



Contact

☎ 010-6577-3479

✉ taraki3639@gmail.com

Projects & Work Experiences

탁성원

최종수정일 : '22-01-18

▶ About Me

- 도전과 몰입을 즐기는 사람
- 소통과 공유를 중요시하는 사람
- 문제해결을 즐기는 사람

A word cloud visualization of the text from the 'About Me' section. The words are arranged in a cluster, with '공유' (Share) and '도전' (Challenge) being the largest and most prominent. Other words include '문제해결' (Problem Solving), '몰입' (Immersion), '소통' (Communication), '수학' (Mathematics), '퍼즐' (Puzzle), '보드게임' (Board Game), '스도쿠' (Sudoku), and '책임감' (Sense of Responsibility). The colors used are primarily purple, blue, and teal.

공유 도전 문제해결 몰입
수학 퍼즐 보드게임 스도쿠 소통
책임감



<https://github.com/SeongwonTak>



<https://taksw222.tistory.com/>

탁성원 | Daram

● Work Experiences

Kakao, Intern DataInsight cell

2021.06.28 ~ 2021.08.31

- 이모티콘 플러스 서비스 현황 분석
- 서비스 신규 가입 유저 예측 Project

● Educations

서울대학교 수리과학부 | 석박통합과정 중퇴

2016.03 ~ 2018.03

중앙대학교 수학과 | 학사 졸업

2012.03 ~ 2016.03

● Skills / Tools

DATA

Python(Pandas, Scikit-Learn) | SQL

OTHERS

Slack, Jira, Git/Github

● Other Experiences

Dacon Basic I 따름이 이용량 예측 1위

2021.06.28 ~ 2021.08.31

- 이모티콘 플러스 서비스 현황 분석
- 서비스 신규 가입 유저 예측 Project

(비개발) 한국전력공사 근무

2019.07.01 ~ 2021.03.30

- 총무, 고객만족도, 수금파트 근무
- 데이터 분야로의 전직을 위한 퇴사

(비개발) 킷출판사 근무

2018.03.08 ~ 2018.09.08

- 수학 도서 교정 교열 및 편집
- 초등부 -> 고등부 변경을 위한 퇴사

▶ Work Experiences

● Kakao Internship

이모티콘 플러스 서비스 현황 분석 & 신규 가입 유저 예측
데이터 인사이트셀 근무, '21.6.28 ~ '21.8.31

유저 집단 구분 및 비교에 따른 주요 변수 추출, 추가 변수 발굴을 통한
모델 성능 개선 경험 (Baseline f1 score 0.3 → 변수 발굴 후 개선 후 f1 score 0.5)



서비스 이용 유저

서비스 이탈 유저

서비스 비이용 유저

Pandas 활용
추가 전처리

SQL 활용
데이터 추출

이모티콘 발화량,
구매 경험 등
주요 변수 발굴

이모티콘 **발화 영향력**,
나를 위한 이모티콘 구매
등의 **추가 변수 발굴**

**회귀와 분류 모델
복합 사용**

불균형 데이터
SMOTE 적용 시도

동일한 발화량에서도
홍보의 수신 여부,
스토어 접속 여부등이
높은 중요도 가짐을 확인

● 따릉이 이용량 예측

Dacon Basic 대회 참가, 따릉이 이용량에 대한 예측 (최종 1위)
Competitions, '21.11.1 ~ '21.11.12

'18년~'20년 4~6월 일간 따릉이 이용량으로 '21년 동 기간 일간 따릉이 이용량 예측 문제

불쾌지수, 일교차, 추운 정도 등의 **기후적 요소**와 **연도별 상승분 반영**을 통한 문제 해결
최초 모델 NMAE 0.58902 → 최종 NMAE 0.25052로 1위 기록

추가 변수 발굴

습도+ 기온 → 불쾌지수

풍속 + 날씨 → 탑승 어려움 지수

일교차 및 풍속 고려 추움 지수



날씨가 나쁜 상황에
이용량이 매우 낮음

모델링

LightGBM Regressor를 활용한 예측

따릉이 이용량 상승 추세 반영
전년도 상승분 적용

● LoL 승부예측 프로젝트

LoL 게임 결과 데이터 바탕, Blue팀과 Red팀 중 승리팀에 대한 예측
Personal Project, '21.4월초, '21.9.27 ~ '21.10.24(개선작업)

상대적으로 비교할 수 있는 변수 생성 후 LightGBM을 통해 **97%** 정도의 정확도로 예측 확인
Decision Tree의 시각화를 통한 주요 7변수 확인 및 이를 반영한 다수결 모델 검토

Remark. 주요 7변수 : 타워 격파 수, 적 챔피언 킬 수, 챔피언 딜량, 미니언 킬 수,
최초 억제기 격파 여부, 억제기 격파 수, 오브젝트 딜량

전처리

Blue vs Red

인게임 동일 지표 비교
값이 더 우수한 팀에 1 부여

분석 / 모델링

결정트리 시각화 통한
중요 변수의 파악

LightGBM 활용 모델링

검증

다른 랭크에도
동일한 모델이 적용되는지 확인

파악한 주요변수들로
다수결 예측기 생성 및 검증



Personal Projects

● 보드게임 추천 프로젝트

보드게임 데이터 기반 보드게임 추천 프로젝트
Personal Project, '21.6월초

플레이 인원수 및 경험한 보드게임 3개의 입력에 따라
경험한 보드게임과 유사한 새로운 보드게임 10개 추천

```
top_list = [idx1, idx2, idx3]
score_list = []

for i in range(0, 2347):
    if i in top_list:
        score_list.append((0, i))
    else:
        score = 0
        score += (pearsonr(category_arr[idx1], category_arr[i])[0])
        score += (pearsonr(category_arr[idx2], category_arr[i])[0])
        score += (pearsonr(category_arr[idx3], category_arr[i])[0])
        score += (pearsonr(mechniac_arr[idx1], mechniac_arr[i])[0])
        score += (pearsonr(mechniac_arr[idx2], mechniac_arr[i])[0])
        score += (pearsonr(mechniac_arr[idx3], mechniac_arr[i])[0])
        score_list.append((score, i))

score_list.sort(reverse = True)
score_list = sorted(score_list[:100], key = lambda x: x[1])
```

대상 및 변수 선택

플레이자들의 평가 수가
1,000건 이상인 보드게임 대상

인원 수, 매커니즘, 장르, 난이도
4가지 사항에 대해 고려

추천 모델링

인원수에 맞는 보드게임 및
경험한 보드게임에서 난이도가 크게 벗어나지 않게 범위 설정

매커니즘과 장르에 유사도 적용 추천



Personal Projects

● 체지방 예측 Model-serving Project

체력 데이터 및 건강검진 데이터 기반 체지방 예측 Model Serving
Personal Project, '21.12월 초

공공 API를 활용한 데이터 수집, 간단한 회귀모델 작성
해당 모델 저장 후 Flask를 통한 Serving 시도

Fat rate(체지방) Prediction

나이 :

키 :

몸무게 :

제자리멀리뛰기 기록(cm) :

셔틀런 기록(개) :

예측 체지방 :

33.81%

참고 : 체지방율별 세부 설명

ACE Body Fat % Chart		
Description	Women	Men
Essential fat	10-13%	2-5%
Athletes	14-20%	6-13%
Fitness	21-24%	14-17%
Average	25-31%	18-24%
Obese	32%+	25%+

[다시 예측하기 가기](#)

데이터 수집

공공API를 통한
체력검정 데이터 수집

연령,

모델 작성 / Flask 활용

선형회귀를 활용한 간단한 모델 작성

Flask를 활용한 예측 모델 연결

프로젝트 발전 방향

Front-End 부분 적용 및 반영

데이터 저장을 통한
지속적 체지방 관리 기능 추가