논리가 기반이 되는 분석을 지향하는 김동우입니다.

About Me

Introduction

- 회귀 분석을 사용한 카쉐어링 스테이션, 자동차 극장 입지 추천과 KBL 승률 예측
- 딥러닝을 활용한 Semantic Segmentation, Object Detection, 문장 내 개체 간 관계 분류, Sequential data 예측 모델링 경험
- 번개장터, 카카오 추천팀 인턴, 쏘카를 통한 분석, 지표 추출, 예측/추천/자연어 모 델링 경험
- 현재 넥슨 재직 중

Contact & Channel

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

Skills

- · Python, Pytorch, R
- SQL



Work Experience (<u>Details</u>)

쏘카

2022.05 ~ 2022.09 | Data Scientist

- 데이터 분석을 통한 차량 배터리 방전 지표 정의
- 차량 배터리 방전 예측 모델링을 통해 차량 운영팀에게 방전 확률이 높은 차량 정 보 제공
- 예측 모델 사용시 장애 조치 비용 20% 감소, 선조치 차량 수 증가

카카오 추천팀 인턴

2022.01 ~ 2021.02 | 분석직군

- 데이터 분석을 통해 검수 관련 컬럼의 요소와 비중의 관계를 파악하고 검수 관련 컬럼과 피드백 데이터를 통해 **콘텐츠 품질의 정량적 지표 설정**
- BERT를 통한 text embedding과 feature engineering을 통해 품질 지표를 예 **측**하고 안전하게 **고품질 컨텐츠를 샘플링**하는 로직 설계

번개장터 인턴

2021.09 ~ 2021.12 | Data Scientist

- 광고 상품의 배치 변경을 위해 위치별 클릭 수와 매출 추정 로직 설계, 분석을 통 해 제시한 변경안으로 AB Test 진행하여 광고 수익 100% 상승 결과 도출
- 검색어 오타 교저을 위해 Transformer Encoder, Decoder를 이용한 오타 생성
- GNN을 통한 Content Based 추천 모델링 진행, 기존 데이터셋을 그래프로 변환 하고 상품 메타 데이터에서 feature 추출, DGL 라이브러리로 모델링 파이프라인 설계

Projects

Naver BoostCamp AI Tech

Deep Knowledge Tracing Competition

2021.05 ~ 2021.06 | Github

- 목적 | 학생의 문제 풀이를 바탕으로 지식 상태를 추적하고 마지막 문제의 정답 여 부를 예측하는 Task
- 목표 | 학습된 모델 Serving
- 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 | Github

- 목적 | 이미지에서 11가지 쓰레기 종류를 분류하는 Task
- 목표 | mmdetection을 사용하지 않은 Baseline 만들기
- 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 추 가하여 학습
- 개체 간 관계 추출 테스크를 Question Answering 테스크로 변형하여 학습
- 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

2017 서울특별시 빅데이터캠퍼스 상시 공모전

2017.12 ~ 2018.01 | 대회 결과

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