

캐글 활용

캐글(Kaggle)

- 데이터 사이언스 경진 대회 플랫폼(kaggle.com)
 - 데이터를 분석하는 대회
 - 각 종 데이터 사이언스 contest가 열리고, 데이터셋 제공
 - 머신러닝, 딥러닝을 연습하고 싶은 분들에게는 최고의 플랫폼
 - 초보자가 접근하기에 다소 어려운 플랫폼

Kaggle.com

The screenshot shows the Kaggle homepage with a banner for the COVID-19 challenge. Below the banner, there's a section titled 'Start with more than a blinking cursor' which describes the no-setup environment. A 'REGISTER WITH GOOGLE' button is visible. An inset shows a notebook titled 'Predict Malicious Websites: XGBoost' with Python code for data loading and preprocessing, and a preview of the dataset table.

Help us better understand COVID-19

There is a large body of research and data around COVID-19. Help the global community better understand the disease by getting involved on Kaggle.

[Get Started](#) [View Contributions](#)

Start with more than a blinking cursor

Kaggle offers a no-setup, customizable, Jupyter Notebooks environment. Access free GPUs and a huge repository of community published data & code.

[REGISTER WITH GOOGLE](#)

[Register with Email](#)

Predict Malicious Websites: XGBoost

Python notebook using data from [Malicious and Benign Websites](#) 4 votes

```
import matplotlib.pyplot as plt

from sklearn.model_selection import train_test_split
from sklearn.metrics import confusion_matrix
from sklearn.utils.multiclass import unique_labels

data = pd.read_csv("../input/dataset.csv")

# clean up column names
data.columns = data.columns.\
    str.strip().\
    str.lower()

# remove non-numeric columns
data = data.select_dtypes(['number'])

# split data into training & testing
train, test = train_test_split(data, shuffle=True)

# peek @ dataframe
train.head()
```

Out[1]:

	url_length	number_of_urls	content_length	log_consensus_percentage	dist_remote_host	remote_ip	app_ip
344	37	9	162.0	1	0	1	66
77	26	6	NAN	0	0	0	0
933	51	10	231.0	7	1	2	769
1008	54	12	NAN	0	0	0	0
882	49	10	69441.0	14	1	3	1387

Try Now

```
In [2]: # split training data into inputs & outputs
X = train.drop(["type"], axis=1)
Y = train["type"]
```















We use cookies on Kaggle to deliver our services, analyze web traffic, and improve your experience on the site. By using Kaggle, you agree to our use of cookies.

[Got it](#) [Learn more](#)

캐글 courses

- Faster Data Science Education

Micro-Courses

	Python Learn the most important language for data science.
	Intro to Machine Learning Learn the core ideas in machine learning, and build your first models.
	Intermediate Machine Learning Learn to handle missing values, non-numeric values, data leakage and more. Your models will be more accurate and useful.
	Data Visualization Make great data visualizations. A great way to see the power of coding!
	Pandas Solve short hands-on challenges to perfect your data manipulation skills.
	Feature Engineering Discover the most effective way to improve your models.
	Deep Learning Use TensorFlow to take machine learning to the next level. Your new skills will amaze you.
	Intro to SQL Learn SQL for working with databases, using Google BigQuery to scale to massive datasets.
	Advanced SQL Take your SQL skills to the next level.
	Geospatial Analysis Create interactive maps, and discover patterns in geospatial data.
	Microchallenges Solve ultra-short challenges to build and test your skill.
	Machine Learning Explainability Extract human-understandable insights from any machine learning model.
	Natural Language Processing Distinguish yourself by learning to work with text data.
	Intro to Game AI and Reinforcement Learning Build your own video game bots, using classic algorithms and cutting-edge techniques.

캐글 Notebook

The screenshot shows the Kaggle website interface. At the top, there's a navigation bar with the Kaggle logo, a search bar, and links for 'Sign In' and 'Register'. Below this, the 'Notebooks' section is prominently displayed with the text 'Explore and run machine learning code with Kaggle Notebooks! Find help in the [Documentation](#).' A red dashed box highlights the '+ New Notebook' button in the top right corner of the main content area. Below the 'Notebooks' header, there are tabs for 'Public', 'Your Work', 'Shared With You', and 'Favorites'. A 'Sort by' dropdown menu is set to 'Hotness'. Below these tabs, there are filters for 'Categories', 'Outputs', 'Languages', and 'Tags', along with a 'Search notebooks' input field. The main list of notebooks shows three items: 'COVID-19 data with SIR model' (272 upvotes, 6h ago), 'PANDA concat tile pooling starter [inference]' (20 upvotes, 7h ago), and 'Panorama do COVID-19 no Brasil' (84 upvotes, 3h ago). Each item includes a brief description, tags, and a 'Py' icon indicating it's a Python notebook. At the bottom of the page, there's a cookie consent banner.

We use cookies on Kaggle to deliver our services, analyze web traffic, and improve your experience on the site. By using Kaggle, you agree to our use of cookies. [Got it](#) [Learn more](#)

노트북

- 원래 주피터(ipython)의 파이썬 웹개발 환경
 - 캐글 Notebook
 - 코랩 Notebook
- 캐글에서 자체 Notebook
 - 캐글에서 정답 제출
 - 수정된 알고리즘을 바로 commit하여 점수를 확인 가능
 - 연습용으로 활용 가능
 - 생성된 Notebook을 서버에 저장해 줍니다.
 - 따로 package 설치가 필요하지 않아도 기본 패키지들이 설치
 - 추가 패키지 설치도 가능
 - 성능이 좋은 GPU를 사용 가능
 - GPU 자원을 '무료' 사용 가능

로그인

- 구글 계정으로 계정 만들어 로그인