

About Me

논리적인 분석으로 사회에 효율적인 방향을 제시할 수 있는 데이터 분석가를 꿈꾸는 유승희입니다.



Introduction

- Regression, RF, XGBoost 등 ML을 활용한 공공자전거 입지추천, 교통사고 위험지역 도출, 음식물 쓰레기 배출량 예측 경험
- BiLSTM, BERT 등 딥러닝을 활용한 텍스트 분류, 감정분석 등 NLP 분석 경험



Contact & Channel

☎ 010-2549-9251

✉ juneyo0823@naver.com

🔗 <https://github.com/JuneSHYoo>



Projects

MBTI별 투자성향 분석 챗봇

데이터 분석 동아리 BOAZ
(2022.01 ~ 2022.07)

- Python, Flask**
- MBTI별 개인 맞춤형 투자 성향 분석 및 가이드라인 제시 챗봇 제작
- 금/부정 답변 유도하는 질문 data 구성 및 약 3000여개의 Dataset 수집과라벨링
- 카카오 i 오픈빌더를 활용한 챗봇** 채널 및 시나리오 제작
- 대화내용 데이터 추출 및 분석을 위한 **웹서버 생성 및 스킴 등록 - Flask & Python** 사용
- 서버로 받은 사용자 답변 전처리 후 **BiLSTM을 이용한 Sentiment Analysis** - test accuracy 79.45
- 사용자 유형 점수 계산 및 결과 제시

한국어 뉴스 토픽 분류

데이터 분석 동아리 BOAZ
(2022.11 ~ 2022.12)

- Python & Keras, Pytorch**
- 한국어 뉴스 헤드라인 주제 분류하는 알고리즘 개발
- 특수문자, 한문 등 자연어 처리 후 **Random Over Sampling**으로 데이터 불균형 문제 해결
- Fasttext를 이용한 word vectorization**
- keras sequential 로 BiLSTM 층 구성, pytorch로 KoBERT 구현 및 성능 비교
- model accuracy : {'BiLSTM' : 0.76 , 'KoBERT' : 0.82}

음식물 쓰레기 배출량 예측

BigContest 2021
(2021.07 ~ 2021.09)

- Python**
- 2021 7,8월 제주시 읍면동별 음식물 쓰레기 배출량 예측
- 데이터 시각화, EDA 기반으로 Feature Engineering 진행
- ARIMA 모델을 사용하여 예측시기의 Feature dataset** 구성
- CatBoost, LGBM, GBR 모델링 학습 및 stacking 진행

어린이 교통사고 위험지역 도출

COMPAS

(2020.12 ~ 2021.1)

- Python, R
- 오산시 어린이 교통사고 위험지역 20개소 제시 및 스쿨존 내 교통안전시설물 우선 설치 지역 20개소 제시
- Isolation Forest로 분류한 이상치(-1)를 y변수로 두어 **Random Forest 활용**
- 변수 중요도 기반 위험도 점수 계산
- 다중공선성 제거후 **stepwise 회귀 분석**을 통한 독립 변수 선택

따릉이 대여소 입지추천

중앙대 경영경제대학 학술제

(2020.09 ~ 2020.11)

- R
- 서울특별시 따릉이 대여소 입지 추천
- 공공데이터 수집 및 탐색, 과제에 맞게 데이터 변형, 결과 시각화 경험
- 다중회귀분석의 **후진제거법**으로 유의한 독립변수 파악



Awards

2021 데이터 분석 공모전

중앙대 통계학과 분석 공모전

(2021.09 ~ 2021.11)

- Python
- 농수산물 가격 예측 모델 해석 및 기후와 가격 간의 인사이트 도출
- 공공 데이터 수집 후 Pandas를 이용한 데이터 EDA, 시각화
- 시차의 영향을 반영하기 위한 각 품목별 lag값 변수 추가 - **pacf 그래프를 통해 window size 결정**
- LSTM 분석 후 모델의 개별 예측값을 설명하기 위한 **LIME 알고리즘 사용**
- 대상 수상

통계학과 학생회장 장학금

(2020.12 ~ 2021.11)

- 코로나로 인한 비대면 학과 행사 기획
- 교내 공공 데이터 분석 공모전 개최



Certificate

- 데이터분석준전문가 (ADsP)
- 컴퓨터 활용 1급



Skills

- Python, R, SQL



Education

- 중앙대학교 통계학과 | 학점 4.11 / 4.5 | 전공평점 4.25 / 4.5