

박기범

[cow-coding](#) [in kbp0237](#) kbp0237@gmail.com [cow-coding.github.io](#)

WHO AM I

- 시스템을 지지하는 든든한 개발자

전체 시스템의 기반이 되는 인프라와 데이터를 구축하는 개발자를 지향합니다. 프로젝트, 협업 과정에서 주도적인 역할이 아닐지라도 제가 설계한 파이프라인, 데이터 모델을 통해 프로젝트가 원활히 수행될 때 가장 큰 성취감을 느낍니다.

- 더 나은 방안을 위한 끊임없는 노력

항상 더 나은 방안을 찾기 위해 고민합니다. 문제해결방향, 적용 기술 뿐만 아니라 협업하는 동료들과의 소통 방법에서 더 효과적인 방법을 찾고자 끊임없이 노력합니다.

- 파이프라인 개발 경험

Monolithic 데이터 파이프라인부터 Producer-Consumer 방식의 파이프라인까지 경험을 했습니다. 기존에 익숙한 방법인 monolithic 파이프라인은 데이터 수집과정에서 문제가 발생하면 데이터 누락이 발생했습니다. 이를 해결하고자 sender와 receiver를 활용한 mail box 방식을 적용하였습니다. 이처럼 데이터를 수집하는 과정에서 발생하는 비동기 문제, 예상치 못한 오류들을 대응하기 위해 다양한 방법들을 적용한 경험을 갖고 있습니다.

EDUCATION

인하대학교

2019.03 - 2023.02 (예정)

- 컴퓨터공학과

- 정보분석전공 (통계학 연계전공)

3.88 / 4.5 (current)

BoostCamp AI Tech 3rd, NAVER

2022.01 - 2022.06

- 추천 시스템 트랙

Machine Learning Engineering for Production (MLOps Specialization)

- Coursera : Certification

COMPETITIONS

CLOVA AI Rush 2021, NAVER

2021.05 - 2021.07

Round 1 - 13rd (회원 키보드, 마우스 기록을 통한 어뷰징 탐지)

Round 2 - 9th (재생 및 메타데이터를 활용한 음악 추천)

TEACHING

학부 조교 팀장, 인하대학교

2021.03 - 현재

- 컴퓨터공학과 학부 조교 총괄

- 강의자료 제작 및 강의 질의응답

GED6003 : 컴퓨팅사고와 데이터분석 기초

GED6004 : 블록코딩과 AI기초

AIE1002 : 객체지향 프로그래밍 2 (2021.09 - 12)

학부 조교, 인하대학교

2020.09 - 2020.12

- 강의 질의응답

GEG3073 : 창의적 사고와 SW 입문 교육

GEG3074 : 창의적 사고와 SW 심화 교육

INTERESTS

Recommender System | MLOps | DataEngineering

SKILLS

Language & Framework

PyTorch | Python | C++ | MySQL

Others

Recommender System | Linux | GCP

PROJECTS

| | |
|--|--------------------|
| 1. BoostCamp AI Tech 3rd | 2022.01 - 2022.06 |
| Final Project: Github you may also like | 2022.05.01 - 06.13 |
| 사용자의 github star, 방문 기록을 통한 repository 개인화 추천 크롬 익스텐션 | |
| - 사용 기술: Python MongoDB GCP Docker Recommender System Airflow | |
| - Github & Post: 데이터 엔지니어링 코드 개발 일지 | |
| - 주요 역할 | |
| • 데이터 ETL 파이프라인 개발 및 관리 | |
| - 추천 모델 학습을 위한 기본 데이터셋 수집 및 데이터 ETL 파이프라인 개발 (user 36만명, item 2700개) | |
| - Docker를 활용한 Message Queue 기반의 producer-consumer 파이프라인으로 데이터 수집과정의 비동기 처리 및 API 호출 제한 문제 해결 | |
| • 배치 학습 파이프라인 개발 | |
| - VAE 모델의 신규 유저 inference 한계 극복을 위한 배치 학습 파이프라인 설계 및 개발 | |
| - Cloud DB의 데이터를 통해 기본 전처리, 학습 전처리, 모델 학습, 모델 정보 저장의 주기적 스케줄링을 위해 Airflow pipeline 구축 | |
| • 데이터베이스 설계 및 관리 | |
| - 프로젝트 전체 데이터베이스 설계 (5개 collection 설계 및 관리) | |
| - MongoDB Atlas 기반의 데이터베이스 구축 | |
| • GCP Management | |
| - 메인 서버인 Google Cloud Platform instance 관리 | |
| - Cloud Storage와 MongoDB Atlas를 연동하여 모델 정보 업로드 및 다운로드를 위한 atomic 코드 개발 | |
| Movie Recommendation: Implicit feedback을 활용한 sequential 영화 추천 대회 | 2022.03.21 - 04.14 |
| - 사용 기술: Python PyTorch Recommender System | |
| - Github: 팀 repository 개인 작업 repository (모델 코드) | |
| - 주요 역할 | |
| • Data EDA | |
| - 대회 데이터 분석을 통한 sequential 유의미함 파악 및 모델 연구 방향성 제시 | |
| • 추천 모델 및 학습 파이프라인 개발 | |
| - 다양한 종류의 추천 모델 학습 파이프라인 코드 개발 | |
| - 주피터 노트북 코드를 재구성하여 python code로 변환하여 하나의 파이프라인으로 재조합 | |
| • Code Refactoring | |
| - inference 코드 refactoring을 통한 inference 시간 단축 (3분 → 1분 미만) | |
| - 학습 파이프라인 개발시 라이브러리 최신 버전에 맞는 코드 refactoring 담당 | |
| Image Classification: 마스크 이미지 분류 대회 | 2022.02.23 - 03.03 |
| - 사용 기술: Python PyTorch | |
| - Github: 팀 repository 개인 작업 repository | |
| - 주요 역할 | |
| • Data EDA | |
| - Data EDA를 통해 기본 통계치 분석 및 Class 불균형 분석, 공유 | |
| - 이를 활용한 '연령' 클래스 분포 변경을 통한 모델 성능향상 | |
| • 베이스 라인 코드 및 학습 파이프라인 개발 | |
| - 팀 베이스 라인 코드 개발 | |
| - 베이스 라인 코드 수정을 통한 실험자의 자유로운 하이퍼 파라미터 변경 코드 개발 (Learning Rate Scheduler, Augmentation 방식 등) | |

2. 데이터베이스 프로젝트 : 알.쓸.선.잡

2021.03 - 2021.06

공공데이터를 활용한 선거 및 국회의원 정보 제공 웹 서비스

- 사용 기술: Python | Node.js | MySQL | GCP
- Github: <https://github.com/cow-coding/V.O.T.E>
- 주요 역할

- Team Lead

- 데이터 ETL 파이프라인 개발

- 공공데이터 포털에서 제공하는 API를 통한 데이터 ETL 파이프라인 설계 (Monolithic)
- 선거관련 API, 국회의원 API 데이터를 활용한 데이터베이스 공동 설계
- GCP에 MySQL 기반 데이터베이스 구축

- 데이터 통신 API 개발

- Node.js를 활용한 데이터 통신 API 개발

3. 여기불어 (ComeHere)

2021.01 - 2021.01

대학생 공동구매 어플리케이션

- 사용 기술: Java | Firebase | Android
- Github: <https://github.com/cow-coding/ComeHere>

- 주요 역할

- Team Lead

- 로그인, 회원가입 기능 개발

- 어플리케이션 로그인, 회원가입 기능 개발

- Firebase DBA 및 전체 어플리케이션 데이터베이스 연결

- Firebase 초기화 및 프로젝트 연동 설정
- 전체 어플리케이션 기능별 firebase 연동 담당
- Firebase 데이터 관리

4. Medipush

2020.08 - 2020.10

의약품 복용 정보 제공 통합 서비스 (환자/약사)

- 사용 기술: Python | MongoDB | AWS
- Github: https://github.com/cow-coding/Medipush_Crawling_Database | <https://github.com/Medipush>

- 주요 역할

- Team Lead

- 데이터 ETL 파이프라인 개발 및 관리

- 의약품 코드 수집을 통한 의약품 데이터 ETL 파이프라인 개발 (Monolithic)

- AWS에 MongoDB 구축 및 데이터베이스 관리