

Kaggle로 스펙을 삼아보자

~~할게 많은데 Kaggle까지 해야한다나...~~



Content

- Kaggle로 스펙을 삼는 Tip
- Kaggle Notebook 필사 그리고 양념 한 스푼
- 취업 준비 Tip



kaggle™

Kaggle로 스펙을 삼는 Tip

Kaggle로 스펙을 삼는 Tip

- Competition과 메달
- 스토리 텔링
- 점수를 높이는 Tip

Competition과 메달

- Tier, 메달은 그 자체가 스펙
 - 전 세계의 뛰어난 경쟁자들
 - 상당한 문제의 난이도
 - ex) Tweet Sentiment Extraction, Generative Dog Images, Lyft Motion Prediction for Autonomous Vehicles
- 무조건 Expert는 달성하라



Competition과 메달

- Notebook을 통한 스터디
- 최신 트렌드를 반영한 코드
- 바로 실행가능한 코드
- 다양한 측면에서의 접근



스토리텔링

어쩐지 오늘은

Machine Learning Engineer

Diary

Data

MLOps

Google Cloud Platform

Development

ETC

About

Q Search

메모가 습관인 데이터쟁이입니다



Uber's Real-time Data Intelligence Platform At Scale: Improving Gairos Scalability/Reliability 번역

14 Mar 2021 in **Data on Mobility**

- Uber's Real-time Data Intelligence Platform At Scale: Improving Gairos Scalability/Reliability 번역 글입니다
 - 전문 번역가가 아니기 때문에 오역이 있을 수 있으며, 의역도 존재합니다
 - 관심이 있으신 분들은 원문을 꼭 보시는 것을 추천드립니다!
 - 실시간 데이터 인텔리전스 플랫폼을 위한 아키텍처가 자세히 나와있습니다
- 키워드: data intelligence platform, real time data intelligence

[Continue reading →](#)

데이터 분석가(데이터 직무), 데이터를 활용하는 다양한 직군 소개

21 Feb 2021 in **Diary on diary**

- 데이터 분석가(데이터 직무), 데이터를 활용하는 다양한 직군 소개 글입니다
 - 키워드: 데이터 분석가, 퍼포먼스 마케터, CRM 마케터, 그로스해커, 비즈니스 분석가, 프로덕트 분석가
- I Want To Study Data Science에 정리한 내용을 보면 "데이터 분석가"를 통합해서 정리했습니다
 - 최근엔 데이터 분석가란 이름으로 통합하는 추세지만, 데이터를 활용하는 직군들이 다양해서 처음 접하는 분들이 혼란이 오는 것 같습니다. 이런 디테일한 이해가 있으면 좋을 것 같아 이 글을 작성합니다

[Continue reading →](#)

towards data science

A Medium publication sharing concepts, ideas and codes.

Follow

568K Followers

Editors' Picks

Features

Explore

Grow

Contribute

About

Elliot Gunn · 3 days ago

READING LIST

The All-time Best Guides to Data Science Writing

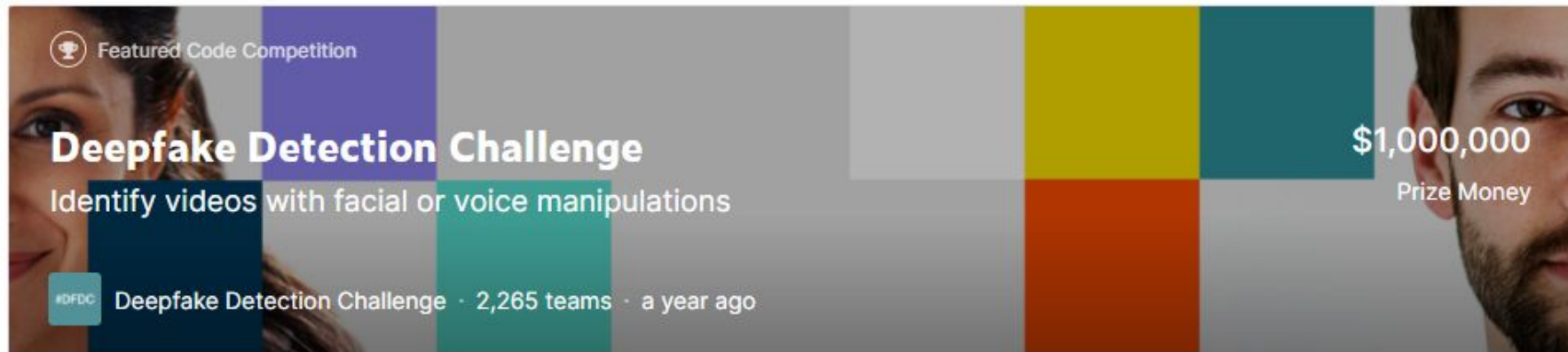
Learn how to write better for your colleagues and peers



Photo by [Freddy Castro](#) on [Unsplash](#)


The data science blogging ecosystem is rich and growing. TDS alone has an archive of more than 20,000 posts across numerous topics. Many experts

스토리텔링



- | | | |
|-------------------------------|---|--------------------|
| 1. 동영상에서 딥러닝을 이용한 판별 | → | - ML/DL Engineer |
| 2. 모델 아키텍처 설계 및 학습 | → | - ML/DL Researcher |
| 3. 생성된 얼굴, 립 모션에 관한 분석, 의미 분석 | → | - Data Scientist |

점수를 높이는 Tip



bestfitting

Joined 6 years ago · last seen in the past day

Followers 12951

Competitions Grandmaster

[Home](#) [Competitions \(50\)](#) [Code \(2\)](#) [Discussion \(259\)](#) [Followers \(12,951\)](#) [Contact User](#) [Unfollow User](#)

Competitions Grandmaster

Current Rank	Highest Rank
2 of 179,916	1

36	11	1
----	----	---

Human Protein ...
1st of 757
- a year ago
Top 1%

Human Protein ...
1st of 2160
- 3 years ago
Top 1%

Cdiscount's Ima...
1st of 626
- 4 years ago
Top 1%

Datasets Contributor

Unranked

0	0	0
---	---	---

No dataset results

Notebooks Contributor

Unranked

1	1	0
---	---	---

i96_1e-4_6e_4_b...
100 votes
- 3 years ago

i96_crop_mode...
27 votes
- 3 years ago

Discussion Expert

Current Rank	Highest Rank
144 of 279,706	23

16	15	133
----	----	-----

A CNN classifier...
607 votes
- 3 years ago

My brief summa...
263 votes
- 4 years ago

Fair Cell Activati...
208 votes
- a year ago

점수를 높이는 Tip

1. 대회 개요 및 데이터 설명을 주의 깊게 읽으십시오.
2. 유사한 Kaggle 대회를 찾으십시오. (비교적 신참인 저는 모든 Kaggle 대회에 대한 기본 분석을 수집하고 수행했습니다.)
3. 유사한 경쟁의 솔루션을 읽어보십시오.
4. 최신 트렌드를 놓치지 않도록 논문을 읽으십시오.
5. 데이터를 분석하고 안정적인 CV(교차검증)를 만드십시오.
6. 데이터 전처리, 특징 공학, 모델 학습.
7. 예측 분포, 오류 분석, 어려운 예제와 같은 결과 분석.
8. 분석을 기반으로 모델을 정교화하거나 새 모델을 설계합니다.
9. 데이터 분석 및 결과 분석을 기반으로 모델을 설계하여 다양성을 추가하거나 어려운 샘플을 해결합니다.
10. 필요한 경우 이전 단계로 돌아갑니다.

점수를 높이는 Tip

Hyper parameter tuning

- 데이터에 대한 이해와 알고리즘의 이론을 바탕으로 파라미터를 조정
- 관련 논문에서 저자들이 시행착오를 찾으려고 노력
- 예측 분포, 영향을 받은 샘플 등과 같은 매개 변수 변경 전후의 결과를 비교

견고한 CV(교차 검증)를 이용한 최종 제출안 선택

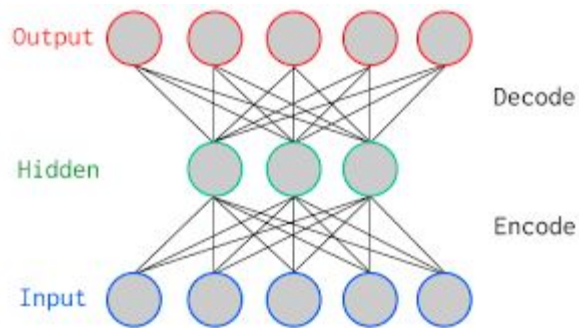
- 좋은 CV는 성공의 절반. 내 모델을 평가할 좋은 방법을 찾지 않으면 다음 단계로 넘어가지 않음
- 안전한 모델의 가중 평균 앙상블 하나와 상대적으로 위험한 모델을 선택

점수를 높이는 Tip

In a few words, what wins competitions?

1. Good CV
2. 다른 대회에서 배우기
3. 관련 논문 읽기
4. 훈련 및 정신적 강인함





Kaggle Notebook 필사

그리고 양념 한 스푼

Kaggle Notebook 필사 그리고 양념 한 스푼

- 노트북을 통한 학습
- 노트북 필사 예시
- 그리고 양념 한 스푼

준지도학습

지도학습

- 레이블이 있는 데이터셋에 대해서 모델을 학습하는 과정
- ex) 분류, 회귀, 예측 등

비지도학습

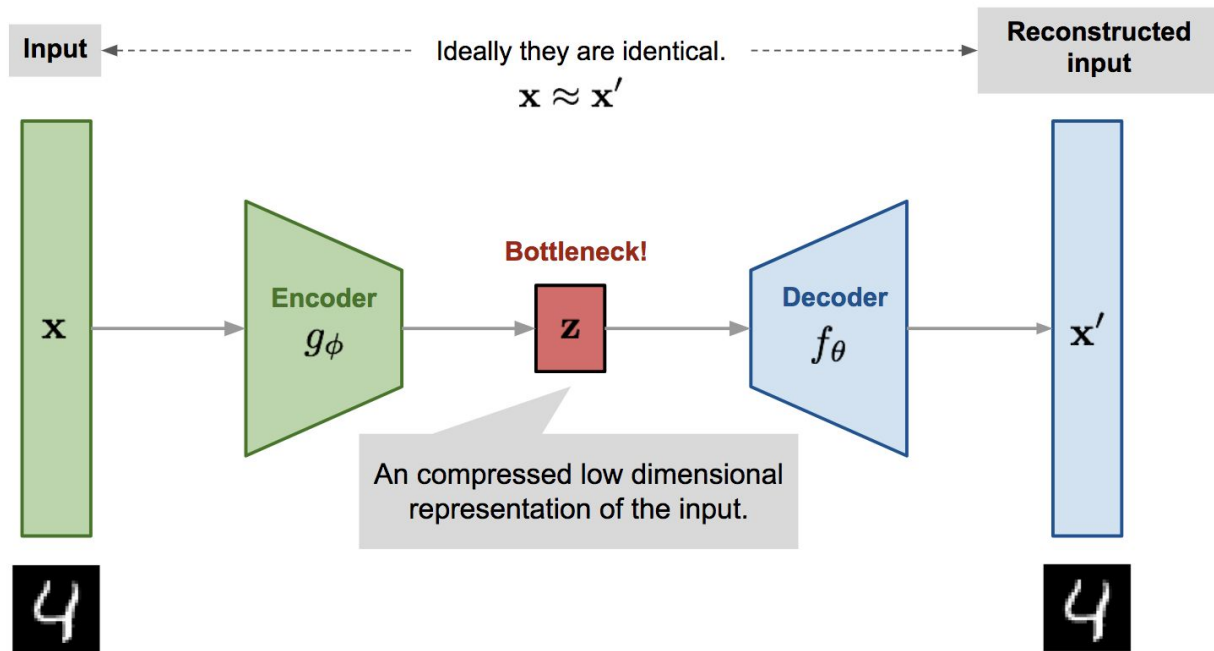
- 목표하는 변수를 알 수 없는 데이터셋에 대해서 모델을 학습하는 과정
- ex) 군집화, 차원 축소, 자기지도학습 등

준지도학습

- 비지도학습 + 지도학습

AutoEncoder

- 가장 대표적인 비지도학습 방법
- 입력을 출력으로 변환하기 위해 표현하는 중간 상태를 학습



Credit Card Fraud Detection



The screenshot shows the top section of a Kaggle dataset page. At the top left, it says 'Dataset'. The title 'Credit Card Fraud Detection' is prominently displayed, followed by the subtitle 'Anonymized credit card transactions labeled as fraudulent or genuine'. Below this, it says 'Machine Learning Group - ULB' and 'updated 3 years ago (Version 3)'. On the right, there's a yellow circle icon and a box showing an upward arrow and the number '7491'. At the bottom of the header, there are tabs for 'Data', 'Tasks (10)', 'Code (3,068)', 'Discussion (82)', 'Activity', and 'Metadata'. To the right of these tabs are two buttons: 'Download (144 MB)' and 'New Notebook'.

- 2013년 9월의 2일간 발생한 유럽 카드 소지자의 신용 카드 거래 데이터셋
- 총 284,807 건의 거래 중 492 건의 사기 (0.172 %)
- 기밀 유지 문제로 인해 본래 데이터의 Feature와 배경 정보를 제공 X -> PCA로 변환한 데이터가 주어짐
- Class가 1이면 사기, 아니면 0



SHIVAM BANSAL · 3Y AGO · 56,523 VIEWS



376

Edit My Copy

465

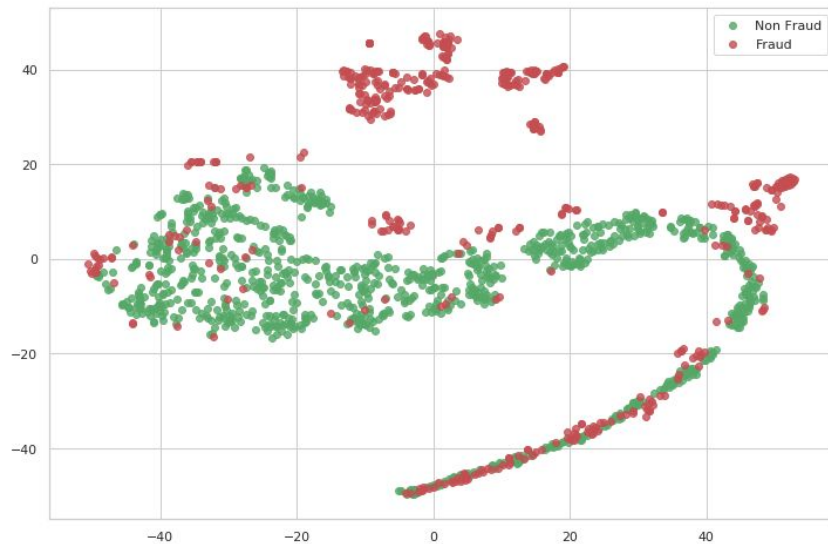


Semi Supervised Classification using AutoEncoders

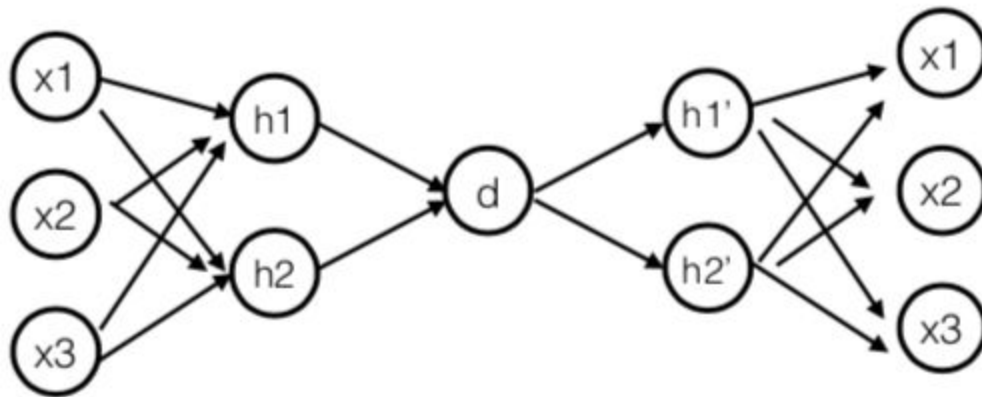
Python · [Credit Card Fraud Detection](#), [Titanic - Machine Learning from Disaster](#)

t-SNE visualization

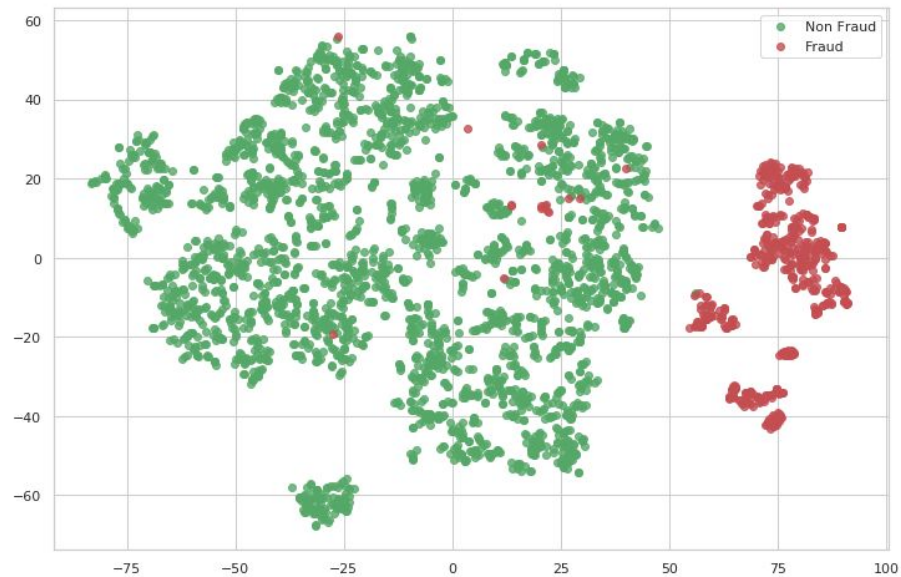
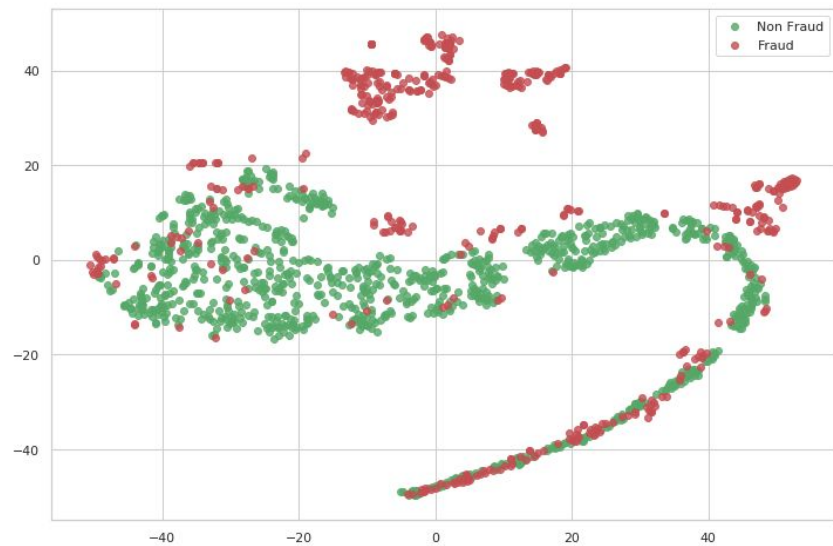
- t-SNE(t-Distributed Stochastic Neighbor Embedding)는 데이터의 차원을 줄이고, 최대 정보를 가진 상위 n 개의 component만 생성하는 데이터셋 decomposition 기술
- t-SNE를 이용해서 정상(녹색)/사기(빨간색) 거래의 특성을 시각화
- 두 축은 t-SNE에 의해 추출된 component



AutoEncoder



Representation (Z)



Simple Classifier

In [13]:

```
train_x, val_x, train_y, val_y = train_test_split(rep_x, rep_y, test_size=0.25)
clf = LogisticRegression(solver="lbfgs").fit(train_x, train_y)
pred_y = clf.predict(val_x)

print("")
print("Classification Report: ")
print(classification_report(val_y, pred_y))

print("")
print("Accuracy Score: ", accuracy_score(val_y, pred_y))
```

```
Classification Report:
              precision    recall  f1-score   support

         0.0         0.98        1.00        0.99        754
         1.0         1.00        0.87        0.93        119

 micro avg         0.98        0.98        0.98        873
 macro avg         0.99        0.94        0.96        873
weighted avg         0.98        0.98        0.98        873
```

```
Accuracy Score:  0.9828178694158075
```

노트북 필사 예시



<> Notebook

[House Price] Tutorial For Korean Beginners (한글커널)

by HONG_YP

3 years ago • 55s to run • Python • ^ 38



<> Notebook

영화 추천 시스템(Movie Recommender Systems)

by alsoj

2 years ago • 1m to run • Python • ^ 4

[(http://labs.criteo.com/wp-content/uploads/2017/08/CustomersWhoBought3.jpg)] 이 커널은 학습용으로 아래 커널을 필사/번역한



<> Notebook

How AutoEncoders Work - korean ver.

by Subin An

3 years ago • 2m to run • Python • ^ 18

이 커널(필사)에서는 오토인코더와 그 구현 방법을 설명합니다. 오토인코더는 이미지 데이터에 많이 사용되고, 다음과 같은 사례에서 사용할 수 있습니다.

[House Price] Tutorial For Korean Beginners (한글커널)

Notebook Data Logs Comments (22)

▲ 38

Copy & Edit 66



Introduction

본 커널은 다른 분들이 올려주신 커널들을 참고하여 만들었습니다.

Regression tutorial을 진행하기 좋은 집값 예측 문제를 titanic처럼 초보자 분들에게 도움될 만한 커널을 만들고자 하였습니다.

본 커널은 높은 점수, 등수를 얻기위함이 목적이 아닌 Baseline 수준의 튜토리얼 커널입니다.

초보자의 입장에서 초보자분들을 위해 커널을 작성하는 만큼, 쉽게 설명하고자 하였습니다.

저 또한 초보적인 실력이기 때문에 실수나 혹은 지적할만한 문제가 있다면 Feedback 해주시면 감사하겠습니다!

본 커널의 순서입니다.

1. Import & Data Check

- 데이터 확인
- 이상치 탐색 및 제거
- 결측 데이터 확인
- Skewness(비대칭도) 확인

Table of Contents



| House Price Predict

Import & Data check

EDA

Feature Engineering

In [20]:

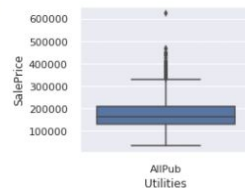
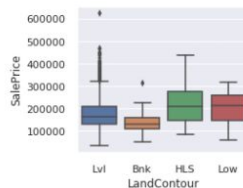
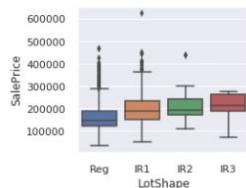
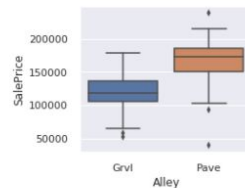
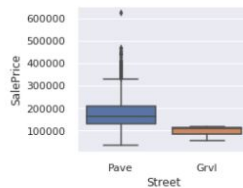
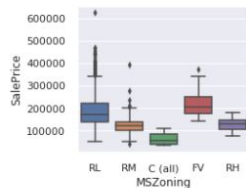
```
li_cat_feats = list(categorical_feats)
nr_rows = 15
nr_cols = 3

fig, axs = plt.subplots(nr_rows, nr_cols, figsize=(nr_cols*4,nr_rows*3))

for r in range(0,nr_rows):
    for c in range(0,nr_cols):
        i = r*nr_cols+c
        if i < len(li_cat_feats):
            sns.boxplot(x=li_cat_feats[i], y=df_train["SalePrice"], data=df_train, ax = axs[r]
[c])

plt.tight_layout()
plt.show()

# BoxPlot을 그려 Categorical Feature와 SalePrice의 관계를 확인합니다.
# 일부 범주는 다른 범주보다 SalePrice와 관련하여 더 다양하게 보입니다.
# Neighborhood 변수는 주택 가격 편차가 매우 크므로 영향이 크다고 생각합니다.
# SaleType 또한 마찬가지입니다.
# 또한 수영장이 있으면 가격이 크게 증가하는 것 같습니다.
# 정리하면 SalePrice에 영향을 많이 미치는 변수로는 'MSZoning', 'Neighborhood', 'Condition2', 'MasVnrType', 'ExterQual', 'BsmtQual', 'CentralAir', 'Electrical', 'KitchenQual', 'SaleType' 등이 있습니다.
```




그리고 양념 한 스푼

1. 데이터셋 바꾸기
2. 하이퍼 파라미터 변경
3. 앙상블
4. 최신 기법 적용
5. 나의 경험

Featured Prediction Competition

H&M Personalized Fashion Recommendations

Provide product recommendations based on previous purchases

 H&M Group · 1,663 teams · a month to go (a month to go until merger deadline)

\$50,000

Prize Money

Overview

Data

Code

Discussion


Leaderboard


Rules

Team

New Notebook

...

 Search notebooks

 Filters


All

Your Work

Shared With You

Bookmarks

Best Score ▾



H&M Ensembling[LB 0.0234]


Notebook copied with edits from AtulVerma · Updated 5d ago

Score: 0.0234 · 0 comments · H&M Personalized Fashion Recommendations +5

▲ 10

Bronze

...




H&M Ensembling - with LSTM

Notebook copied with edits from AtulVerma · Updated 7d ago

Score: 0.0232 · 0 comments · H&M Personalized Fashion Recommendations +3

▲ 4

...



H&M : Framework for Partitioned Validation


Updated 6d ago

Score: 0.0232 · 5 comments · H&M Personalized Fashion Recommendations +1

▲ 46

Silver

...



H&M Ensembling - with LSTM


Notebook copied with edits from Giba · Updated 12d ago

Score: 0.0232 · 10 comments · H&M Personalized Fashion Recommendations +5

▲ 46

Silver

...



H&M Ensembling - How to

Notebook copied with edits from a private notebook · Updated 23d ago

Score: 0.0231 · 11 comments · H&M Personalized Fashion Recommendations +5

▲ 80

Gold

...

Ablation study

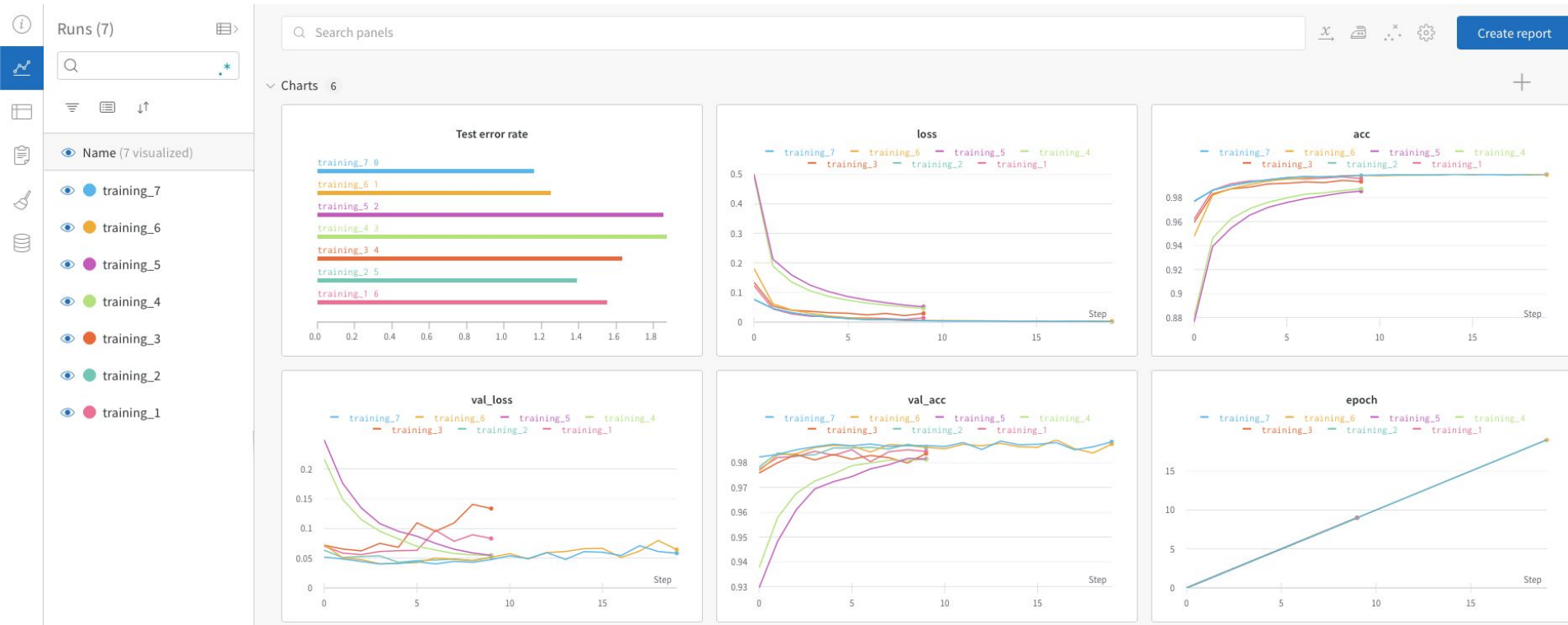
Hierarchically fused layers from:						mAP
conv4-3s	conv7s	conv8-2s	conv9-2s	conv10-2s	conv11-2s	
✓						27.6
✓	✓					37.1
✓	✓	✓				36.8
✓	✓	✓	✓			36.8
✓	✓	✓	✓	✓		36.6
✓	✓	✓	✓	✓	✓	36.0

w&b, neptune



Experiment tracking and model registry for
production teams

Weights & Biases (W&B)



Weights & Biases (W&B)



ANY FRAMEWORK



TENSORFLOW



PYTORCH



KERAS



SCIKIT-LEARN



HUGGING FACE



XGBOOST

```
import wandb

wandb.init(project="visualize-sklearn")

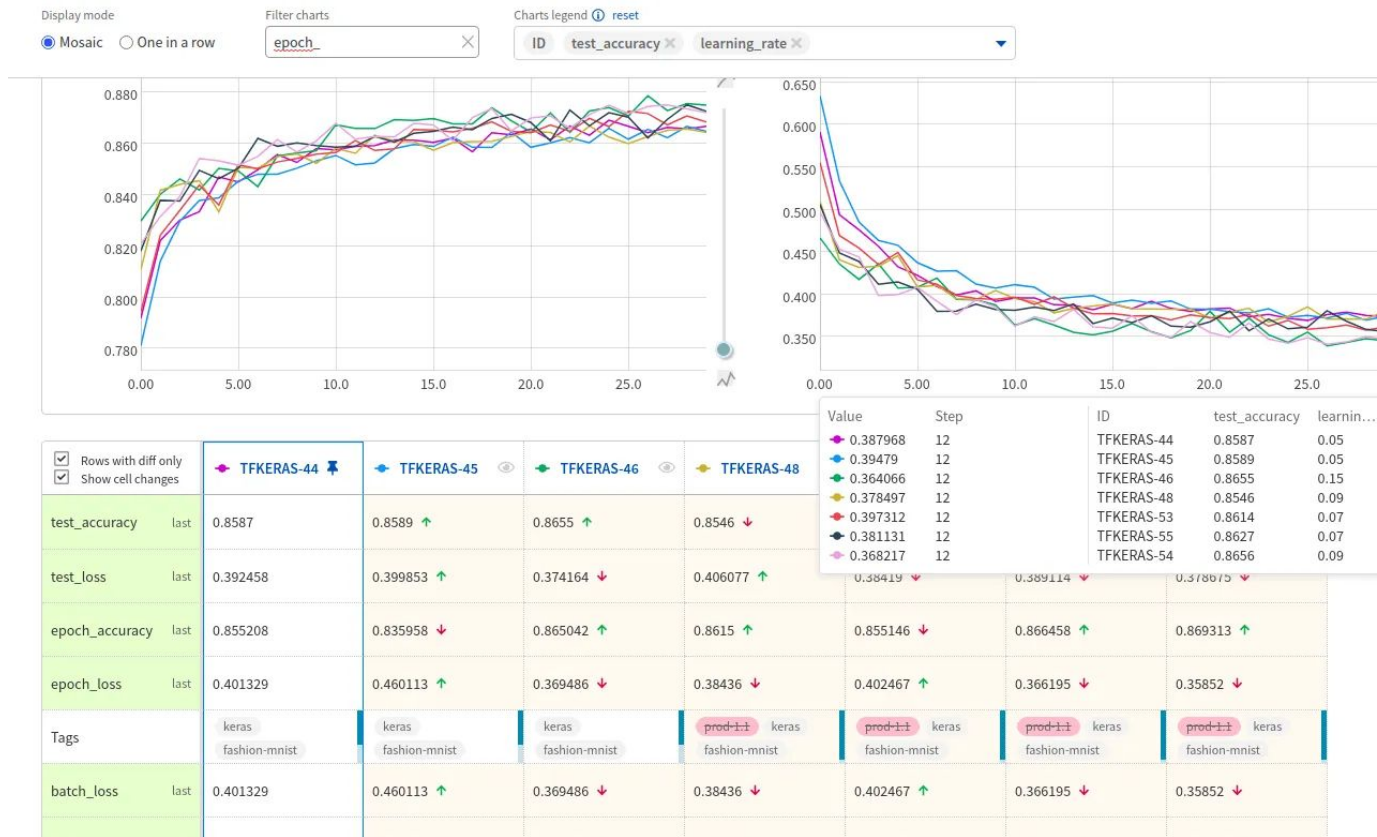
# Model training here

# Log classifier visualizations
wandb.sklearn.plot_classifier(clf, X_train,
                             X_test, y_train, y_test, y_pred, y_probas,
                             labels, model_name='SVC', feature_names=None)

# Log regression visualizations
wandb.sklearn.plot_regressor(reg, X_train,
                             X_test, y_train, y_test, model_name='Ridge')

# Log clustering visualizations
wandb.sklearn.plot_clusterer(kmeans, X_train,
                              cluster_labels, labels=None, model_name='KMeans')
```

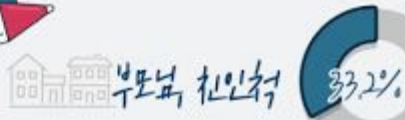
neptune.ai



2020년 취업 응원합니다!

(복수응답) 2019년 취업 준비 시
도움 받은 경험이 있는 곳은?

친구, 선배
49.4%



취업준비 Tip

취업준비 Tip

1. 코딩 테스트
2. 분명한 개성의 프로젝트

코딩 테스트

Category - All



programmers

탐프로그래머스 실력 체크 개발자 채용 테크 피드 코딩테스트 연습 프로그래밍 강의

[코딩테스트 고득점 Kit](#) [SQL 고득점 Kit](#) [모든 문제](#)

Pick One



Weekly Contest 233

Sunday, Mar 21
2:30 - 4:00AM UTC

Register

코딩테스트 고득점 Kit

코딩테스트에는 어떤 알고리즘/자료구조가 출제될까요?

사람들은 어떤 문제를 어려워할까요? 국내에서 코딩테스트를 가장 많이 운영해온 프로그래머스 팀이 [코딩테스트](#) 결과를 분석해서 만든 [고득점 Kit](#). 코딩테스트에 자주 나오는 유형, 사람들이 많이 틀리는 유형을 간추렸습니다.

해시

해시는 Key-value형으로 데이터를 저장하는 자료구조입니다.

출제 빈도
높음

평균 점수
보통

문제 세트
0 / 4

스택/큐

LIFO, FIFO, push & pop 스택과 큐를 이용해서 문제를 풀어보세요.

출제 빈도
보통

평균 점수
높음

문제 세트
0 / 4

정렬

정렬을 이용해서 문제를 효율적으로 풀어보세요.

출제 빈도
높음

평균 점수
높음

문제 세트
0 / 3

완전탐색

무식해 보아도 사실은 최고의 방법일 때가 있지요.

출제 빈도
높음

평균 점수
높음

문제 세트
0 / 3

code jam

What's involved?

Code Jam is Google's longest running global coding competition, where programmers of all levels put their skills to the test. Competitors work their way through a series of online algorithmic puzzles to earn a spot at the World Finals, all for a chance to win the championship title and \$15,000.

[Learn more](#)



분명한 개성의 프로젝트


GettingStarted Prediction Competition

Titanic - Machine Learning from Disaster

Start here! Predict survival on the Titanic and get familiar with ML basics

Kaggle · 14,142 teams · Ongoing

Overview Data Code Discussion Leaderboard Rules Team My Submissions **Submit Predictions** ...


 NBME - Score Clinical Patient Notes

Identify Key Phrases in Patient Notes from...

Featured

Code Competition · 965 Teams

\$50,000 1mo to go


 Happywhale - Whale and Dolphin Identification

Identify whales and dolphins by unique ch...

Research

1343 Teams

\$25,000 18d to go

 U.S. Patent Phrase to Phrase Matching

Help Identify Similar Phrases in U.S. Paten...

Featured

Code Competition · 359 Teams

\$25,000 3mo to go

Analytics Competition

2021 Kaggle Machine Learning & Data Science Survey

The most comprehensive dataset available on the state of ML and data science


Kaggle · 4 months ago

\$30,000 Prize Money

Overview Data **Code** Discussion Rules **New Notebook** ...


Q Search notebooks **Filters**

All Your Work Shared With You Bookmarks Most Votes ▾




How are the 🧑 Ladies and the 🧑 Gents doing?
Updated 4mo ago
65 comments · 2021 Kaggle Machine Learning & Data Science Survey +2

248
Gold ...




What Takes to be a Data Scientist: Story of Robert
Updated 4mo ago
69 comments · 2021 Kaggle Machine Learning & Data Science Survey +1

183
Gold ...




日本語でやってみたかった。Trying this in Japanese. [en sub]
Updated 4mo ago
29 comments · 2021 Kaggle Machine Learning & Data Science Survey

129
Gold ...




Kaggle survey EDA and visualization without plots
Updated 5mo ago
52 comments · 2021 Kaggle Machine Learning & Data Science Survey

106
Gold ...



What's up Kaggle? 🤔 [Kaggle Survey 2021]
Updated 5mo ago
63 comments · 2021 Kaggle Machine Learning & Data Science Survey

103
Gold ...



Kaggle ML & DS Survey Analysis 📊
Updated 5mo ago
27 comments · 2021 Kaggle Machine Learning & Data Science Survey

100
Gold ...

분명한 개성의 프로젝트

타겟팅

- 원하는 회사에 맞게
- 원하는 직무에 맞게
- 원하는 미래에 맞게
- 원하는 타이밍으로

커뮤니티 활동

- 최신 키워드 습득
- 스터디 그룹
- 팀 프로젝트
- 다양한 행사 참여

네트워킹

- 회사 내부 정보 습득
- 추천 입사
- 이직에 도움될
가능성
- 가스라이팅 방지

타겟팅

[Blind] 2022년 IT업계 테크직군 초봉 / 86개 기업 / 학사 신입 연봉 기준								수정일
Copyright 2021. 2022. Blind IT엔지니어 라운지 일동. All rights reserved.								2022/04/01 03:17:49 PM
기업명	분야	계약 종류	보너스	RSU	스톡옵션	연봉 인상률	포괄	비고
구글 코리아	검색 / 미디어	6000	15%	5500만	?	?	O	L3 기준, RSU는 1년 클리프 당 금액으로 총 4년
퀄컴 코리아	통신	추정 6200	?	?	?	?	?	
SAP랩스 코리아	DB / 클라우드	추정 5000	0 ~ 3%	?	?	?	X	2021년 전직원 월급 2% 인상
오라클 코리아	DB / 클라우드	최소 5000	?	?	?	?	?	
크래프톤	게임	원전 8000	?	?	?	?	O	
NCSOFT	게임	원전 6000	최소 800	?	?	?	X	복포 250만
넥슨	게임	원전 6000	?	?	?	?	X	복포 250만, 게임 프로그래밍 및 분석 외 테크 직군 4500
넷마블	게임	5000 ~ 6000	?	?	?	?	X	복포 250만, 식대 240만
NHN	게임	5000 ~ 5500	?	?	?	?	O	식대 약 200만
풀어비스	게임	4300 ~ 4400	?	?	?	?	X	현직 800만 인상 및 200만 보너스
룩업스	게임	4100 ~ 4200	10 ~ 15%	?	?	?	?	복포 240만, 논테크 구분 없음
스마일게이트	게임	4000 ~ 5000	0 ~ 10%	?	?	평균 2.5%	X	현직 평균 800만 인상
베이글코드	게임	3900 ~ 4100	?	?	1100만	?	?	
베스파	게임	3900 ~ 4000	?	?	?	?	?	
11번가	커머스	6500	400 ~ 500	?	?	200 ~ 300	O	복포 400만, 현금성 약 500만
브랜드	커머스	6000	0	?	?	?	?	현금성 240만
위메프	커머스	6000	?	?	?	2 ~ 3%	X	
티켓몬스터	커머스	6000	?	?	?	?	?	
우아한형제들	배달	5500	최소 14%	?	?	?	X	주 32시간, 주식 중여 평균 5000만원, 이직 시 사이닝 보너스 직전 계약 연봉 20%
달리버리히어로 코리아	배달	5300	?	?	?	3%	?	2021년 연봉 인상률 3배
당근마켓	중고거래/금융	5100	없음	?	있음	?	O	사이닝 보너스 있음
비바리퍼블리카	금융	5000	최소 10%	있음	-	3%	X	통신비 월10, 체력단련비 월10, 점심/저녁비용 법인카드 100%지원
두나무	금융	5000	있음	?	10%	?	?	초봉 SKY 기준
뱅크샐러드	금융	5000	있음	-	-	?	?	리드급 스톡옵션 최소 1억, 경력 이직 기본급 1.5배
직방	부동산	5000	?	?	?	?	O	이직 시 사이닝 보너스 직전 계약 연봉 100% (최대 1억), 원격 근무
리디북스	전자책	5000	?	?	?	?	?	시니어/PM 이직 시 사이닝 보너스 5000만, 직전 연봉 30% 인상 (현직 포함)
야놀자	레저	5000	?	?	?	?	O	우리사주 1000만, 사이닝 보너스
디어유	메신저	5000	?	?	?	?	?	
쏘카	카셰어링	5000	?	?	?	8 ~ 10%	?	
그린랩스	스마트팜	5000	?	?	있음	?	?	이직시 직전 계약 연봉 최대 130%, 최대 5000만원 사이닝 보너스
네이버	검색 / 플랫폼	5000	15 ~ 25%	1000만	1000만	3 ~ 10%	X	2021년부터 3년 간 연 1000만원 상당 주식
카카오	메신저 / 플랫폼	5000	5 ~ 20%	500만	?	?	X	현직 200 ~ 300만 인상
카카오엔터테인먼트	컨텐츠	5000	5 ~ 20%	500만	있음	3 ~ 10%	X	현직 500만원, 안식 휴가비 200만원, 복포 360만, 명절 100만, 통신비 지원
라인플러스	메신저 / 플랫폼	5000	평균 15%	있음	있음	3 ~ 10%	X	영구 재택, 현직자 리텐션보너스 500 ~ 2000만, 복포 256만, 신입 사이닝 보너스
TmaxSoft	미들웨어	5000	0	?	?	?	?	연구원 기준 (이후 5년 동결)

타겟팅

ML Engineer



Coupang
South Korea, 서울특별시

Apply on 프로그래머스

🕒 Over 1 month ago 🏢 Full-time


• 쿠팡은 세계에서 가장 빠르고 크게 성장하는 Ecommerce 기업입니다. 우리는 고객, 직원, 파트너 그리고 우리를 둘러싼 모든 사람들의 일상을 어떻게 혁신할 수 있을지 매 순간 고민합니다. 우리는 아직 아무도 풀지 못한 문제를 해결함으로써 사람들이 이렇게 묻는 세상을 만들고자 합니다. "쿠팡 없이 어떻게 살았을까?" 쿠팡은 서울뿐만 아니라 베이징, LA, 시애틀, 상하이와 실리콘밸리 등에 오피스를 두고 있는 글로벌 기업입니다.

FLEX팀은 쿠팡맨 및 쿠팡 플렉서의 로켓 배송 업무 지원을 담당하고 있습니다. 배송 관리 시스템에서부터 배송 모바일 웹을 제공하여 로켓 배송의 엔드 투 엔드(end to end)의 비즈니스를 구현하여 수준 높은 배송 서비스를 제공하고 있습니다.

업무 내용:

- Java, Kotlin 등을 활용한 Back-end 비즈니스 어플리케이션 설계 및 개발
- 여러 도메인과 dependency를 도출하고 시스템 연동 개발
- 배송 경로 최적화 관련 ML 알고리즘 및 어플리케이션 개발
- 머신 러닝 시스템 디자인 및 기존 ML library, framework 확장

커뮤니티 활동



Google 검색

로그인

회원가입

Q&A

Tech

커뮤니티

칼럼

Jobs

Github Issues

Editor's Choice

- 2022 전자정부표준프레임워크 컨트리어뷰션 참가자 모집
DPIT # 1K 4일 전
- 개발을 시작하려는, 시작한 분들을 위해 써보는 글입니다.
자비스를만들고... # 749 13일 전
- SI,정부과제 vs 서비스 vs 솔루션 장단점
하마 # 8K 18일 전
- 한 줄 더 없어서 안줄은 이력서가 되는 경우 (+ 무료로 이력서 첨삭해드립니다)
후하하 # 8K 1일 전
- 신입들의 큰 착각: 님이 하는 언어를 10년 후에도 쓰고 있을까요?
후하하 # 8K 2일 전

Weekly Best

- (장문주의)편일하다가 깨달
30대 비전공 문과출신 백
인프런 강의 올리고 기본
MSA는 80%정도의 희망
Q. 불게 팔까요, 갠게 팔
하세요.

Q&A

- [개발 로직] 블로그, 동영상 등의 게시글을 특정 사용자에게만 접근 허용하도록 로직을 구현하는 방법
남해발바다 # 451 2분 전
- java while 반복문 질문!
Effortvij # 10 29분 전
- API 카카오톡 주소창 질문
개조보1123 # 30 41분 전
- 자바 스프링에서 dto 순환참조에 대해서 궁금한게 있습니다.
jinwoo lee # 52 43분 전
- 자바스크립트 중복값 제거
검색도검색 # 20 1시간 전

가짜연구소 (Pseudo Lab)

공개 그룹 · 멤버 526명

그저 첫 발걸음을 떼면 된다. 계단 전체를 올라다 볼 필요도 없다.
그저 첫 발걸음을 떼면 된다. - 마틴 루터 킹

가짜연구소 디스코드: <https://discord.gg/HeHbFAvmSZ>

이루고 싶은 것, 해야할 것, 해야만 하는 것들을 바라볼 때면.
너무나도 높아 보이고, 멀어 보이고, 심지어 수 많은 장애물에 가로막혀 보이지 않을 때도 있습니다.
그저 첫 발걸음을 떼면 되는 데, 그 움직임이 참 어렵죠.

저희 공동체의 움직임도 그렇습니다.
첫 발걸음이 필요합니다!

이 글이 누군가의 삶을 송두리째 변화시키는 엄청난 존재가 되지는 못하더라도, 누군가의 열정에 불을 붙이는 불씨와 같은 역할, 기나긴 삶의 여정에서의 첫 발걸음에 도움이 되는 역할이 되었으면 좋겠습니다.

가짜연구소는 기존에 슬랙을 기반으로 커뮤니티를 형성하였습니다. 천여명이나 함께하고 계신데, 디스코드로 이전을 시작했습니다! 가짜연구소의 여러분들의 첫 발걸음은 무엇인가요? 기대하겠습니다!

가짜연구소 디스코드: <https://discord.gg/HeHbFAvmSZ>



DISCORD.COM

Join the PseudoLab Discord Server!

Check out the PseudoLab community on Discord - hang out with 116 other members and enjoy free voice and text chat.

질문-게시판

안녕하세요! 최근에 clustering 진행해 보신 분 계신가요? text와 categorical 변수를 함께 활용하고싶는데 레퍼런스를 찾지못해서 질문드립니다. 랜질은 방법이나 레퍼런스 자료가 있을까요?

안녕하세요! 최근에 clustering 진행해 보신 분 계신가요? text와 categorical 변수를 ...

어떤 데이터를 사용하실까요? 제가 document clustering은 많이 해봐서 도움 드릴수 있을 것 같습니다 😊

어떤 데이터를 사용하실까요? 제가 document clustering은 많이 해봐서 도움 드릴수...

최종적으로 컨텍스트 클러스터링을 하는게 목표이고 사용하는 데이터는 해당 컨텍스트의 text데이터(제목, 본문), 컨텍스트의 종류 및 특징, 작성자 정보등이 될 것 같습니다!

텍스트 클러스터링 메시지 5개, 스레드에 새 메시지가 없어요.

님! 텍스트 클러스터링 스레드를 시작했어요(스레드 모두 보기). 2022.03.28.

안녕하세요. 혹시 주피터 노트북 파일을 복사해서 쓸 때 copy와 duplicate의 차이... copy 기능을 쓰면서 원본이 변경이 되었던 경우는 못보던거 같긴합니다!

copy 기능을 쓰면서 원본이 변경이 되었던 경우는 못보던거 같긴합니다!

감사합니다 😊

2022년 3월 29일

2022.03.29

대학원 진학 준비를 지금부터 하려고 하는데 혹시 포트폴리오는 대부분 자유 방식인가요? 포트폴리오에 어떤 내용이 들어가야되고, 어느정도에 디테일을 작성해야하는 지 모르겠습니다 T

대학원진학 메시지 12개, 스레드에 새 메시지가 없어요.

님! 대학원진학 스레드를 시작했어요(스레드 모두 보기). 2022.03.29.

2022.03.29

혹시 맥북M1으로 학습들리시는 분이 계신가요? 계시다면 어떤 프레임워크(tensorflow, pytorch) 쓰시는지 궁금합니다!

맥북M1 딥러닝 프레임워크 메시지 10개, 스레드에 새 메시지가 없어요.

님! 맥북M1 딥러닝 프레임워크 스레드를 시작했어요(스레드 모두 보기). 2022.03.29.

2022.03.29

여재 오진 10:57


혹시 라이엇게임즈 api로 재미삼아 해보신 분 계신가요..

여재 오진 10:59

혹시 라이엇게임즈 api로 재미삼아 해보신 분 계신가요..

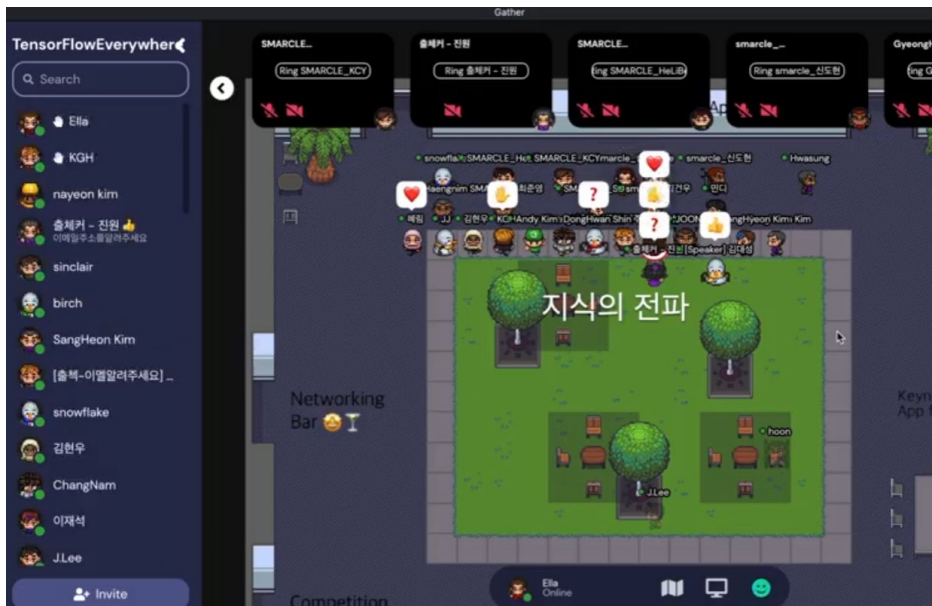
애전에 api call을 한두번? 정도 날려 보았습니... ㅎㅎ 라이엇게임즈 third party 개발자 커뮤니티 디스코드 채널도 별도로 있으니 참고하시면 좋을 것 같습니다 ㅎㅎ <https://discord.gg/riotgamesdevrel>

서버에 초대받았어요

 Riot Games Third Party Developer
4,940명 온라인 • 멤버 19,136명

참가

네트워킹



#Weekly_AI_Arxiv #클럽하우스 #Stanford_AI_Index_2022 특집

매주 일요일 밤 10시에 한주간 재미있는 AI뉴스와 연구 소식을 깃털처럼 가볍게 논의하는 클럽하우스 채널인 Weekly AI ArXiv 45회차가 3월 27일에도 진행 됩니다.

지난 주 예고헤드린대로 이번주는 Stanford AI Index Report 2022 (<https://hai.stanford.edu/research/ai-index-2022>)를 모더레이터 분들이 공유해주실 예정입니다.

각 챗터 별로 아래와 같습니다.

- 👉 Ch.1: Research & Development (김영도)
- 👉 Ch.2: Technical Performance (이준형)
- 👉 Ch.3: Technical AI Ethics (Nick Lee)
- 👉 Ch.4: The Economy and Education (이경훈)
- 👉 Ch.5: AI Policy and Governance (하정우)
- 👉 Discussion & Feedback: 전중홍, 이진원, 이수정, Terry Taewoong Um

많이 청취해주시고~ 늘 그럴듯이 야심한 밤이 부담스러우신 분들은 다시듣기 가능합니다! 내일 컨디션이 좋기를 비나이다 비나이다~

👉 자료: <https://github.com/jungwoo-ha/WeeklyArxivTalk/issues/45>

👉 클럽하우스: <https://www.clubhouse.com/.../naver.../NP553IbN/xpaN19vL...>



CLUBHOUSE.COM

Weekly AI arXiv 45회차 - Stanford AI Index 2022 특집 - NAVER AI & CLOVA

Sunday, March 27 at 10:00pm KST with Jung-Woo HA, Jonghong Jeon, Joonhyung Lee, Terry T...

Q & A