



김형준 Hyeonjun Kim

학력 연세대학교 (2013. 03 - 2020. 08 졸업 예정)
경영학과 / 컴퓨터 과학과 복수 전공 (3.74 / 4.3)

이메일 hjkim2246@gmail.com

블로그 yeomko.tistory.com

깃 협 github.com/yeomko22

관심 분야 Data Engineering, MLOps, Recommendation

**자기 소개 및
지원 동기** 데이터 엔지니어를 꿈꾸는 학부생 개발자입니다. 머신러닝 서비스를 구현하기 위한 다양한 엔지니어링에 관심이 많으며, 특히 데이터 엔지니어링, MLOps 분야에서 전문성을 키우고 싶습니다. 풀스택 개발자를 목표로 공부를 해왔으며 프론트엔드, 백엔드, 데이터 베이스 등 전반적인 개발 분야에 모두 익숙합니다. 대규모 데이터를 직접 다루는 경험을 쌓고 성장하길 바랍니다.

경력

Nota AI, 개발 인턴 (2019. 09 - 2020. 02)

- 카드 거래 데이터 분석 및 통계 지표 추출 시스템 설계 및 개발

Naver Clova AI, 개발 인턴 (2018. 11 - 2019. 02)

- 인공지능 광고 이미지 생성 시스템 백엔드 개발

Yonsei Database Lab, 학부생 인턴 (2019. 03 - 2019. 06)

- 빅데이터 개인정보 비식별화 프로젝트 백엔드 개발

비영리 민간 의료 봉사 단체 Freemed, IT 기획팀장 (2019. 03 - 2020. 03)

- 전자 문진 기록(EMR), 리크루팅 페이지, 홈페이지 유지 보수

**주요 활동 및
수상 실적**

- 2019 연세대학교 Software Capstone Project 우수상 (컴퓨터 과학과 1위)
- 2018 인공지능 R&D 챌린지 본선 진출 (12위)
- 2018 한국 정보처리학회 최우수논문상 (3 저자)
- 2017 SK 스마트 관광앱 공모전 우수상 (3위)

보유 기술

Programming Language

- Python, Java, Javascript, Go, PHP

Frontend & Backend Framework

- ReactJS, NodeJS, Flask, Django, Celery, Faiss

Deeplearning Framework

- Pytorch, Tensorflow

Infra

- AWS, GCP, Docker

**자격증 및
어학**

- Tensorflow Developer Certificate (2020. 04)
- IBT TOEFL 95 (2020. 02.)

프로젝트 1

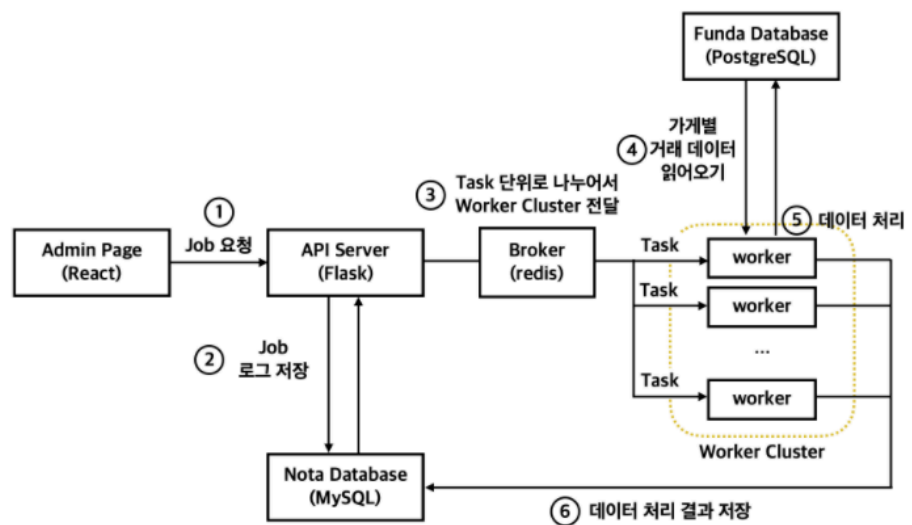
(2019. 9 ~ 2019. 12)

(주) Nota AI
소프트웨어 엔지니어

· 카드 거래 빅데이터 분산 처리 파이프라인

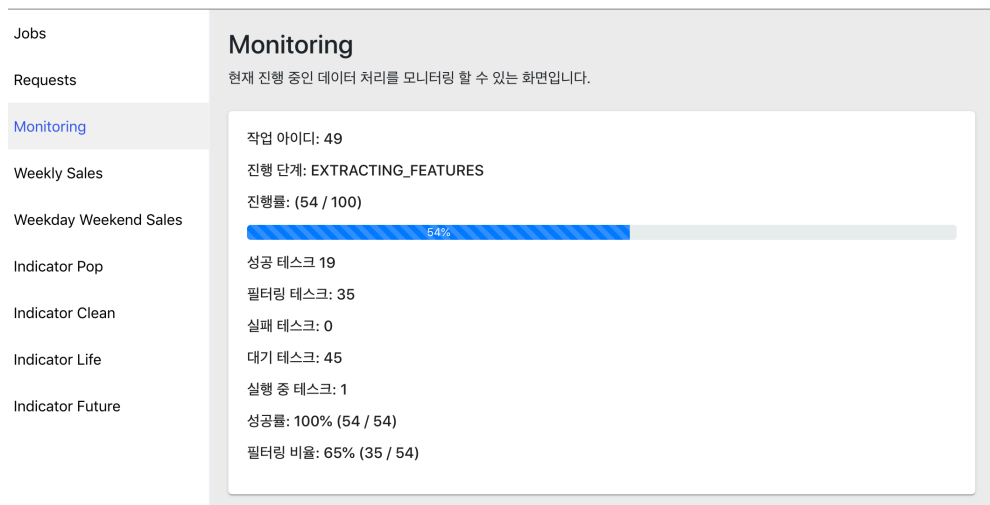
Nota AI 개발 인턴십 과정에서 진행한 정부 과제 프로젝트입니다. 매달 발생하는 1억건의 카드 거래 데이터를 분석하여 가게별 고객 생애 가치 지수, 미래 가치 지수 등의 지표를 추출하는 시스템입니다.

- ✓ 전체 서비스 구조 설계 및 구현
- ✓ Flask 기반 API 서버 구축
- ✓ Celery, Pandas, Redis 기반 분산 데이터 처리 워커 구현
- ✓ MySQL 기반 통계 데이터 저장 데이터 베이스 구축
- ✓ ReactJS 기반 관리자 페이지 구축
- ✓ Github Actions, Docker, GCP 기반 DevOps 파이프라인 구축



[도표 1] 데이터 처리 파이프라인 구조

nota Leading Startup - 빅데이터 플랫폼 관리자 페이지



[도표 2] 관리자 페이지 - 모니터링 탭

프로젝트 2

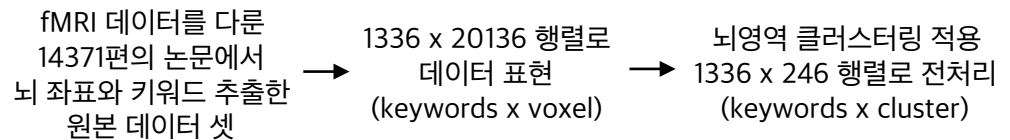
(2019. 3 ~ 2019. 7)
연세대학교 컴퓨터과학과
팀장

• Neurocast

연세대학교 컴퓨터 과학과 졸업 프로젝트 과목에서 진행하였으며 팀장으로 핵심 알고리즘 구현 및 실험 진행을 담당하였습니다. 특정 키워드와 뇌의 활성화 영역 좌표가 담겨진 데이터로 추천 알고리즘을 학습시킨 뒤, 아직 밝혀지지 않은 뇌 영역 좌표와 키워드 간의 상관관계를 추론하는 프로젝트입니다.

- ✓ [Demo]: <http://15.164.48.188:8000/>
- ✓ Collaborative Filtering 알고리즘 구현
- ✓ 실험 설계 및 모델 검증
- ✓ 데이터 전처리 및 추론 결과 시각화 구현
- ✓ 우수상 수상 (컴퓨터 과학과 1위)

[데이터 전처리]

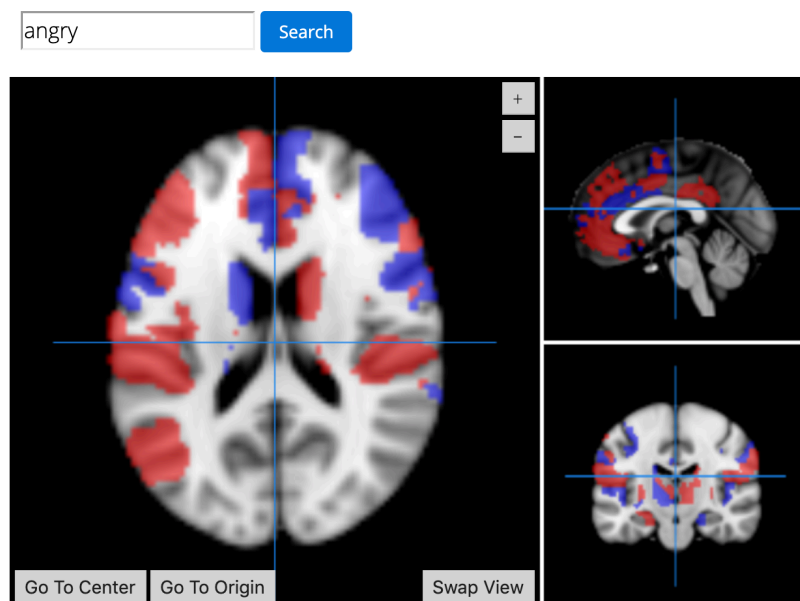


[모델 학습]

전처리를 통해 얻은 1336 x 246 행렬을 입력받아 키워드와 뇌 영역에 해당하는 Latent Matrix를 학습시키는 Matrix Factorization 기법을 적용하였습니다. 구현에 필요한 세부사항은 다음 논문¹을 참고하였습니다.

[결과 시각화]

Dashboard Demo



¹ Hu et al, "Collaborative Filtering for Implicit Feedback Datasets", 2008

• 2018 인공지능 R&D 챌린지

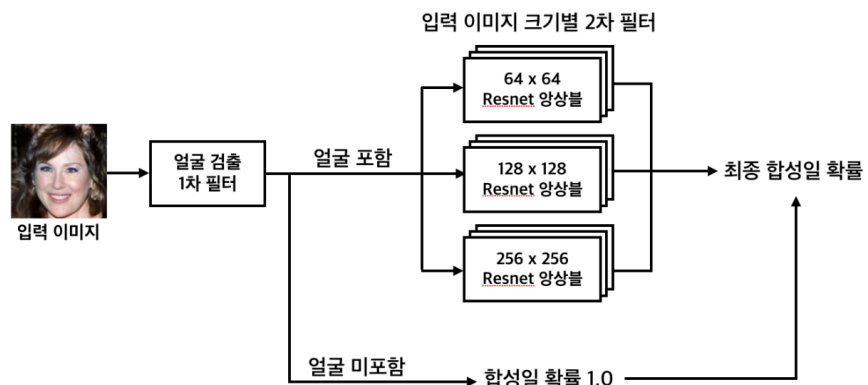
한국 과학기술정보통신부 주관의 2018 인공지능 R&D 챌린지에 팀장으로 참여하였습니다. 주어진 두 가지 과제에 대하여 각각 합성 이미지 탐지 모델을 설계 및 개발하였습니다.

- ✓ 임무 1: GAN을 통해 생성된 얼굴 이미지 탐지
 - ✓ 임무 2: 얼굴 전체 혹은 일부분이 합성된 얼굴 이미지 탐지
 - ✓ 기업과 대학 연구소를 포함한 전체 47개 팀 가운데 **최종 12위 기록**
- 임무 1 AUROC: 0.8953
임무 2 AUROC: 0.5489

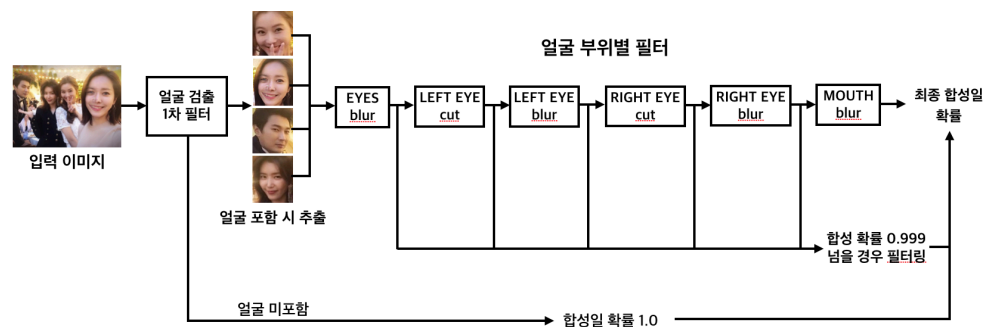
[데이터 셋]

- ✓ 임무 1: progressive GAN을 통해서 생성 얼굴 이미지 데이터 셋을 구성
- ✓ 임무 2: openCV, dlib 등을 활용하여 합성 얼굴 이미지 데이터 셋 구성

[임무 1 모델]



[임무 2 모델]



프로젝트 4

(2017. 11 ~ 2018. 6)
개인 프로젝트

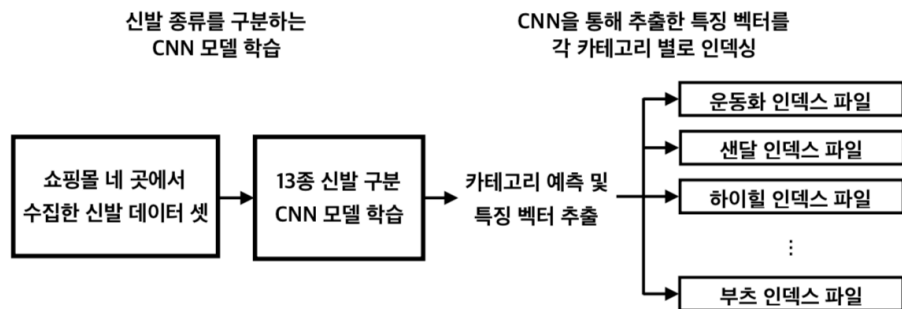
• 이미지 검색 신발 쇼핑몰, 슈피

이미지만으로 유사한 신발을 검색할 수 있는 쇼핑몰입니다. CNN을 활용하여 실생활에서 유용하게 사용할 수 있는 서비스를 개발해 볼 수 없을까라는 개인적인 호기심에서 시작하게 된 프로젝트입니다.

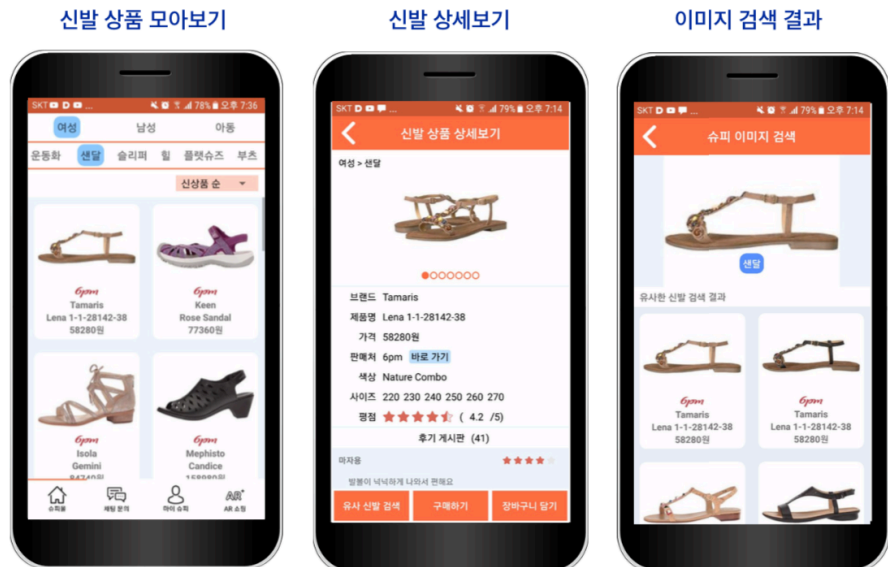
- ✓ [Demo]: <https://www.youtube.com/watch?v=M8jPjzrUXY8>
- ✓ 크롤링을 통해 4곳의 신발 쇼핑몰에서 신발 이미지 데이터 수집
- ✓ pyTorch와 faiss를 이용해 이미지 벡터 추출 및 인덱싱 구현
- ✓ Flask 기반 이미지 검색 API 서버 구축
- ✓ 안드로이드 기반 프론트엔드 앱 구현

[모델 학습 및 인덱싱]

- ✓ 신발 이미지를 13가지 종류로 분류하는 resnet50 모델 학습
- ✓ 각 카테고리 별로 모델에서 추출된 특징 벡터를 Vector Quantization 적용
- ✓ 이미지 검색 쿼리가 들어오면 특징 벡터 추출 뒤, L2 거리가 가까운 이미지를 찾은 뒤, 결과 반환



[결과 시각화]



프로젝트 5

(2017. 7 ~ 2017. 12)
과학기술정보통신부
팀원

• 2017 인공지능 R&D 챌린지

한국 과학기술정보통신부 주관의 2017 인공지능 R&D 챌린지에 팀원으로 참여하였습니다. 주어진 두 가지 과제에 대하여 각각 가짜 탐지 모델을 설계 및 개발하는데 참여하였습니다.

- ✓ 임무 1: 제목과 상관 없는 뉴스 탐지
- ✓ 임무 2: 맥락과 상관 없는 문장이 삽입된 뉴스 탐지
- ✓ 뉴스 데이터 수집 및 전처리 담당
- ✓ 학습용 가짜 뉴스 데이터 생성 알고리즘 개발
- ✓ 기업과 대학 연구소를 포함한 전체 24개 팀 가운데 **최종 21위 기록**

[프로젝트 논문화]

해당 프로젝트의 논문화 작업에 공저자로 참여했습니다. 해당 논문은 2018 한국 정보처리학회 춘계 학술대회에서 **최우수 논문상을 수상**하였습니다. 또한 해외 JIPS 저널에 추천을 받고 게재 되었습니다.

- ✓ JIPS 저널 제출 논문 링크: <http://www.jips-k.org/q.jips?cp=pp&pn=707>
- ✓ 한국 정보처리학회 논문 제출 링크: <http://www.koreascience.or.kr/article/CFKO201826259815374.pub>

[가짜 뉴스 탐지 모델]

기사의 제목과 본문을 각각 Fasttext를 통해 임베딩을 추출합니다. 여기에 wide and shallow CNN을 적용하여 특징 벡터를 추출하여 해당 뉴스가 가짜일 확률을 구합니다.

