

윤도연

다 잘하고 싶은 끈기 만렙 4년차 Data Scientist 윤도연입니다.

Contact

Github: github.com/ydy8989

E-mail: ydy89899@gmail.com

Blog: (구)ydy8989.github.io, (신)yooniversai.github.io

About Me

포항공대 인공지능연구원 인턴을 시작으로 NLP, CV, RecSys 등 다양한 데이터 과학 분야에서 경험을 쌓았습니다.

에이치비(현. Scatter X)에서는 반도체 이상탐지 모델을 성공적으로 구현하여 삼성전자 전 공정에 적용하는 성과를 이뤘으며, 현재 재직중인 라이엔캐처스에서는 사내 도서추천 서비스 '비블리'의 고도화에 기여하였습니다.

사내 DS팀과 함께 지식을 공유하며 지속적인 스터디를 운영해 팀 내부의 NLP 역량을 강화했고, 문체 변환 서비스 프로젝트를 통해 사내 자산을 마련했습니다. 이 과정에서 프리A 투자 유치에 기여했으며, 현재는 네이버, 이스트소프트와의 LLM 구축 사업에 투입되어 RAG 시스템 구축을 진행 중입니다.

Data Science와 ML Engineering 분야에서 성장하며 다양한 프로젝트에 가치를 더하고자 합니다.

Skills

Frameworks & Libraries & SDK

- Tensorflow | Pytorch | Scikit-learn | Pandas | Numpy
- HuggingFace | Ollama | Langchain
- FastAPI | BentoML | Cog | Langserve
- matplotlib | seaborn | streamlit

Infrastructure & Databases

- MongoDB | Redis
- Milvus | ChromaDB | FAISS
- Docker

Tools & IDEs

- Jira | Height
- Github | Gitlab | Bitbucket
- Confluence | Notion
- Jupyterlab

Work Experience

라이엔캐처스

Data Scientist

2021.11 ~ Present
(2년 10개월)

네이버, 이스트소프트 글로벌 AI 교육 플랫폼

2024.07 ~ Present

- AI 튜터, AI 코디, AI 조교 등의 기능을 지원하는 LLM 기반 교육 솔루션 플랫폼 구축 사업
- Langchain을 사용한 HyperclovaX 모델 커스텀 및 이를 활용한 RAG 시스템 구축
- FastAPI를 이용한 벡터스토어 api와 Langserve를 활용한 RAG 서비스 서빙
- PoC를 기반으로 한림대학교 제안 수주 성공

휴노 - AI 채용서류 평가 자동화

2023.06 ~ 2024.04

- 취준생 서류 자동 평가를 위한 AI 모델링 업무 수행
- 총 14개 위반 및 검출 행위에 대하여 위반 사항 검출률 Precision 0.99, f1-score 0.98 수준으로 달성
- 특정 위반 사항(4종류)에 대하여 pretrained 언어모델을 사용한 text classification 진행
- 표절 검사시 속도 및 메모리 이슈 해결을 위해 클러스터 단위로 잘라 병렬처리로 문제 해결
- ML 관련 기술력과 문제 해결능력을 인정 받아 KEIT 및 NIA 주관 휴노의 다음 프로젝트에 투입
- 경영난으로 힘들었던 시기에 사내 재정 확보에 보탬이 됨

라이언캐처스 - 지식거래 플랫폼 유사 문서 검색 및 랭킹

2023.01 ~ 2023 03

- 자사 신규 서비스에 들어갈 "문서 클러스터링" 기능 구현에 필요한 모델링 수행
- 문서간 유사성을 계층적으로 계산하도록 구현하여, 문서의 길이에 상관없이 유사도를 정의할 수 있는 척도 마련
- 자사 서비스 내 문서 유사도 관련 기술에 대한 이해도 향상 및 VectorDB와의 차별성 고려한 산출물 확보

중소벤처기업부 - 대스타 해결사 플랫폼 데이터 증강 대회

2022.09 ~ 2022.12

- 공개되지 않은 의도 분류 언어 모델에 강건한 증강 데이터를 생성하는 대회
- 빠른 베이스라인 구축을 통한 실험 환경 구축으로 테스트 및 학습 소요 시간 절약
- 미리 개발해둔 문체 변환 모델을 적극 활용하여 빠른 시간 내에 많은 방식의 데이터 증강 진행
- 그 외 증강 처리 기법들을 앙상블하여 성능 향상(Back translation, TEM)
- 결선 진출 및 최종 준우승(9천만원) (<https://www.startupn.kr/news/articleView.html?idxno=31137>)

라이언캐처스 - 추천솔루션 AI:MD

2022.04 ~ 2022.12

- 자사 추천 프로젝트의 경험을 정리하고 표준화하여 솔루션으로 제품화 가능한 기반 제공
- 오픈소스 Recbole(<https://github.com/RUCAIBox/RecBole>)의 커스텀 및 기능 정리 진행
- 애자일 스크럼 마스터 역할을 수행하여 프로젝트 성과 및 산출물 확보

와인나라 - 추천시스템

2021.12 ~ 2022.04

- 와인 온라인 사이트에 배포할 아이템 기반 & 구매이력 기반 추천 시스템 구현
- cold start에 취약한 implicit feedback 데이터를 기반으로 추천 모델 구현
- 기존에는 사용하지 않던 text 메타 데이터에 pre-training한 언어모델 embedding을 적용해 아이템 추천 성능 향상

- 사내 추천시스템 기술 역량 성장 및 산출물 확보를 통한 이후 LINE Wallet Platform Feature Store 수주의 발판이 됨.

ScatterX

Data Scientist

2019.06 ~ 2020.03
(10개월)

반도체 공정 스마트 인터락 진/가성 분류 모델 개발

2019.09 ~ 2020.03

- 반도체 공정 센서의 Interlock 타입 자동화 분류(이상탐지) 프로젝트의 AI 모델 개발 담당
- 기존 휴먼 리소스에 의존하던 분류 방식에서 AI 자동화 적용 테스트
- Phase 1 pilot project 종료 후 성능을 인정받아 전 공정에 적용 성과 달성

포항공대 인공지능연구 원(PIAI)

Intern Data
Scientist

2018.11 ~ 2018.12
(2개월)

아산병원 "HeLP Challenge 2018 Contest" 참가

2018.11 ~ 2019.02

- Medical Image(Cardiac) Segmentation 경진대회 참가
- 3D segmentation baseline 구현(3d-Unet)을 통한 실험 환경 구축
- .mha format 파일의 전처리 진행
- 최종 Top 5

인근 중소기업 기술 애로사항 해결을 위한 연구부 기술 지원 업무

2018.11 ~ 2018.12

- 논문 survey 및 seminar 발표를 통한 연구부 활동 지원
- 리뷰 주제 : semi-supervised learning
- 관련 논문 : learning by association

Other Experience

Side Project

Chat PDF
Clone coding

2023.12 ~ 2024.01

- Langchain 숙련도를 올리기 위한 클론코딩
- FAISS를 활용하여 PDF 파일 인덱싱 후 간단한 RAG chain으로 구현
- Streamlit을 활용하여 배포

비-글(Bigle) 문
체 변환 서비스

2022.01 ~ 2022.10

- 글 또는 대화를 입력했을 때 성경(Bible) 말투로 변환해주는 서비스. 이후 급식체, 할아버지체, 등등 다양한 문체로 말투를 바꿔주는 모델을 학습하였음.
- KoBart 모델을 Fine tuning 후 BentoML을 이용하여 서빙 진행
- 전사 개발 조직을 위한 사이드 프로젝트로서 추후 "대스타 해결사 플랫폼" 대회 진행시 만들어 놓은 문체들을 활용하여 데이터 증강을 빠르게 진행하는데 도움이 됨

Training

Naver

Boostcamp AI Tech

2021.01 ~ 2021.06

- 약 6개월간 네이버 현업자분들과 함께하는 교육 이수 (NLP, CV, GNN 등)
- 3개월의 온라인 강의 기간 수료 + 3개월의 대회 형식으로 리더보드 순위를 가르는 프로젝트 진행
- 대회 주제 4가지(Relation Extraction | Dialogue State Tracking | Deep Knowledge Tracing | Image Classification)
- 그 중 Dialogue State Tracking에서 리더보드 Public 1위, Private 1위 달성, Deep Knowledge Tracing에서 리더보드 Public 2위, Private 4위 달성

포항공대 PIRL(현 PIAI)

AI/Big Data Advanced Course

2018.09 ~ 2018.10

- 2개월간의 교육 (Tensorflow, Pytorch, 인공지능 개론 등)과 함께 팀 프로젝트를 진행
- 뉴스 기사를 활용한 주가 상하락 예측 앱 구현 프로젝트에서 전통적인 방법의 모델링으로 예측 정확도 75% 달성
- Soynlp를 활용한 토큰나이징, Fasttext를 활용한 임베딩 그리고 1DCNN을 활용한 모델링과 더불어 라이브러리 LIME을 활용하여 설명가능한 상/하락 예측 앱 구현
- 앱 데모 : <https://www.youtube.com/watch?v=syQfQGFAAZ0>

Education

- 광운대학교 일반대학원 수학과

2015.03 ~ 2017.02

- 광운대학교 수학과

2013.03 ~ 2015.02

- 서울과학기술대학교 기계시스템디자인공학과

2008.03 ~ 2013.03