

자동화된 ML, 나도 해보자

Episode 6: 자동화된 ML 코드로 돌려보기

한석진
마이크로소프트

Episode 6

자동화된ML

코드로 돌려보기

입맛대로 샘플 고르기

컴퓨팅 인스턴스 만들기

샘플 가져오기 / 샘플 실행하기

결과 확인하기 (학습결과, 모델설명, 배포까지)

입맛대로 샘플 고르기

<https://aka.ms/azpls/azureml-automl-samples>

분류

- **V신용카드 사기 탐지**
 - Dataset: Kaggle's credit card fraud detection dataset
 - <https://aka.ms/azpls/azureml-automl-sample-ccard-local>
- **은행 정기예금 개설 예측**
 - Dataset: UCI's bank marketing dataset
- **뉴스 기사 텍스트로 주제 분류**
 - Dataset: 20 newsgroups text dataset

회귀분석

- **하드웨어 부품 성능 예측**
 - Dataset: Hardware Performance Dataset

시계열

- **에너지 수요 예측**
 - Dataset: NYC energy demand data
- **오렌지 주스 매출 예측 (Multi-Series)**
 - Dataset: Dominick's grocery sales of orange juice
- **자전거 대여 서비스 수요예측**
 - Dataset: Bike demand data
- **Forecast 함수 사용 예제 (Recursive 등 포함)**
 - Dataset: 샘플 데이터
- **음료 생산 예측**
 - Dataset: Monthly beer production data
- **연결하여 재학습하기 예제**
 - Dataset: NOAA weather data (Azure Open Datasets)

컴퓨팅 인스턴스 만들기

Microsoft Azure Machine Learning

≡

+ 새로 만들기

홈

작성자

Notebooks

자동화된 ML

디자이너

자산

데이터 세트

실험

파이프라인

모델

엔드포인트

관리

컴퓨팅

데이터 저장소

데이터 레이블 지정

aml-handson-202001 > 컴퓨팅

컴퓨팅 목록

컴퓨팅 인스턴스

컴퓨팅 클러스터유추 클러스터연결된 컴퓨팅

① Microsoft는 COVID-19에 대응하기 위해 응급 의료요원, 의료 및 응급상황 관리 서비스, 중요한 정부 인프라 및 기존 유료 고객 사용에 ...

+ 새로 만들기

↺ 새로 고침

▶ 시작

⏹ 중지

↺ 다시 시작

...

🔍 검색하여 항목 필터링...

이름	상태	애플리케이션 URI	가상 머신 크기
ci-sehan-ds3v2	실행 중	JupyterLab Jupyter RStudio SSH	STANDARD_DS3

원하는 샘플을 직접 어디서든지 가져오거나...

The screenshot displays the Microsoft Azure Machine Learning web interface. The left sidebar contains a navigation menu with options like '새로 만들기' (Create), '홈' (Home), 'Notebooks', '자동화된 ML' (Automated ML), '디자이너' (Designer), '데이터 세트' (Data Set), '실험' (Experiment), '파이프라인' (Pipeline), '모델' (Model), '엔드포인트' (Endpoint), '컴퓨팅' (Compute), '데이터 저장소' (Data Store), and '데이터 레이블 지정' (Data Labeling). The main area shows the 'aml-handson-202001' workspace with a '컴퓨팅 목록' (Compute List) and a 'Notebooks' section. The 'Notebooks' section lists various notebooks, including 'dem108ht', 'sehan', 'automl-load-test', 'AzureML-BERT', 'azureml-compute-instance-mgmt', 'computervision-recipes', and 'designer-exploration'. A terminal window is open on the right, showing the 'Welcome to Azure Machine Learning Terminal' and the command prompt 'azureuser@ci-sehan-ds3v2:~/cloudfiles/code/Users/sehan\$'.

Microsoft Azure Machine Learning

aml-handson-202001 > 컴퓨팅 목록

컴퓨팅 목록

Microsoft는 COVID-19에 대응하기 위해 Azure Machine Learning을 활용하고 있습니다.

+ 새로 만들기 새로 고침

이름

ci-sehan-ds3v2

Microsoft Azure Machine Learning

aml-handson-202001 > Notebooks

Notebooks

파일 샘플 Notebook

3:ci-sehan-ds3v2

컴퓨팅: ci-sehan-ds3v2 - 실행 중

Welcome to Azure Machine Learning Terminal

Type "git clone [url]" to clone a repo
Type "git --help" to learn about Git CLI
Type "az ml --help" to learn about Azure ML CLI

azureuser@ci-sehan-ds3v2:~/cloudfiles/code/Users/sehan\$

애저ML 샘플이라면 바로 옆에서 가져오기

The screenshot displays the Microsoft Azure Machine Learning web interface. On the left, a sidebar contains navigation options such as '새로 만들기' (Create new), '홈' (Home), 'Notebooks', '자동화된 ML' (Automated ML), '데이터 세트' (Data sets), '실험' (Experiments), '파이프라인' (Pipelines), '모델' (Models), '엔드포인트' (Endpoints), '컴퓨팅' (Computing), '데이터 저장소' (Data stores), and '데이터 레이블 지정' (Data labeling). The main area is divided into two panels. The left panel shows the 'aml-handson-202001' workspace with a 'Notebooks' section. Under the 'Samples' folder, the '1.18.0' subfolder is expanded, and the 'automated-machine-learning' folder is highlighted with a red box. The right panel shows the 'Notebooks' section of the same workspace. The '샘플 Notebook' (Sample Notebook) tab is selected, and the 'local-run-classification-credit-card-fraud' folder is highlighted with a red box. Below this, the file 'auto-ml-classification-credit-card-fraud-local.yml' is listed. On the far right, a sidebar displays the 'auto-ml-classification-credit-card-fraud-local.yml' file details, including a warning '편집을 시작하시겠습니까?' (Do you want to start editing?) and a button '이 Notebook 복제' (Copy this Notebook), both highlighted with red boxes. Below the warning, the copyright notice 'Copyright (c) Microsoft Corporation. All rights reserved. Licensed under the MIT License.' is visible. At the bottom, the 'Automated Machine Learning' section is shown, with the title 'Classification of credit card fraudulent transactions with local run' and a 'Contents' list: 1. Introduction, 2. Setup, 3. Train, 4. Results, 5. Test, 6. Explanation.

주피터 노트북 실행하기 (Python 기반)

The screenshot displays the Microsoft Azure Machine Learning web interface. The left sidebar contains navigation options: '새로 만들기' (Create), '홈' (Home), '작성자' (Author), 'Notebooks', '자동화된 ML' (Automated ML), '디자이너' (Designer), '자산' (Assets), '데이터 세트' (Data Set), '실험' (Experiment), '파이프라인' (Pipeline), '모델' (Model), '엔드포인트' (Endpoint), '관리' (Manage), '컴퓨팅' (Compute), '데이터 저장소' (Data Store), and '데이터 레이블 지정' (Data Labeling). The main area shows the 'aml-handson-202001' workspace with a 'Notebooks' tab. A file explorer on the left lists folders like 'forecasting-energy', 'forecasting-forecast', 'forecasting-orange', and 'local-run-classification', along with files 'auto-ml-classification', 'automl_env.yml', and 'automl_env_linux.yml'. The 'auto-ml-classification' notebook is open, showing a Jupyter interface with a toolbar and a code cell. The code cell contains the following Python code:

```
1 import logging
2
3 from matplotlib import pyplot as plt
4 import pandas as pd
5
6 import azureml.core
7 from azureml.core.experiment import Experiment
8 from azureml.core.workspace import Workspace
9 from azureml.core.dataset import Dataset
10 from azureml.train.automl import AutoMLConfig
11 from azureml.interpret import ExplanationClient
```

Below the code cell, a message states: "This sample notebook may use features that are not available in previous versions of the Azure ML SDK." The top status bar indicates the notebook is running on 'ci-sehan-ds3v2' using 'Python 3.6 - AzureML'.

실행 결과 확인 (학습결과, 모델설명, 배포까지)

Microsoft Azure Machine Learning Studio interface showing the workspace with various components like Notebooks, Datasets, and Models. The 'aml-han' workspace is selected, and the 'Notebooks' tab is active. The 'ci-seh' notebook is highlighted in the list.

```
[30] 1 # Serialize the first row of the test data into json
      2 X_test_json = X_test_df[:1].to_json(orient='records')
      3 print(X_test_json)
      4
      5 # Call the service to get the predictions and the engineered and raw explanations
      6 output = aks_service.run(X_test_json)
      7
      8 # Print the predicted value
      9 print('predictions:\n{}\n'.format(output['predictions']))
     10 # Print the engineered feature importances for the predicted value
     11 print('engineered_local_importance_values:\n{}\n'.format(output['engineered_local_importance_values']))
     12 # Print the raw feature importances for the predicted value
     13 print('raw_local_importance_values:\n{}\n'.format(output['raw_local_importance_values']))
     14
```

```
[{"Time":1.0,"V1":-0.966271712,"V2":-0.185226008,"V3":1.79299334,"V4":-0.863291275,"V5":-0.01030888,"V6":1.24
7203168,"V7":0.23760894,"V8":0.377435875,"V9":-1.387024063,"V10":-0.054951922,"V11":-0.226487264,"V12":0.1782
28226,"V13":0.50775687,"V14":-0.287923745,"V15":-0.631418118,"V16":-1.059647245,"V17":-0.684092786,"V18":1.96
5775003,"V19":-1.23262197,"V20":-0.208037781,"V21":-0.108300452,"V22":0.005273597,"V23":-0.190320519,"V24":-1
.175575332,"V25":0.647376035,"V26":-0.221928844,"V27":0.062722849,"V28":0.061457629,"Amount":123.5}]
```

```
predictions:
[False]
```

```
engineered_local_importance_values:
```

```
[[-0.00016403975366892305, -3.611681645118468e-05, 6.415124728881989e-05, -0.00011283265709443348,
-0.0019114408081337174, 7.176774833342308e-05, -4.5091258534032926e-05, -0.00035454702428313294,
-0.0002065916944927191, 0.000473752297398849, -0.0024285077789540266, -0.0002772885056132418,
-0.002490733985921676, -8.341464327372632e-05, -0.007198314466574828, 0.00013114908753018528,
-4.400396764750505e-05, -0.0006338343268349717, -9.104375933792205e-05, -0.00010424632598777514,
0.00018462347423919907, -4.363143130284942e-05, -3.887407366631882e-05, -1.6775605272163988e-05,
```

Episode 6

자동화된ML

코드로 돌려보기

입맛대로 샘플 고르기

컴퓨팅 인스턴스 만들기

샘플 가져오기 / 샘플 실행하기

결과 확인하기 (학습결과, 모델설명, 배포까지)