1. Kaggle 소개

Kaggle 플랫폼에 대해 알아보자

한기용

keeyonghan@hotmail.com

programmers

목차

programmers

Kaggle이란?

Kaggle의 역사부터 제공 서비스까지 알아보자

Kaggle 소개

- 호주에서 시작한 국제적인 기계학습 컨테스트를 제공하는 서비스
- 2017년에 구글에 인수됨 (Google Cloud)
- 데이터 과학자들에게 자신의 능력을 검증해보면서 몸값을 올릴 수 있는 좋은 기회를 제공
  - Kaggle Code
  - Kaggle Ranking

≡ kaggle

Kaggle Code

- 이전에는 Kaggle Kernel이라 부름
- 데이터 과학 관련한 모듈이 모두 설치된 주피터 파이썬 노트북
  - 사용이 무료! 4 CPU, 16GB memory, 1GB of disk
  - 여기서 만든 노트북을 다른 이들과 공유 가능
- 다른 이들이 만든 노트북을 보고 학습하기에 최적

Code

Explore and run machine learning code with Kaggle Notebooks.
Find help in the Documentation.

kaggle이란?

Kaggle Ranking

| Competitions | Datasets | Notebooks | Discussions | | Learn more about rankings |

| 318 Grandmasters | 2,029 Masters | 9,389 Experts | 68,693 Contributors | 116,502 Novices |

Rank	Tier	User	Medals			Points
1		Dieter	38	16	3	171,950
2		Psi	34	8	0	169,072
3		bestfitting	42	13	2	142,954

1. Novices
2. Contributors
3. Experts
4. Masters
5. Grandmasters

랭킹 시스템
이해하기

Kaggle 기타 주요 기능들

- Competitions
- Datasets
- Models
- Discussions
- Learn

- **Competitions**는 뒤에서 자세히 설명
- 나머지는 간단히 기능 설명
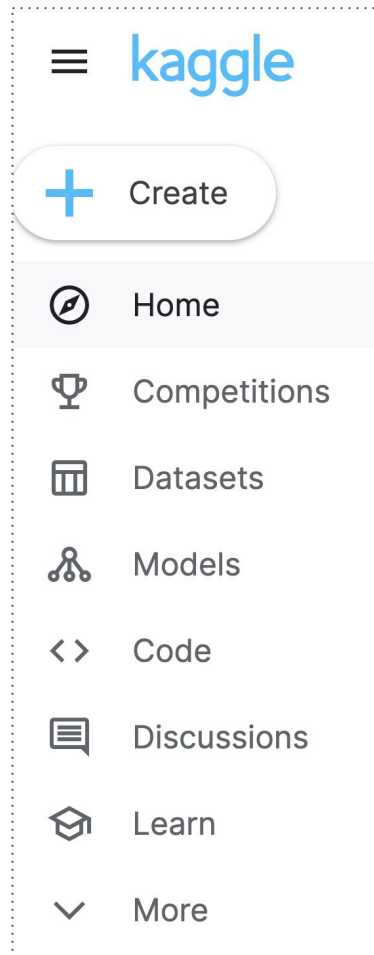
Kaggle 대회 이해

Kaggle Competitions에 대해 이해하고 어떻게 내 커리어 발전에 사용할 수 있는지 알아보자

<u>Kaggle Competitions</u>의 목적은?

- 참가자 관점에서의 목적
  - 기술 습득
  - 문제 해결
  - 금전/경력 인센티브
- 조직/회사 관점에서의 목적
  - 혁신 장려
  - 인재 발굴
  - 벤치마킹
  - 연구 기여
- Kaggle 관점에서의 목적
  - 커뮤니티 빌딩
  - 매출 증대

Competitions

Grow your data science skills by competing in our exciting competitions. Find help in the <u>documentation</u> or learn about <u>Community Competitions</u>.

Host a Competition

Kaggle Competitions 종류: 회사 관점 (대회 호스팅)

Kaggle Competition Features

- Dataset Hosting
- Real-time Leaderboards
- Preloaded Metrics
- Discussion Forums
- Automated Scoring
- Kaggle Notebooks

Community

For educators, small businesses, or ML enthusiasts

1

Featured

For businesses or organizations with demanding ML problems

2

Research

For Academic / Research Institutions or Non-Profits

3

문제 해결과 인재
채용/발굴

Kaggle Competitions 종류 - 개인 관점 (학습,경험,이력,금전)

All Competitions ☰ Everything, past & present	**Featured** ☆ Premier challenges with prizes	**Getting Started** ⚑ Approachable ML fundamentals
Research ⚗ Scientific and scholarly challenges	**Community** 👥 Created by fellow Kagglers	**Playground** 🎉 Fun practice problems
Simulations 🤖 Train bots to navigate environments	**Analytics** 📊 Open-ended explorations	
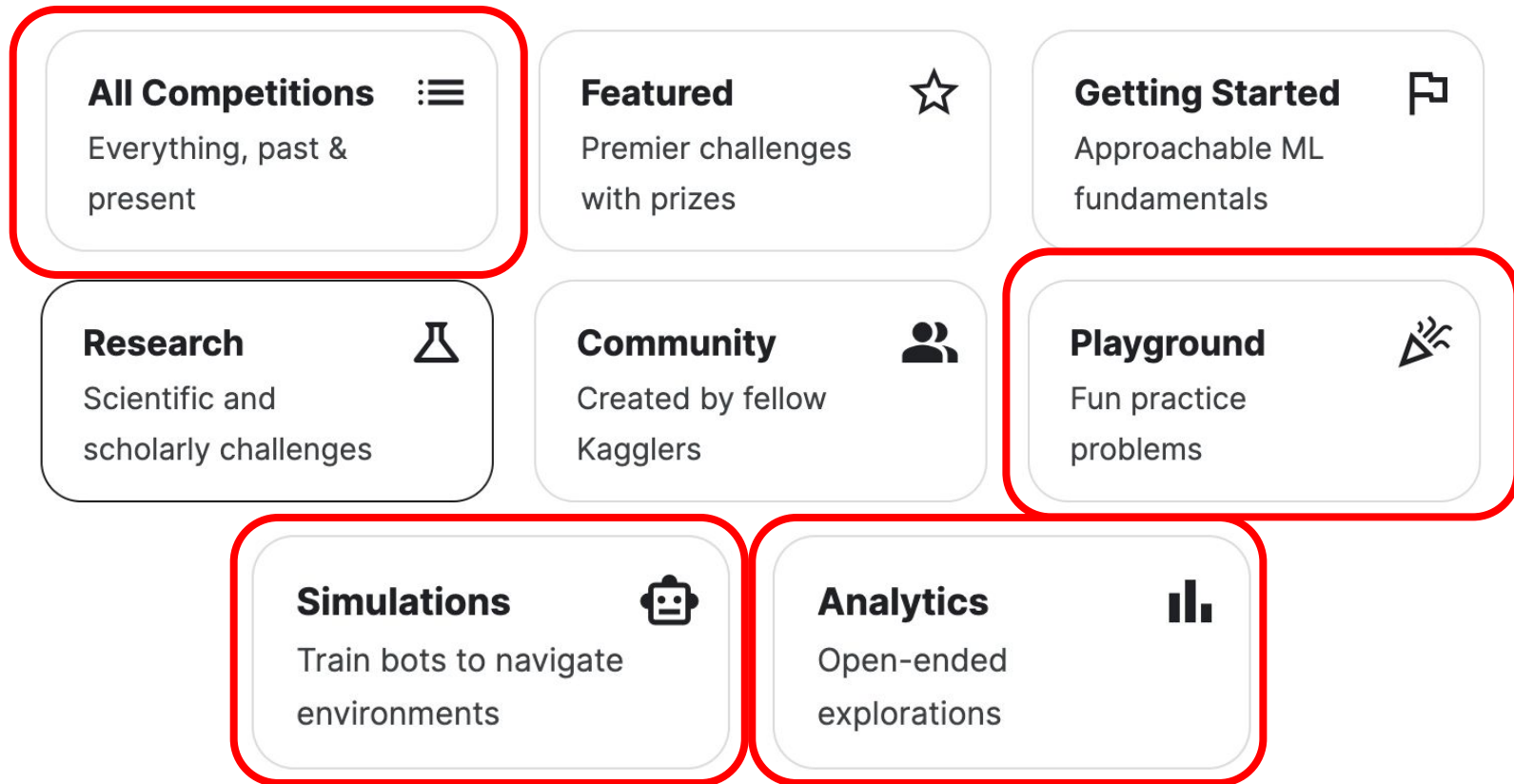
[Kaggle Competitions](): All Competitions
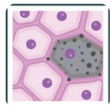
Kaggle Competitions: Playground

Playground ✕

Results

Total Teams ▾ ▦

Bike Sharing Demand
Forecast use of a city bikeshare system
Playground · 3242 Teams · 9 years ago

Knowledge

⋯

San Francisco Crime Classification
Predict the category of crimes that occurred in the city by the bay
Playground · 2331 Teams · 8 years ago

Knowledge

⋯

Don't Overfit! II
A Fistful of Samples
Playground · 2315 Teams · 5 years ago

Swag

⋯

Tabular Playground Series - Sep 2021
Practice your ML skills on this approachable dataset!
Playground · 1942 Teams · 2 years ago

Swag

⋯

[Kaggle Competitions](): [Simulations]()

Simulations ✕

Results

Total Teams ▾

Rock, Paper, Scissors
Shoot!
Playground · Simulation Competition · 1662 Teams · 3 years ago

Prizes
· · ·

Lux AI
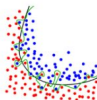Gather the most resources and survive the night!
Featured · Simulation Competition · 1178 Teams · 2 years ago

$10,000
· · ·

Halite by Two Sigma
Collect the most halite during your match in space
Featured · Simulation Competition · 1139 Teams · 3 years ago

Swag
· · ·

Google Research Football with Manchester City F.C.
Train agents to master the world's most popular sport
Featured · Simulation Competition · 1138 Teams · 3 years ago

$6,000
· · ·

Hungry Geese
Don't. Stop. Eating.
Playground · Simulation Competition · 879 Teams · 2 years ago

Prizes
· · ·

[Kaggle Competitions](#): [Analytics](#)

Analytics ✕

Results   Recently Launched ▾   ⊞

NFL Big Data Bowl 2024   **$100,000**
Help evaluate tackling tactics and strategy
Analytics · 0 Teams · 10 days to go   •••

LLM Prompting with MakerSuite   **Swag**
Design useful, innovative prompts for LLMs using MakerSuite
Analytics · 0 Teams · 2 months ago   •••

2023 Kaggle AI Report   **$70,000**
Essays on the state of machine learning in 2023
Analytics · 220 Teams · 5 months ago   •••

NFL Big Data Bowl 2023   **$100,000**
Help evaluate linemen on pass plays
Analytics · 0 Teams · a year ago   •••

2022 Kaggle Machine Learning & Data Science Survey   **$30,000**
The most comprehensive dataset available on the state of ML and data science
Analytics · 0 Teams · a year ago   •••

캐글 랭킹이 어떻게 개인의 커리어에 도움이 될까**?**

- 그랜드마스터는 전세계적으로 몇명되지 않음
- 이력서에 적기 아주 좋은 경력

AI 타임스 기사: https://www.aitimes.com/news/articleView.html?idxno=140608

Kaggle 대회 참여해보기

Kaggle에서 가장 유명한 대회인 타이타닉 승객 생존 예측에
참여하면서 Kaggle 환경을 이해해보자.

Kaggle 타이타닉: https://www.kaggle.com/competitions/titanic

- 프로그래밍 언어로 치면 "Hello, World"에 해당하는 유명한 모델링 문제
- 이 문제를 풀고 제출해봄으로써 Kaggle 환경을 익히는 것이 이번 섹션의 목적
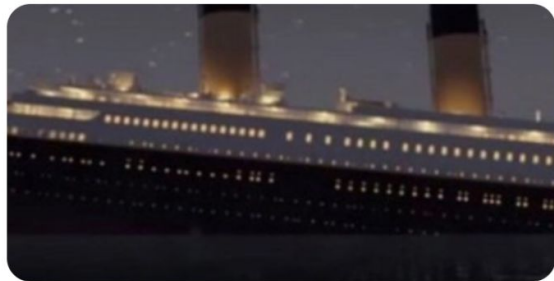
Kaggle API

- https://www.kaggle.com/docs/api
- Kaggle 노트북 등을 사용하지 않고 본인의 환경에서 모델링을 하고 싶은 경우
  - pip install kaggle
- 아래 일을 수행 가능
  - Interacting with Competitions
    - kaggle competitions list
    - kaggle competitions download -c [COMPETITION]
    - kaggle competitions submit -c [COMPETITION] -f [FILE] -m [MESSAGE]
  - Interacting with Datasets
  - Interacting with Notebooks

실습

- 먼저 회원 가입하고 아래 링크 방문
- https://www.kaggle.com/competitions/titanic

programmers

숙제

오늘 배운 것을 정리해보도록 하자

오늘의 숙제

- 캐글에 사용자 등록하기: Novice 랭킹
- 캐글 Contributor 되기

Contributor

You've completed your profile, engaged with the community, and fully explored Kaggle's platform.

☐ Run 1 notebook or script

☐ Make 1 competition submission

☐ Make 1 comment

☐ Give 1 upvote

programmers

Q & A

오늘 강의에 대해서 궁금한 부분이 있으면 알려주세요!