분석으로 유익한 사회를! 김동우입니다.

About Me

Introduction

- 회귀 분석을 사용한 카쉐어링 스테이션, 자동차 극장 입지 추천과 KBL 승률 예측 경험
- 딥러닝을 활용한 Semantic Segmentation, Object Detection, 문장 내 개체 간 관계 분류, Sequential data 분석 경험
- 학습한 모형으로 재학습 가능한 무중단 배포 파이프라인 구축 경험

Contact & Channel

- Phone | 010-3106-7593
- Email | zxcvbnm1997@hanmail.net
- Github | https://github.com/ddooom



중앙대학교 통계학과

- 소프트웨어학부 복수전공
- Major GPA | 4.2 / 4.5



Naver BoostCamp Al Tech

Deep Knowledge Tracing Competition

2021.05 ~ 2021.06

<u>Github</u>

- 목적 | 학생의 문제 풀이를 바탕으로 지식 상태를 추적하고 마지막 문제의 정답 여부를 예측하는 Task
- 목표 학습된 모델 Serving
- 데이터 EDA를 통해 30여 개의 Feature 생성, LGBM에 적용
- BentoML를 이용한 Inference API 생성
- Flask을 이용하여 BentoML과 클라이언트 랜더링
- 재학습 가능한 모델 배포 파이프라인 구축

재활용 품목 분류 Segmentation, Object Detection Competition

2021.04 ~ 2021.05

<u>Github</u>

- 목적 | 이미지에서 11가지 쓰레기 종류를 분류하는 Task
- 목표 | mmdetection을 사용하지 않은 Baseline 만들기
- 팀원과 실험 공유를 위한 config file, wandb 등 실험 관리 툴 사용
- 특정 이미지의 한 개체를 다른 이미지에 추가하여 데이터 증강 (Classmix)
- 데이터의 Class Imbalance와 모델 일반화를 위한 Loss, Augmentation 실험
- Seg private mIoU 0.6892, 3등

문장 내 개체 간 관계 추출 Competition

2021.04

Github

- 목적 | 문장 내 개체 간 42가지 관계를 추출하는 Task
- 목표 | Huggingface를 사용한 Pytorch Baseline 만들기
- 개체 인식에 도움을 주는 Additional Special tokens와 Entity embedding를 추가하여 학습
- 개체 간 관계 추출 테스크를 Question Answering 테스크로 변형하여 학습 시 Γ
- Baseline model accuracy 71.5에서 78.7로 성능 향상



2020 빅콘테스트 퓨처스리그

2020.08 ~ 2020.11

대회 상세정보

- KBO 정규시즌 팀별 승률, 타율 및 방어율 예측
- Pandas를 이용한 데이터 EDA, 시각화
- 데이터의 높은 분산으로 데이터를 Aggregation하여 WLS로 예측
- 우수상 (빅데이터포럼의장상) 수상

통계학과 부학생회장 장학금

2019.12 ~ 2020.11

- 코로나로 인한 온라인 대체 행사 진행
- 제 1회 경영경제대학 학술제 개최

2017 서울특별시 빅데이터캠퍼스 상시 공모전

2017.12 ~ 2018.01

<u>대회 결과</u>

- 서울특별시 카쉐어링 스테이션 입지 추천
- 공공 데이터를 수집 및 탐색하여 분석에 맞게 변형하고 결과를 시각화 파이프라 인 경험
- R 을 사용하여 데이터 변형, 시각화
- 회귀분석의 Coefficient로 독립 변수의 중요도를 파악
- 우수상 수상



Machine Learning

- Python
- · Pytorch, Pandas, Numpy
- R

Model Serving Framework

- BentoML
- Flask