

윤도연

Data Scientist

ydy89899@gmail.com



Be a sniper, not a machine gun

안녕하세요. 수학을 좋아하며 AI를 공부하는 윤도연입니다.

현업에서 필요한 수학적 사고와 공학적 사고의 경계를 명확히 할 수 있습니다.

다양한 도메인(의료, 금융, 자연어, 추천시스템)에 관심이 많으며, 몇몇 Toy 프로젝트 경험을 통해 해당 도메인에서의 Key considerations을 이해할 수 있었습니다.

현재 재직중인 회사에서는 추천솔루션 및 자연어 task를 담당하고 있습니다.

현업에서의 추천시스템을 경험하면서, 그리고 최근 MLM의 LLM화와 같은 흐름을 보면 MLOps에도 관심이 생겼습니다.

해당 분야에서도 역량을 갈고 닦아 궁극적으로는 모델링, MLOps를 가리지 않는 ML Engineer로 성장하는 것이 목표입니다.

업무 경험

라이앤캐쳐스

2021.11 ~ [재직 중](#)

Data Scientist · AI팀

- 💡 와인나라 추천 솔루션 개발
 - User/Contents based 추천 시스템 구현
 - Model Serving 구현
- 💡 사내 문체 변환 토이 프로젝트 모델링
 - KoBART 모델을 활용한 문체 변환
 - Bentoml 모델 서빙
 - 도커라이징
- 💡 교보문고 추천 솔루션 MVP
 - GA4
 - 추천 모델 개발 중

HBEE

2019.06 ~ 2020.03

AI research engineer · AI팀

- 💡 반도체 공정 센서 스마트 인터랙 진/가성 분류 모델 개발(삼성전자 Pilot)
 - 역할 : 공정 센서의 인터랙 타입 자동화 분류 프로젝트의 AI 모델 개발 담당
 - 데이터 Type : 반도체 공정 데이터 (Image data + Time series sensor data)
 - Models : Semi supervised GAN, Auto-Encoder + Linear regression ensemble
 - Skills : Flask, Keras, Tensorflow, Tensorflow serving
 - 성과 : Phase 1 pilot project 성공적 종료 후 전 공정에 적용.(Precision 99%, f1-score 83%)

포항공과대학교 정보통신연구소

2018.11 ~ 2018.12

인턴 연구원 · 연구부

- 💡 Medical image(심장) Segmentation contest
 - 심장의 세 부위를 정확하게 segmentation하는 contest.
 - 역할 : 모델 구현 및 preprocessing 담당
 - 데이터 : 3D CT-images(.mha format)
 - Skills : Docker, Keras, Tensorflow
 - Model architecture : 3D-UNet
 - 성과 : Top 5(수상 X)

💡 인근 중소기업 기술 애로사항 해결을 위한 기술 지원

- 논문 서베이 및 세미나, baseline 구현

- subject : semi-supervised learning

프로젝트

Deep Knowledge Tracing

2021.05 ~ 2021.06

네이버 부스트캠프 AI Tech

🏆 리더보드 순위 : Public 2위 | Private 4위

💡 기간 : 4주

💡 대회 개요 :

DKT는 유저별 "지식 상태"를 추적하는 task입니다. 유저별 시험데이터를 통해 지식 상태를 추적하고 이를 통해 다음 문제의 정답 여부를 예측하는 task입니다.

💡 기술 개요 :

- Input : 7442명에 대한 230만 X 6 시계열 데이터와 target 레이블

- Output : "특정 유저"의 "특정 시점"에서의 target

- 모델 : LightGBM, SAINT 모델 양상을

- Skills : Feature engineering, EDA, Model implementation

🔗 <https://github.com/bcaitech1/p4-dkt-decayt>

Dialogue State Tracking

2021.04 ~ 2021.05

네이버 부스트캠프 AI Tech

🏆 리더보드 순위 : Public 1위 | Private 1위

💡 기간 : 4주

💡 대회 개요 : 유저와 시스템(챗봇)간의 대화에서 대화 상태(Dialogue State)를 매 턴마다 추론하는 Task

💡 기술 개요 :

- 데이터 : 9000개 대화(train : public : private = 7 : 1 : 1)

- input : 유저 발화 문장

- output : 유저 발화에서 얻을 수 있는 정보(Domain-Slot-Value)

- 모델 : SUMBT, TRADE, SOM-DST 모델 양상을

💡 역할 :

- 팀에서 SOM-DST 모델 구현 및 implementation

🔗 <https://github.com/bcaitech1/p3-dst-teamed-st>

Relation Extraction

2021.04 ~ 2021.04

네이버 부스트캠프 AI Tech

🏆 리더보드 순위 : 46/135등

💡 기간 : 2주

💡 대회 개요 :

관계 추출(Relation Extraction)은 문장의 단어(Entity)에 대한 속성과 관계를 예측/분류하는 문제

💡 기술 개요 :

- 데이터 : train set 9000문장, test set 1000문장

- input : 문장, 단어1, 단어2

- output : 단어1과 단어2의 41개 관계 Class

- 모델 : Roberta(hugging face pretrained model), kobart, base-bert 양상을

💡 전처리 :

- EDA(Easy Data Augmentation)

- Back Translation

- TEM(Typed Entity Marker)

🔗 <https://github.com/bcaitech1/p2-klue-ydy8989>

마스크 착용 상태 분류(18 Classes)

2021.03 ~ 2021.04

네이버 부스트캠프 AI Tech

🔥 리더보드 순위 : 152/223등

💡 기간 : 2주

💡 대회 개요 : 마스크를 끈 사람의 성별, 나이, 마스크 유무 총 18개 Class로 이미지 분류

💡 기술 개요 :

- 데이터 : 384 X 512 images

- 모델 : EfficientNet b0 ~ 4

- 전처리 : EDA를 통한 labeling 조정, Augmentation, Label filtering,

- 학습 전략 : data sampler, Stratified K-Fold

🔗 <https://github.com/bcaitech1/p1-img-ydy8989>

뉴스 document classification을 통한 익일 주가 상/하락 예측

2018.09 ~ 2018.10

포항공대 정보통신연구소

- 같은 뉴스도 기업의 주가에 각각 다르게 영향을 미칠 것이라는 점에서 프로젝트 진행. 특정 기업의 뉴스 기사를 통해 다음날 주가 상승 하락을 예측하는 프로젝트.

- 역할 : project manager 및 전처리(feature engineering), 모델 구현

- 데이터 : 뉴스 데이터 크롤링 5년치

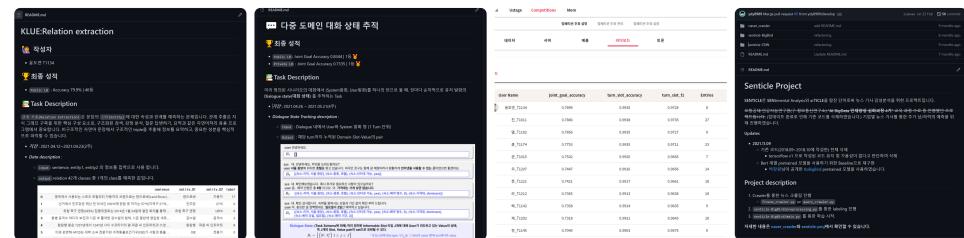
- Skills : Tensorflow, LIME, NLP

- Model architecture : 1DCNN

- Demo : <https://youtu.be/syQfQGFAAZ0>

🔗 <http://www.github.com/ydy8989/senticle-proj>

포트폴리오



파일

🔗 [Doyeon_Yoon_Portfolio.pdf](#)

링크

😺 <https://github.com/ydy8989/senticle-proj>

😺 <https://github.com/bcaitech1/p2-klue-ydy8989>

😺 <https://github.com/bcaitech1/p3-dst-teamed-st>

😺 <https://github.com/bcaitech1/p4-dkt-decayt>

전문 분야

스킬



산업 지식



학력

광운대학교

수학과 졸업

2015.03 ~ 2017.02

Combinatorics(A+), Graph Theory(A+)

세부전공 : Combinatorics

학위논문 : The anti-bandwidth of triangular grid graph

광운대학교

수학과 졸업

2013.03 ~ 2015.02

Linear Algebra(A+), Algebra, Graph Theory(A+) etc.

서울과학기술대학교

기계시스템디자인공학 중퇴

2008.03 ~ 2013.02

수료 및 자격증

부스트캠프 AI Tech

네이버커넥트재단

2021

AI·Big Data 인재양성 심화과정

포항공과대학교 정보통신연구소

2018

URL

<https://www.notion.so/whydo/Doyeon-Yoon-05603016086c4b3ca954cf2b6c64e46f>

<https://github.com/ydy8989>

<https://ydy8989.github.io/>