

# 구건모

TEL : (+82) 10-5022-9423 | EMAIL : ggm1207@gmail.com | GITHUB : github.com/ggm1207

자연어 처리, 범용 인공지능에 관심이 있으며 잘 설계된 인공지능은 사회를 더 이롭게 변화시킬 수 있다고 생각합니다. 인공지능의 발전으로 인해 발생하는 윤리적인 문제에도 관심이 많으며 어떻게 해결해야 할지 종종 생각하는 편입니다. 잘 동작하는 코드보다 잘 짜인 코드를 선호합니다.

## SKILL

Python, PyTorch, C++, JavaScript, Linux, Vim,  
MySQL, RocksDB, MongoDB, React

## PRIZE

[한국소프트웨어협회] 2018 SW Challenge - 입상(4등)  
[강원대학교] Capstone Design Festival - 은상(4등)

## EDUCATION

### Boostcamp AI Tech, Naver Connect & Upstage

2021년 1월 - 현재

- AI 지식, 수학, 통계, 그래프 이론, 자연어 처리, 컴퓨터 비전 등 다양한 머신러닝/딥러닝 지식을 이론과 실습으로 배우는 교육과정 프로그램
- 테크 토크 발표자: 모델링 자동화와 실험 검증 전략 [\[Link\]](#)

### B.S in Computer Science, Kangwon National University

2014년 3월 - 2021년 2월

- 전공 학점: 4.0/4.5 (5번의 성적 우수 장학금)
- 컴퓨터 그래픽스 학사 연구원: 1년간 카메라로 캡처된 Real Motion Data에 내재한 잡음을 딥러닝 모델을 사용하여 제거하는 연구 진행 [\[DEMO\]](#)

## WORK EXPERIENCE

### Kakao — Intern

2019년 12월 - 2020년 2월

#### 멜론 유사곡 추천 시스템 고도화 (Team of 3)

- (전처리) FastText로 노래의 메타 정보를 고려하는 Song2Vec 과제 수행, 유사곡 추천 시 잡음에 강한 추천 결과 생성
- (후처리) 장르 유사도를 사용하여 추천 결과를 재정렬하는 작업을 수행, 늘어난 연산량을 비트연산, 분산처리, RocksDB를 사용하여 해결
- 2주간의 배포 결과 기존모델 (Word2Vec) 대비 클릭 수 증가

#### 정량 평가 시스템 구축 (Team of 3)

- 정성 평가 및 A/B 테스트의 위험성을 해결하기 위해 서비스 배포 전에 추천 시스템의 정량적 지표를 확인할 수 있는 시스템을 구축
- 추천의 의외성을 측정하는 지표인 Serendipity, Novelty를 도입, Plotly를 사용하여 Interactive한 시각화 생성 후 Html에 이식

### ScalaWox — Intern

2019년 9월 - 2019년 12월

#### 서울 교통 공사, 모터 이상 상태 판별 시스템 구축

- 분산처리를 사용하여 대규모 모터 신호 데이터를 Peak 기준으로 분할 및 오류 데이터를 제거하여 학습 가능한 데이터로 정제하는 과정을 수행
- 모터 신호 코퍼스 데이터로 사전 학습된 CNN 분류 모델을 전이 학습하여 95%의 정확도로 모터의 이상 상태를 판별하는 시스템을 구축

## PROJECT

### Deep Knowledge Tracing (Team of 6) [\[Link\]](#)

2021년 5월 - 2021년 6월

- 사용자의 문제 풀이 데이터가 주어지면 마지막 문제의 정답 유무를 예측하는 모델 개발
- 다양하고 체계적인 실험을 위한 유연한 파이프라인 구축
- Ray[Tune]을 사용한 PBT 하이퍼파라미터 최적화 기능 구현

### Open Domain Question Answering (Team of 5) [\[Link\]](#)

2021년 4월 - 2021년 5월

- 질문이 주어지면 위키피디아와 같은 대규모 비정형 문서에서 문서를 검색하고 답변을 예측하는 모델을 개발
- Sparse Retriever와 Dense Retriever를 조합한 Hybrid Retriever 모델을 구현하여 문서 검색에서 TOP-3 정확도 기준 90%를 달성
- 협업 가능한 베이스라인 코드의 기반 설계, 4주간의 기간 동안 모델 추상화 및 리팩토링을 통해 기술 부채를 줄여 발생 가능한 재작업 비용을 줄임

### Relation Extraction [\[Link\]](#)

2021년 3월 - 2021년 4월

- 문장 내 두 개의 Entity 간의 관계를 예측하는 모델을 개발
- Ray[Tune]을 사용한 하이퍼파라미터 최적화 기능 구현 및 Sqlite를 사용하여 모델 성능에 기반을 둔 전략 샘플링 기능을 구현, 최종적으로 샘플링 & 학습 & 추론 & 제출이 반복되는 자동화 파이프라인 구축

### Mask Wearing Status Classification [\[Link\]](#)

2021년 3월 - 2021년 4월

- 마스크 착용 여부, 나이, 성별을 예측하는 모델을 개발
- 모델의 추론 근거를 시각화할 수 있는 GradCAMPP, 혼동 행렬 및 Confidence 등 다양한 시각화를 구현
- WandB Sweep을 사용하여 하이퍼파라미터 최적화 기능 구현 및 시각화 자료가 WandB에 자동 로깅되도록 파이프라인을 구축