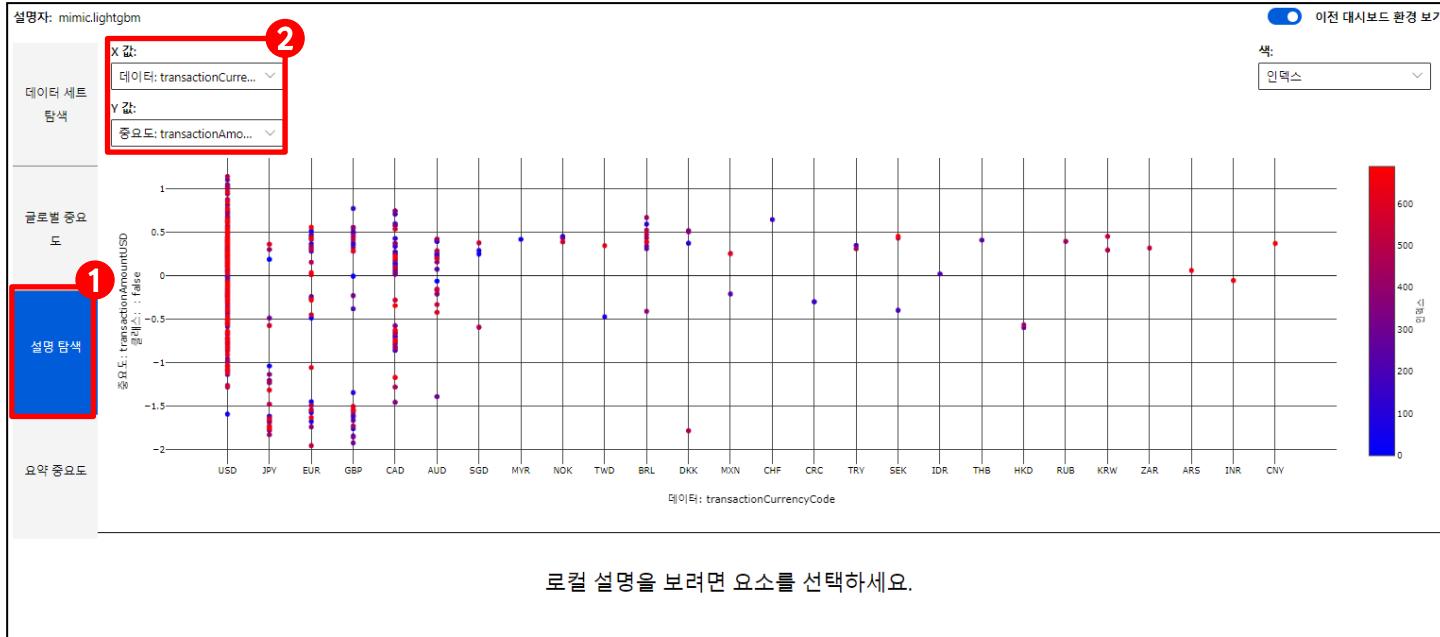
1. <글로벌 중요도> 버튼을 클릭
2. 모델에 직접적인 영향을 주는 변수들의 글로벌 중요도를 확인할 수 있음

## 글로벌 중요도

1. 최상위 K-개의 변수 중요도를 전역적으로 보여줌

2. 그림에 따라 transactionAmountUSD이라는 변수가 0.489의 글로벌 중요도로 보였으며 이 변수가 중요한 역할로 모델에 영향을 많이 미침

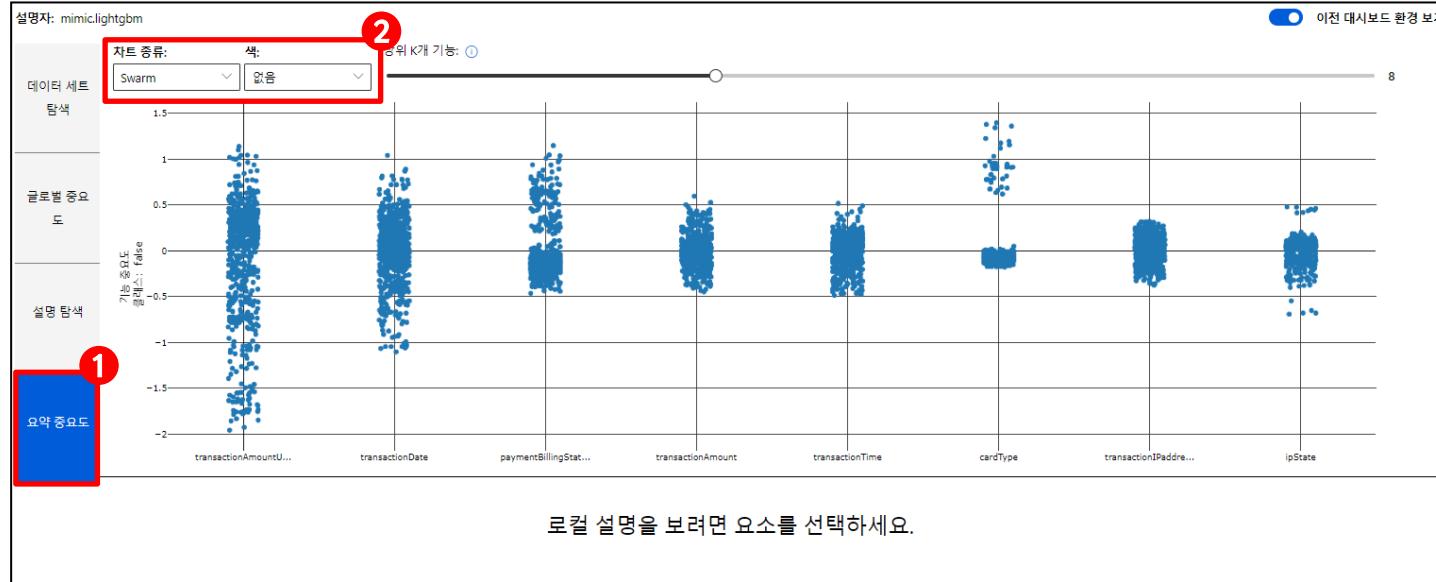
# 34 | 실험 결과 확인 (12/13)



실험 결과 확인

1. <설명 탐색> 버튼을 클릭
  2. X 값과 Y 값을 선택하여  
변수들의 중요도와 기준  
데이터 간의 관계를 확인할 수  
있음

# 35 | 실험 결과 확인 (13/13)



## 실험 결과 확인

1. <요약 중요도> 버튼을 클릭
2. 차트 종류 및 색깔을 선택하여 기능들의 중요도를 확인할 수 있음

## 요약 중요도

1. 모든 데이터 요소에서 로컬 변수 중요도 값을 사용하여 예측 값에 대한 각 변수 영향의 분포를 표시함.
2. 분포의 범위와 점의 밀도에 따라 transactionAmountUSD 변수가 이 모델에 가장 중요한 역할로 보였음

# 36 | 데이터 가드 레일

세부 정보 **데이터 가드 레일** 모델 출력 + 로그 자식 실행 스냅샷

자동 기능화를 사용하도록 설정하면 자동화된 ML에서 데이터 보호책이 실행됩니다. 데이터 보호책은 입력 데이터를 대상으로 수행하는 검사 시퀀스로서 모델을 학습시키는 데 고품질 데이터가 사용되는지를 확인합니다.

| 유형                   | 상태 | 설명  |
|----------------------|----|---|
| 클래스 균형 검색            | 통과 | 입력이 분석되었으며, 학습 데이터에서 모든 클래스가 균형 상태입니다.<br><a href="#">클래스 균형 데이터에 대해 자세히 알아보세요.</a> |
| 유형<br>기능 값 대체 누락     | 통과 | 학습 데이터에서 값이 누락된 기능이 감지되지 않았습니다.<br><a href="#">값 대체 누락에 대해 자세히 알아보세요.</a>           |
| 유형<br>높은 카디널리티 기능 검색 | 완료 | 높은 카디널리티 기능이 입력에서 검색되고 처리되었습니다.<br><a href="#">높은 카디널리티 기능 검색에 대해 자세히 알아보세요.</a>    |

[+ 추가 세부 정보 보기](#)

## 데이터 유효성 검증 / 데이터 가드 레일 (Data Guardrails)

1. 클래스 균형 검사 (Class Balancing Detection)
2. 결측값 대체/처리 (Missing Value Imputation)
3. 높은 카디널리티 변수 검사 (High Cardinality Feature Detection)

# 37 | 기타 모델 확인



| 세부 정보   | 데이터 가드 레일             | 모델        | 출력 + 로그  | 자식 실행 | 스냅샷                  |        |     |
|---|-----------------------|-----------|----------|-------|----------------------|--------|-----|
| <span>&gt; 배포</span> <span>다운로드</span> <span>🔍 모델 설명</span> |                       |           |          |       |                      |        |     |
| 알고리즘 이름   | 설명됨                   | AUC 가중치 ↓ | 샘플링 ⓘ    | 실행    | 만들어짐                 | 기간     | 상태  |
| VotingEnsemble  | <a href="#">설명 보기</a> | 0.84435   | 100.00 % | 실행 64 | 2021년 2월 16일 오후 5:41 | 1분 9초  | 완료됨 |
| StackEnsemble   |                       | 0.84041   | 100.00 % | 실행 65 | 2021년 2월 16일 오후 5:41 | 1분 29초 | 완료됨 |
| StandardScalerWrapper, XGBoostClassifier                    |                       | 0.82859   | 100.00 % | 실행 39 | 2021년 2월 16일 오후 5:34 | 50초    | 완료됨 |
| SparseNormalizer, XGBoostClassifier                         |                       | 0.82822   | 100.00 % | 실행 55 | 2021년 2월 16일 오후 5:38 | 57초    | 완료됨 |
| SparseNormalizer, XGBoostClassifier                         |                       | 0.82770   | 100.00 % | 실행 24 | 2021년 2월 16일 오후 5:30 | 56초    | 완료됨 |
| SparseNormalizer, LightGBM                                  |                       | 0.82375   | 100.00 % | 실행 47 | 2021년 2월 16일 오후 5:36 | 50초    | 완료됨 |
| StandardScalerWrapper, XGBoostClassifier                    |                       | 0.82361   | 100.00 % | 실행 54 | 2021년 2월 16일 오후 5:38 | 56초    | 완료됨 |
| SparseNormalizer, XGBoostClassifier                         |                       | 0.82225   | 100.00 % | 실행 56 | 2021년 2월 16일 오후 5:39 | 53초    | 완료됨 |
| StandardScalerWrapper, XGBoostClassifier                    |                       | 0.82191   | 100.00 % | 실행 30 | 2021년 2월 16일 오후 5:32 | 54초    | 완료됨 |
| StandardScalerWrapper, XGBoostClassifier                    |                       | 0.82189   | 100.00 % | 실행 43 | 2021년 2월 16일 오후 5:35 | 50초    | 완료됨 |
| StandardScalerWrapper, XGBoostClassifier                    |                       | 0.81919   | 100.00 % | 실행 57 | 2021년 2월 16일 오후 5:39 | 53초    | 완료됨 |

## 기타 머신러닝 모델 확인

최적 모델 외에는 기타 실행된 머신러닝 모델 및 해당 분석 결과를 확인 가능

# 38 | 모델 배포 (1/2)

The screenshot shows the Azure Machine Learning studio interface. On the left, there's a sidebar with various options like '새로 만들기' (Create New), '노트북' (Notebook), '자동화된 ML' (Automated ML), etc. The main area shows a pipeline named '실행 64' (Run 64) with a status of '완료됨' (Completed). A red circle labeled '1' highlights the '배포' (Deployment) button in the top navigation bar.

**모델 배포** (Model Deployment) dialog box is open on the right:

- 이름:** automl-model
- 설명:** 온라인 거래 사기 데이터 분류 모델
- 컴퓨팅 형식:** Azure Container Instance
- 인증 사용:** 키는 앤드포인트 세부 정보 페이지에서 확인할 수 있습니다.
- 사용자 지정 배포 자산 사용:** 사용자 지정 배포 자산 사용 (unchecked)

A red box labeled '2' surrounds the dialog box. A red box labeled '3' surrounds the '배포' (Deploy) button at the bottom right of the dialog.

## 모델 배포

1. <배포> 탭을 클릭
2. 모델 배포 정보 입력
  - 모델이름
  - 모델관련 설명
  - 컴퓨팅 형식 선택
    - Azure Kubernetes Service(AKS)
    - **Azure Container Instances(ACI)** ✓
    - “인증 사용” 체크
3. <배포>버튼을 클릭하여 배포 작업을 진행

# 39 | 모델 배포 (2/2)

The screenshot shows the Azure Machine Learning Studio interface. On the left, there's a sidebar with various options like '새로 만들기', '작성자', 'Notebooks', '자동화된 ML', etc. A red box labeled '4' highlights the '엔드포인트' (Endpoints) option in the sidebar. The main area shows a list of endpoints under the '엔드포인트' tab. One endpoint is selected, highlighted by a red box labeled '5'. Below the list, a modal window titled '세부 정보' (Detailed Information) is open for the selected endpoint. It shows details like '서비스 ID: automl-model', '설명: --', '배포 상태: Healthy', '컴퓨팅 형식: ACI', and '모델 ID: AutoML\_3cb8a9b3-e0a5-4322-aba1-ee99026929fd\_59'. A red box labeled '6' highlights the 'REST 엔드포인트' (REST Endpoint) section, which displays the URL: `http://3a0f0e62-b670-4410-82d0-da51bad3b3a3.koreacentral.azurecontainer.io/score`. At the bottom of the modal, another red box labeled '6' highlights the 'REST 엔드포인트' section again, showing the same URL.

## 모델 배포

4. <엔드포인트> 탭을 클릭
5. 배포된 모델을 확인하여 클릭
6. <세부 정보> 탭에서 REST 엔드포인트에서 scoring uri를 확인할 수 있음

# 40 | 배포된 모델 테스트

automl-model

**1** 세부 정보 **테스트**

실시간 엔드포인트를 테스트하려면 데이터를 입력하세요. **테스트**

**data**

|                         |               |
|-------------------------|---------------|
| transactionID           | example_value |
| accountID               | example_value |
| transactionAmountUSD    |               |
| transactionAmount       |               |
| transactionCurrencyCode | example_value |
| transactionDate         |               |
| transactionTime         |               |
| localHour               |               |
| transactionIPaddress    |               |

automl-model

**1** 세부 정보 **테스트** 사용 배포 로그

실시간 엔드포인트를 테스트하려면 데이터를 입력하세요. **테스트**

**data**

|                         |                                      |
|-------------------------|--------------------------------------|
| transactionID           | 40A1F735-520A-4885-9273-AFA2CB8DE4FF |
| accountID               | A1688853516417580                    |
| transactionAmountUSD    | 798.86022                            |
| transactionAmount       | 6198                                 |
| transactionCurrencyCode | HKD                                  |
| transactionDate         | 20130611                             |
| transactionTime         | 40813                                |
| localHour               | 13                                   |
| transactionIPaddress    | 153.129                              |

automl-model

세부 정보 **테스트** 사용 배포 로그

실시간 엔드포인트를 테스트하려면 데이터를 입력하세요. **테스트**

**data**

아래에 쉼표로 구분된 값을 입력하세요.

```
transactionID,accountID,transactionAmountUSD,transactionAmount,transactionCurrencyCode,transactionDate,transactionTime,localHour,transactionIPaddress,ipState,ipPostcode,ipCountryCode,isProxy,IP,browserLanguage,paymentInstrumentType,cardType,paymentBillingPostalCode,paymentBillingState,paymentBillingCountryCode,cvv,VerifyResult,digitalItemCount,physicalItemCount
40A1F735-520A-4885-9273-AFA2CB8DE4FF,A1688853516417580,798.86022,6198,HKD,20130611,153.129,tokyo,100-0000.jp,False,zh-HK,CREDITCARD,MC,NA,NA,HK,M,2,1
```

**4** 테스트 결과

```
{
  "result": [
    true
  ]
}
```

## 모델 테스트

- <테스트> 탭을 클릭
- 테스트할 데이터를 박스형식이나 CSV형식으로 입력
- <테스트> 버튼을 클릭하여 테스트 실행
- 테스트 결과를 확인

# 41 | 모델 호출 (1/5)

**automl-model**

세부 정보 테이블 사용 배포 로그

**기본 사용 정보**

REST 엔드포인트  
[ http://3a0fe62-b670-4410-82d0-da51bad3b3a3.koreacentral.azurecontainer.io/score ]

인증 유형  
 키 사용  토큰 사용  
 기본 키  
 3RUM01ROCLKLw7a10uahWE1Bjh0ynMut [다시 생성](#)  
 보조 키  
 UVSaqQUMAwrmRlYAgGxh2ropMYWDUdfW [다시 생성](#)

**사용 옵션**

소비 유형  
 C#  Python  R

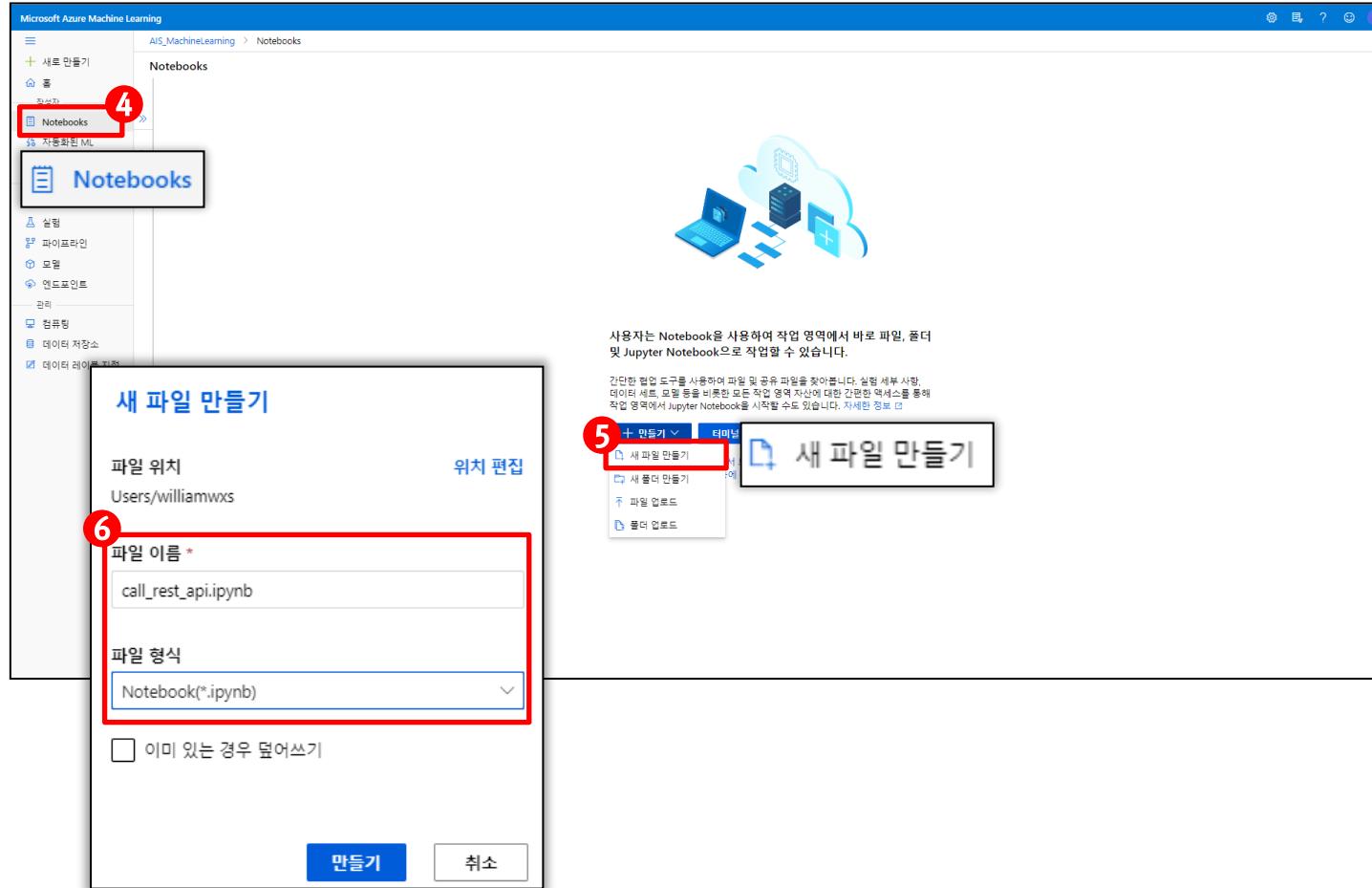
```

1 import urllib.request
2 import json
3 import os
4 import ssl
5
6 def allowSelfSignedHttps(allowed):
7     # bypass the server certificate verification on client side
8     if allowed and not os.environ.get('PYTHONHTTPSVERIFY', '') and getattr(ssl, '_create_unverified_context', None):
9         ssl._create_default_https_context = ssl._create_unverified_context
10
11 allowSelfSignedHttps(True) # this line is needed if you use self-signed certificate in your scoring service.
12
13 data = {
14     "data": [
15         {
16             "transactionID": "example_value",
17             "accountID": "example_value",
18             "transactionAmountUSD": "0",
19             "transactionAmount": "0",
20             "transactionCurrencyCode": "example_value",
21             "transactionStatus": "0"
22         }
23     ]
24 }
```

## 모델 호출

- <사용> 탭을 클릭
- "기본 사용 정보"에서 REST 엔드포인트 및 모델 호출 기본 키를 확인할 수 있음
- "사용 옵션"에서 <Python> 탭을 클릭하여 모델 호출 코드를 확인 및 복사

# 42 | 모델 호출 (2/5)



## 모델 호출

4. <Notebooks> 탭을 클릭
5. <만들기> 버튼을 클릭하고 <새 파일 만들기>를 선택
6. "새 파일 만들기" 창에서 파일 이름 입력 (\*.ipynb 까지 입력) 및 파일 형식 "Python Notebook" 선택

# 43 | 모델 호출 (3/5)

The screenshot shows the Azure Machine Learning Studio interface with the 'Compute' tab selected. A red box highlights the list of available compute instances:

| 이름 ↑   | 범주      | 워크로드 유형                                      | 사용 가능한 할당량 | 비용        |
|--|---------|--|------------|-----------|
| <input type="radio"/> Standard_DS2_v2<br>코어 2개, 7GB RAM, 14GB 스토리지             | 범용      | Notebook(또는 다른 IDE)에서 개발 및 경량 테스트            | 코어 20개     | \$0.16/시간 |
| <input checked="" type="radio"/> Standard_DS3_v2<br>코어 4개, 14GB RAM, 28GB 스토리지 | 범용      | 클래식 ML 모델 학습, AutoML 실행, 파이프라인 실행(기본 컴퓨팅)    | 코어 20개     | \$0.33/시간 |
| <input type="radio"/> Standard_DS12_v2<br>코어 4개, 28GB RAM, 56GB 스토리지           | 메모리 최적화 | 대규모 데이터 세트(1GB 초과) 병렬 실행 단계, 일괄 처리 추론에 대한 학습 | 코어 20개     | \$0.40/시간 |
| <input type="radio"/> Standard_F4s_v2<br>코어 4개, 8GB RAM, 32GB 스토리지             | 컴퓨팅 최적화 | 실시간 추론 및 기타 대기 시간이 중요한 작업                    | 코어 20개     | \$0.19/시간 |

## 모델 호출

7. <새 컴퓨팅> 탭을 클릭
8. 가상 머신 유형, 가상 머신 크기 등 정보 선택

## 44 | 모델 호출 (4/5)

설정 구성  
선택한 가상 머신 크기의 컴퓨팅 인스턴스 설정을 구성합니다.

| 이름              | 범주 | 코어 | 사용 가능한 할당량 | RAM   | 스토리지  | 비용/시간     |
|-----------------|----|----|------------|-------|-------|-----------|
| Standard_DS3_v2 | 범용 | 4  | 코어 20개     | 14 GB | 28 GB | \$0.33/시간 |

컴퓨팅 이름 \* ⓘ  
notebook-compute

SSH 액세스 사용 ⓘ

> 고급 설정 표시

뒤로 만들기 자동화에 대한 템플릿 다운로드 취소

## 모델 호출

9. "설정 구성"에서 선택한 가상 머신 정보를 확인하고 컴퓨팅 이름 입력
10. <만들기> 버튼을 클릭하여 Notebooks 컴퓨팅을 생성

# 45 | 모델 호출 (5/5)

```

Notebooks
call_rest_api.ipynb
notebook-compute - Jupyter 커널 유통 상태
Python 3.6 - AzureML

1
2
import urllib.request
import json
import os
import ssl

def allowSelfSignedHttps(allowed):
    # bypass the server certificate verification on client side
    if allowed and not os.environ.get('PYTHONHTTPSVERIFY', '') and
       ssl._create_default_https_context == ssl._create_unverified_context:
        ssl._create_default_https_context = ssl._create_unverified_context

allowSelfSignedHttps(True) # this line is needed if you use self-signed certificate in your scoring service.

data = {
    "data": [
        {
            "transactionID": "example_value",
            "accountID": "example_value",
            "transactionAmountUSD": "0",
            "transactionAmount": "0",
            "transactionCurrencyCode": "example_value",
            "transactionDate": "0",
            "transactionTime": "0",
            "localHour": "0",
            "transactionIPaddress": "0",
            'ipState': "example_value",
            'ipPostcode': "example_value",
            'ipCountryCode': "example_value",
            'isProxyIP': "False",
            'browserLanguage': "example_value",
            'paymentInstrumentType': "example_value",
            'cardType': "example_value",
            'paymentBillingPostalCode': "example_value",
            'paymentBillingCountryCode': "example_value",
            'ipCountryCode': "example_value",
            'cvvVerifyResult': "example_value",
            'digitalItemCount': "0",
            'physicalItemCount': "0",
        },
    ],
}

```

## 모델 호출

- 생성된 컴퓨팅을 실행중인 상태를 확인할 수 있음
- 복사된 파이썬 코드를 실행창에 붙여넣기

# 46 | 호출 결과 확인

Notesbooks

call\_rest\_api.ipynb

2. 실행 버튼

```

1 import urllib.request
2 import json
3 import os
4 import ssl
5
6 def allowSelfSignedHttps(allowed):
7     # bypass the server certificate verification on client side
8     if allowed and not os.environ.get('PYTHONHTTPSVERIFY', '') and getattr(ssl, '_create_unverified_context', None):
9         ssl._create_default_https_context = ssl._create_unverified_context
10
11 allowSelfSignedHttps(True) # this line is needed if you use self-signed certificate in your scoring service.

```

1. JSON 테스트 데이터

```

{
    "transactionID": "40A1F735-520A-4885-9273-AFA2CB8DE4FF",
    "accountID": "A1688853516417580",
    "transactionAmountUSD": "798.86022",
    "transactionAmount": "6198",
    "transactionCurrencyCode": "HKD",
    "transactionDate": "20130611",
    "transactionTime": "40813",
    "localHour": "13",
    "transactionIPaddress": "153.129",
    "ipState": "tokyo",
    "ipPostcode": "100-0000",
    "ipCountryCode": "jp",
    "isProxyIP": "False",
    "browserLanguage": "zh-HK",
    "paymentInstrumentType": "CREDITCARD",
    "cardType": "MC",
    "paymentBillingPostalCode": "NA",
    "paymentBillingState": "NA",
    "paymentBillingCountryCode": "HK",
    "cvvVerifyResult": "M",
    "digitalItemCount": "2",
    "physicalItemCount": "1",
},

```

true

```

{
    "transactionID": "D11F3504-26D2-4FCF-AE33-0B96E80452F0",
    "accountID": "A1055521409411370",
    "transactionAmountUSD": "139.99",
    "transactionAmount": "139.99",
    "transactionCurrencyCode": "USD",
    "transactionDate": "20130527",
    "transactionTime": "201016",
    "localHour": "13",
    "transactionIPaddress": "75.7",
    "ipState": "colorado",
    "ipPostcode": "80113",
    "ipCountryCode": "us",
    "isProxyIP": "False",
    "browserLanguage": "en-US",
    "paymentInstrumentType": "CREDITCARD",
    "cardType": "AMEX",
    "paymentBillingPostalCode": "80108",
    "paymentBillingState": "CO",
    "paymentBillingCountryCode": "US",
    "cvvVerifyResult": "M",
    "digitalItemCount": "1",
    "physicalItemCount": "0",
},

```

false

3. 테스트 결과: b'{"result": [true, false]}'

## 모델 호출 코드 실행

- 테스트할 데이터를 json 형식으로 입력
- <셀 실행> 버튼을 클릭
- 모델 호출 코드를 실행하여 테스트 결과 확인

# Q & A