# 백주엽

이메일: wnduq08217@naver.com
연락처: +821050276059

안녕하세요. 3년차 AI 엔지니어인 백주엽입니다.

스타트업에서 AI 엔지니어로서 일하며 모델 개발과 MLOps 분야에서 경험을 쌓았습니다.
개발에 대한 지식과 수리적 역량이 강점이며 항상 개선점을 생각하는 습관을 가지고 있습니다.
제 강점과 경험으로 팀의 효율성 향상과 프로젝트 성과에 기여했습니다.

최근에는 LLM과 MLOps 기술에 큰 관심을 가지고 있어 꾸준히 공부하고 있습니다.
아래 링크를 통해 제 GitHub와 블로그를 확인하실 수 있습니다.

GitHub: https://github.com/100jy
블로그 (Notion): jy100space.notion.site

---

경력

**주식회사쓰리빌리언** 2021.07 – 현재 재직중
AI팀 / AI engineer

- **유전병 진단을 위한 MIL(Multiple Instance Learning)기반의 변이 추천 시스템 개발 및 배포**
  2023.03 – 2023.10

  관련 기술:

  Pytorch, Rank system, MLflow, FastAPI

  문제상황:

  - 유전병 환자들은 대체로 2만개의 유전 변이를 가지며, 이 중 하나라도 병인이 되는 변이가 있다면 유전병으로 판단합니다.
  - 전체 유전 변이 중에서 rule based 방법으로 필터링한 대략 1000개의 변이 중 1~2개가 실제 원인변이로 False positive가 자주 발생하는 문제가 있었습니다.
  - 라벨이 부여된 유전 변이는 전체의 0.1%에 불과했고, 유전병 여부뿐만 아니라 해당하는 병인 유전자를 예측해야 하는 어려운 과제였습니다.

  역할 및 기여:

  - 해당 프로젝트에서 저는 모델링 및 기술 조사, API 개발에 기여했습니다.

해결방법:

- 모델 설계: Multi-modal 구조와 다중 인스턴스 학습을 활용하여 모델을 설계하였습니다.
- False Positive 대응: Sparsemax, gating mechanism, Attention layer를 결합하여 False positive 문제를 효과적으로 해결했습니다.
- 성능 향상: RankNet loss를 적용하고 추가 feature를 도입하여 모델의 성능을 향상 시켰습니다.
- API 배포: FastAPI를 이용하여 내부용 API를 배포하였습니다.
- 트렉킹 및 실험 관리: MLflow를 활용하여 실험 과정을 효율적으로 추적하고 실험을 체계적으로 관리했습니다.

성과:

- Top 5 hit-rate 80%에서 98%까지 향상.
- AUROC 기준 81.에서 88.까지 향상
- 국내 특허 등록 및 해외 출원
- 기술 특례 상장 핵심기술로 활용

- **Milvus DB 구축 및 실험 관리 웹 어플리케이션 개발** 2023.01 - 2023.03

관련 기술:

- Milvus, Docker, AWS EC2
- Bootstrap, JavaScript, Flask, MongoDB

문제 상황:

- 새로운 화합물과 특허 등록된 구조 간의 유사도 검색 시스템이 필요했습니다.
- 2억 개의 화합물 구조를 담은 DB에서 매 실험마다 5만 개의 구조를 검색해야 했습니다.
- 화합물 구조는 1024비트의 bit array로 이루어져 있습니다.
- 더불어, 실험 및 평가 결과가 매번 문서 형식으로 보고되어 실험 관리가 어려웠습니다.

역할 및 기여:

- 해당 프로젝트에서 저는 기술 조사, API 및 웹 개발에 기여했습니다.

해결 방법:

- Milvus 오픈소스 벡터 DB를 Docker 컨테이너로 on-premise 서버에 구축하였습니다.
- Milvus DB를 활용하여 화합물의 신규성을 평가하는 서비스를 개발하였습니다.
- EC2 서버에 Docker 이미지 형태로 내부 서비스를 배포하였습니다.
- Flask 기반의 웹 어플리케이션을 개발하여 실험 및 평가 결과를 기록할 수 있도록 하였습니다.

성과:

- 기존의 mysql+파이썬 코드 대비 쿼리당 속도를 5초에서 0.019초 수준으로 개선했습니다.
- 실험 결과 계산 속도를 7시간에서 2분으로 대폭 단축시켰습니다.
- 실험 기록의 편의성을 크게 향상시켰습니다.

- **화합물 대사 안정성 및 독성 평가 모델 개발** 2022.09 - 2022.12

gnn기반의 화합물 독성 예측기와 xgboost 기반의 대사안정성 예측 모델을 개발하였습니다.

관련 기술
- pytorch, scikit-learn, xgboost, gnn, torch-geometric

역할 및 기여:

- 해당 프로젝트에서 저는 모델링에 기여했습니다.

- **지식 그래프 기반 약물 타겟 유전자 탐색 시스템 개발** 2022.01 - 2022.03

gene ontology, reactome DB를 기반으로 하여 약물의 타겟이 될 수 있는 유전자를 그래프 알고리즘을 이용하여 찾는 시스템을 개발하였습니다.

관련 기술:
- 네트워크 분석, networkx, owleady2, spaql

역할 및 기여:
- 해당 프로젝트에서 저는 DB파싱, 알고리즘 고안 및 개발에 기여했습니다.

- **강화학습 기반 약물 구조 생성 모델 개발** 2021.07 - 2021.12

"LEARNING TO NAVIGATE THE SYNTHETICALLY ACCESSIBLE CHEMICAL SPACE USING REINFORCEMENT LEARNING"에서 제시된 방법에 따라 Twin Delayed DDPG 및 Ray를 이용하여 약물구조 생성 모델을 재현하였습니다.

관련 기술
- 강화학습, pytorch, Ray, Twin Delayed DDPG

역할 및 기여:

- 해당 프로젝트에서 저는 논문조사 및 모델링에 기여했습니다.

관련자료:

https://github.com/100jy/TD3-from-scratch

---

## 악사손해보험 2021.04 - 2021.07

Data science 팀 / 데이터 분석가

- **Anomaly detection 기반 보험 사기 탐지 모델 개발** 2021.04 - 2021.07

  관련 기술
  - SQL, Pytorch, Deep-SAD, tabnet

  문제 상황:

  - 보험금 청구건 중 사기 의심 건을 예측하는 분류 모델이 필요했습니다.
  - 전체 청구 건 중 사기 건의 비율이 매우 낮았습니다.(천건당 1건 수준)

  해결 방법:

  - 준지도학습 방법을 고려하여 Deep SAD를 선택하여 조사를 진행하였습니다.
  - 공개된 소스코드를 기반으로 사내 데이터에 맞춰 모델을 구현하였습니다.
  - 사내 데이터를 집계하여 Dataframe을 구성하고 EDA 및 전처리를 하였습니다.
  - 인코더를 tabnet encoder로 교체하여 성능을 소폭 개선했습니다.

  성과:

  - 기존 모델 대비 소폭의 성능 개선 확인.

  관련자료:
  https://github.com/100jy/anomaly-research

---

## 가우디오랩 2020.10 - 2020.12

AI팀 / AI 연구 인턴

- **CNN 기반 음원 분리 모델 개발** 2020.10 - 2020.12

  음원에서 가수의 목소리와 반주를 분리하는 모델을 개발하기 위해 기존의 알고리즘을 조사하고 새로운 모델을 개발하였습니다.

  관련 기술
  - Tensorflow, Pytorch, Torchaudio, ONNX, Densenet

역할 및 기여:
- 해당 프로젝트에서 저는 논문조사 및 모델링에 기여했습니다.

관련자료:
https://github.com/100jy/source_separation

---

학력

### 부산대학교 2014.03 - 2020.08

통계학과(자연)
3.39 / 4.5

학부연구생 활동:
- LSTM 모델을 활용한 유동인구 예측을 통한 옥외광고 타겟팅 모델
- 관련 자료: https://github.com/100jy/lstm_com

관련 수업:
- 회귀분석/3.0/A+
- 시계열분석/3.0/A+
- 베이지안 통계학/3.0/A+
- 통계 프로그래밍 언어/3.0/A+

---

스킬

**Python, Pytorch, Git, SQL, Linux, Docker, Github, AWS, JavaScript, Scikit-Learn, Flask, FastAPI**

---

수상 및 기타

### 논문

- 제목: Novel Variant Prioritization leveraging Multiple Instance Learning for Jointly prioritizing SNV and CNV and classifying reportability in rare disease: 3ASC
- 현재 작성 중

### 특허

- 제목: 다중 인스턴스 학습에 기반한 유전성 질환 예측 및 질병 유발 유전변이 발굴 시스템
- 등록번호: 1025847700000(한국)
- 출원번호: 18/495.539(미국), 2023-180434(일본), EP23205209.2(영국, 스위스)

### 기타 이력

공모전) 음성 중첩 데이터 분류 AI 경진대회 (2020.07)

- 전체 85팀 중 13등

스터디) 학과 딥러닝 스터디 (2020.01)
이안 굿펠로우의 *Deep learning 서적을 이용하여 스터디 모임을 하였습니다.
*Goodfellow, Ian, Yoshua Bengio, and Aaron Courville. Deep learning. MIT press, 2016.

공모전) 데이터 청년 캠퍼스 우수 프로젝트(2019.08)
- 효율적인 미세먼지 관련 정책 시행을 위한 spark 빅데이터 시스템 구축
- 우수상(6등)

교육) 빅데이터 청년 인재 교육 과정(2019.07)
- 자바, 스프링 프레임워크, 오라클 디비, 스파크 교육을 수료하였습니다.

---

외국어 **영어**
비즈니스회화

---

링크 https://github.com/100jy

jy100space.notion.site