

PORTFOLIO



이다은, Daeun Lee

✉ goodgpt@korea.ac.kr

Tel 010 5093 1892

<https://daeun-computer-uneasy.tistory.com/>

<https://github.com/daeunni>

이다은, Daeun Lee

학력

고려대학교(안암) 통계학과 재학

2019.03 ~
GPA (3.87 / 4.5)

딥러닝을 위한 통계적 모델링 | A+
통계적 머신러닝 | A+
학술논문입문 | A+
통계계산 소프트웨어 | A+

경력

미래에셋생명 | 디지털 혁신팀

2021.01 ~ 2021.02
인턴십 (인사이트 도출 및 분석)

Nyx | R&D

2021.06 ~ 2021.11
인턴십 (연구 보조 및 연구 수행)

연구

(CVPR) Trajectory prediction

2021.08 ~ 2021.11
제 1저자 (Main 실험 진행)

(CVPR) Colorization using DG

2021.07 ~ 2021.08
DG model + Colorization 실험 support

2022 CVPR Reviewer

대외활동

빅데이터 연합동아리 BOAZ

2020.01 ~ 2021.01
데이터 분석팀

Vision & AI LAB

2021.06 ~
학부연구생 (지도교수 : 김진규)

수상내역

2020 금융 빅데이터 페스티벌 | 최종1위

2020.08 ~ 2020.12
보험금 청구분야

2020 Kakao arena | 상위 2%

2020.05 ~
플레이리스트 추천시스템

산업혁신 빅데이터 분석 공모전 | 우수상

2021.05 ~ 2021.06
NLP를 통한 업종 코드 분류

Tools

Python

PyTorch

Sklearn

SQL

R

SAS

Keras

Blog

다은이의 컴퓨터 공부

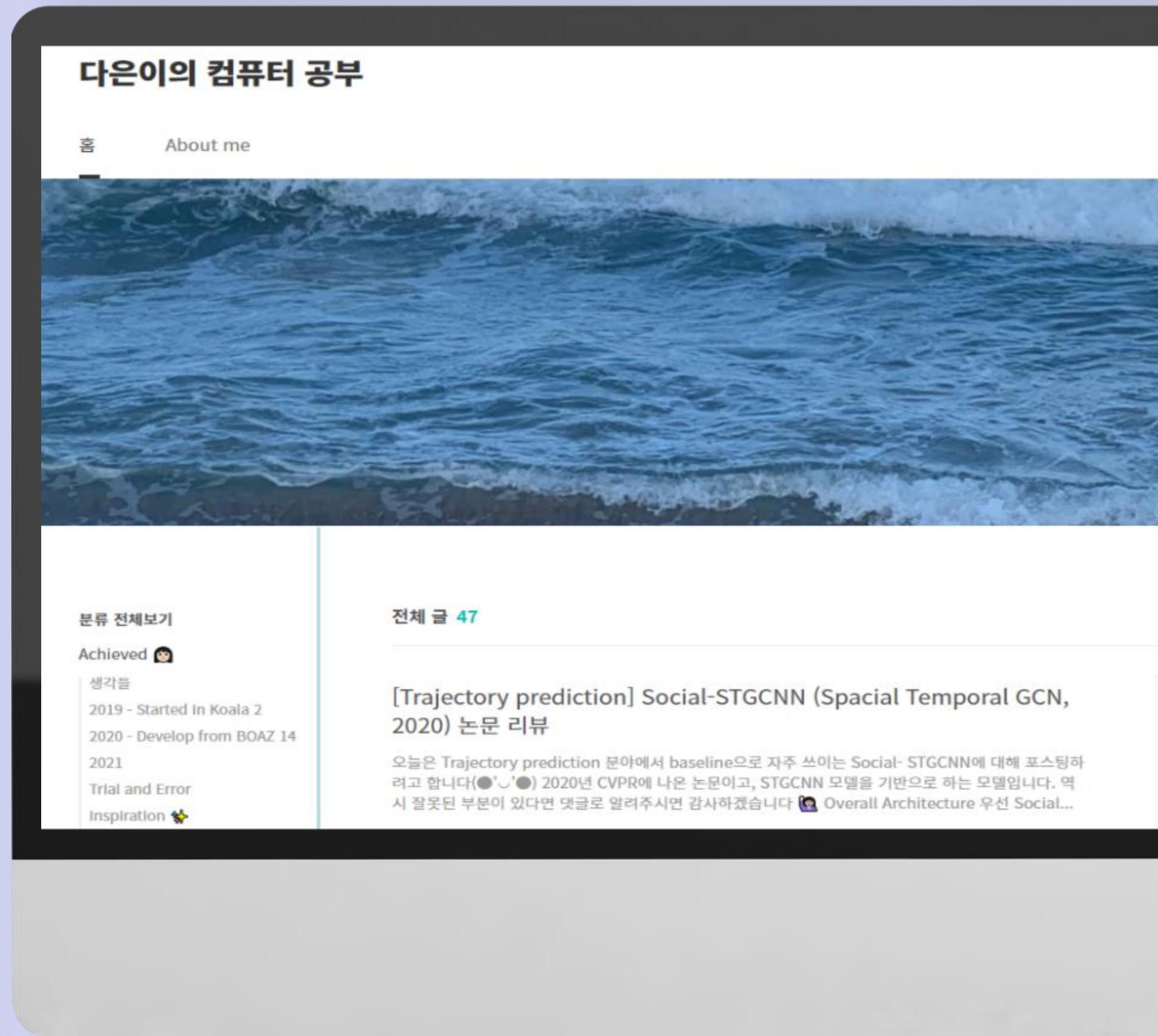
<https://daeun-computer-uneasy.tistory.com/>



- 하루 방문자 수 : 약 80 ~ 100명
- 누적 방문자 수 : 6700 (21.12.03 현재)

< 인기글 순위 >

- [CV] Self-supervised learning(자기주도학습)과 Contrastive learning
- [XAI] OpenAI CLIP 논문 리뷰[2] - Zero shot & Representation learning
- [XAI] OpenAI CLIP 논문 리뷰[1] - 전반적인 아키텍처
- [XAI] OpenAI CLIP 논문 리뷰[3] - Domain Generalization



CONTEST

- 참여했던 많은 공모전 중 인상 깊은 공모전만 수록하였습니다.

1. Kakao Arena - Melon playlist continuation (TOP 2%)

기여도 80% ●●●●●

✓ <https://github.com/daeunni/kakao-arena>
✓ 2020.04.30 ~ 2020.07.26

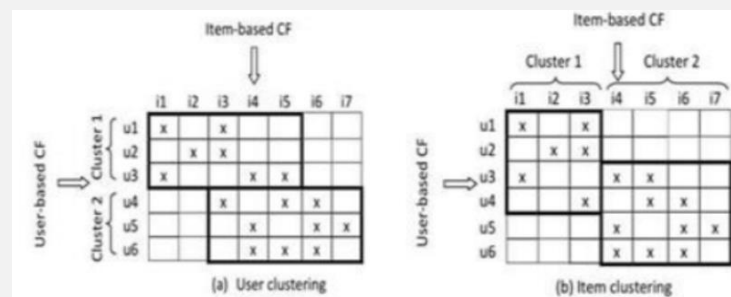
Recommender system

Problem

플레이리스트에 숨겨진 노래와 태그를 추천시스템을 활용해 예측

Main idea

KNN + CF (Collaborative Filtering)



KNN + CF Algorithms

CF 진행 전 KNN으로 Clustering을 진행하고, 이 후 CF를 적용하여 각 Active user와 유사한 song id를 출력하도록 설계

Side information 삽입

```
vali = val_song * val_tag * val_title_genre * val_genre  
cand_song = tra_song_sp_T.dot(vali) # 기존 행렬에 dot 연산으로 곱함
```

플레이리스트의 제목, 발매 일자, 가수 명, 앨범이름, 장르 등의 song side information을 자유롭게 CF에 첨가

Inverse Coefficient 사용

$$s_{uv} = \sum_{i \in I} ((f_i - 1)^\rho + 1)^{-1} \frac{r_{ui} r_{vi}}{\|R_u\|_2 \|R_v\|_2}$$

자주 등장하지 않는 Minor 한 곡들의 Similarity에 가중치를 주기 위함

Reranking & Denoising

- XGBoost를 Reranking에 사용
- Reranking과 Denoising의 중요성을 낮게 깨달아 아쉬움이 남음

2. 금융 빅데이터 페스티벌 - 보험금 청구 가능성 예측 및 BI (1위)

기여도 90% ●●●●●

✓ <https://github.com/daeunni/Insurance-contest>

✓ 2020.08.01 ~ 2020.11.27

ML

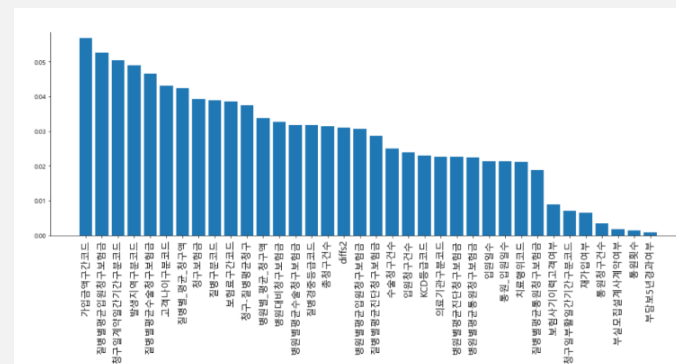
Business Intelligence

Problem

보험금 청구건 다중분류모델 competition + 비즈니스 활용방안 고안

Main idea

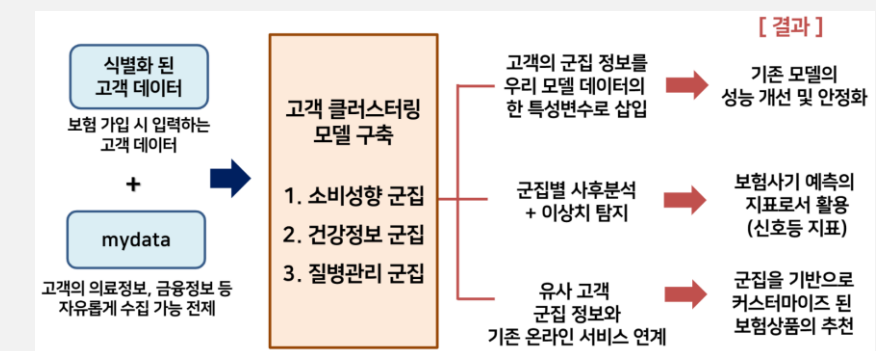
Extra tree + KNN model



Feature importance

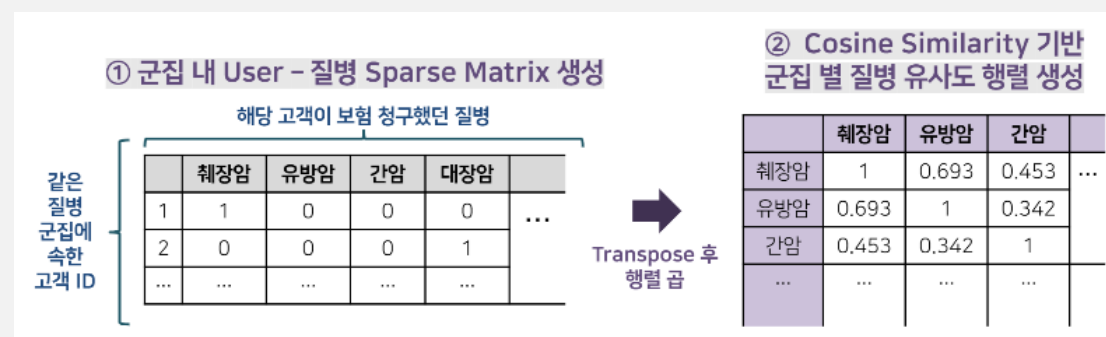
Randomness가 심한 Extra Tree의 특성을 보완하고 안정성을 높이하고자 Random state를 다르게 설정한 3가지 모델 + KNN 앙상블

고객 Segmentation process 제안



마이데이터 사업과 연계하여 고객 Segmentation 전략과 활용 방안에 대하여 피력

CF를 활용한 고객 유사도 기반 향후 예측 서비스 제안



군집 기반 협업필터링(Collaborative Filtering)을 활용하여 군집 내 고객의 질병 유사도 기반 질병예측 모델 제안

3. 디지털 산업혁신 빅데이터 공모전 - 업종코드 예측 (2위)

기여도 85% ●●●●●

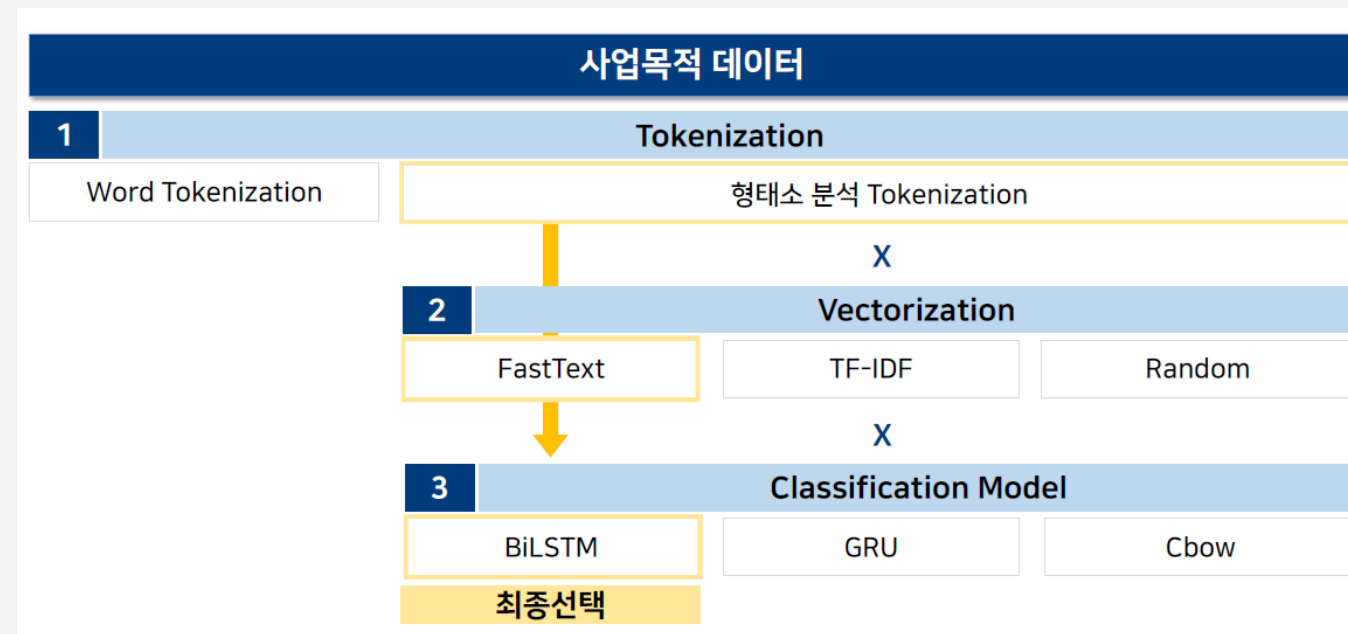
✓ https://github.com/daeunni/KED_Project
✓ 2021.04.09 ~ 2021.06.12

NLP

Problem 사업목적을 활용한 업종코드 예측

Main idea

Model flow chart



- Tokenization : Mecab (형태소 단위)
- Vectorization : FastText
- Classification Model : BiLSTM

1000여개의 target label 분류

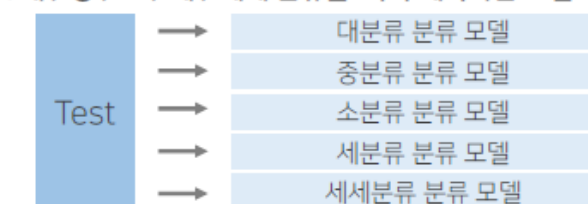


학습방법

1. 예측 모델로 분류된 대분류를 기준으로 중 / 소 / 세 / 세세 분류를 각각 예측하는 모델 학습



2. 대 / 중 / 소 / 세 / 세세 분류를 각각 예측하는 모델 학습



약 1000여개의 multi classification 을 주안점으로 고려하여 Model train method를 구상

PROJECT

- 많은 프로젝트 중 인상 깊은 프로젝트만 수록하였습니다.
- 현재 아직 Submitted 상태인 paper(CVPR 2022)의 실험은 넣지 않았습니다.

1. Pose estimation을 활용한 AI home trainer

기여도 75% ●●●●●

✓ youtu.be/gqZZXzuefs4
✓ 2020.08 ~ 2021.01

CV

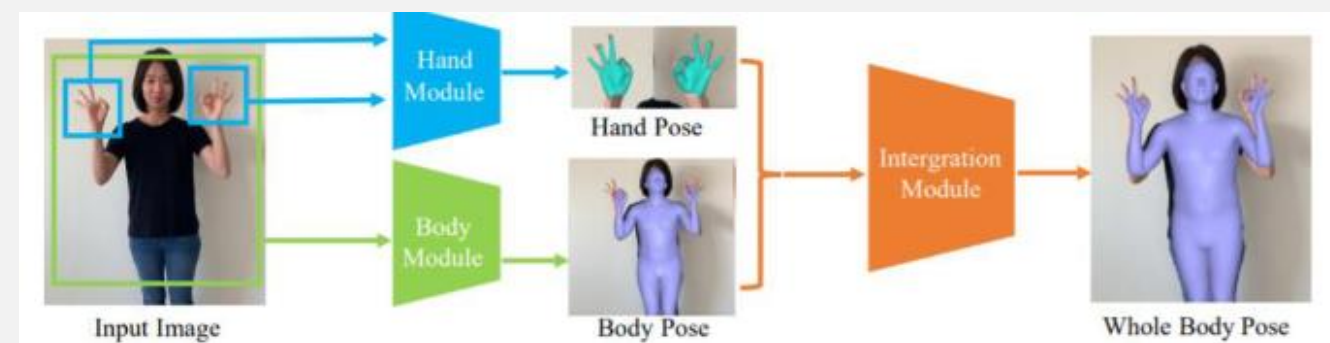
Pose Estimation

Problem

3D pose estimation 을 활용해 특정 운동 자세에서 개선이 필요한 부위를 실시간 음성으로 도출해 코칭

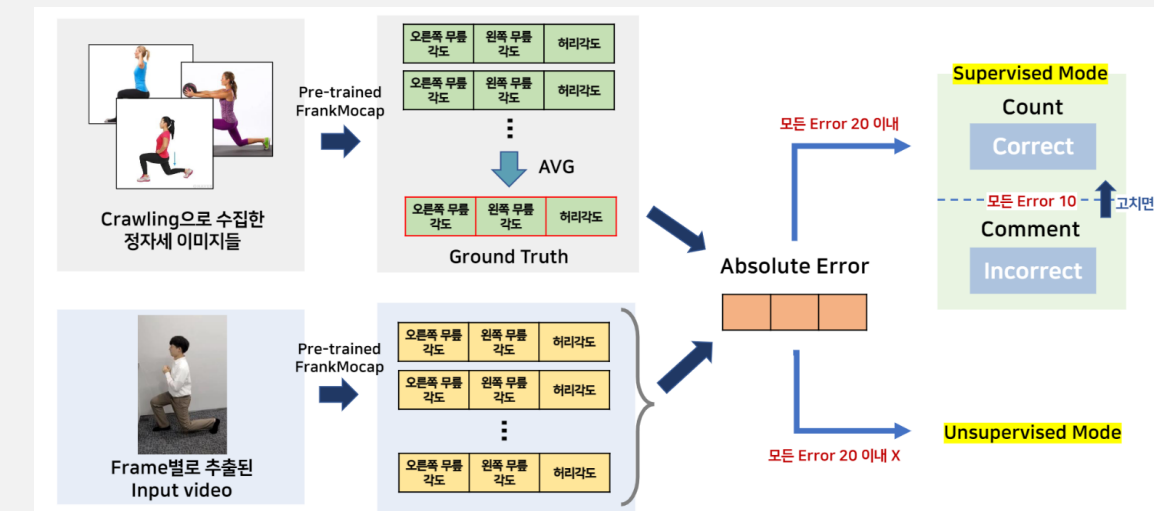
Main idea

3D pose estimation - Frank Mocap



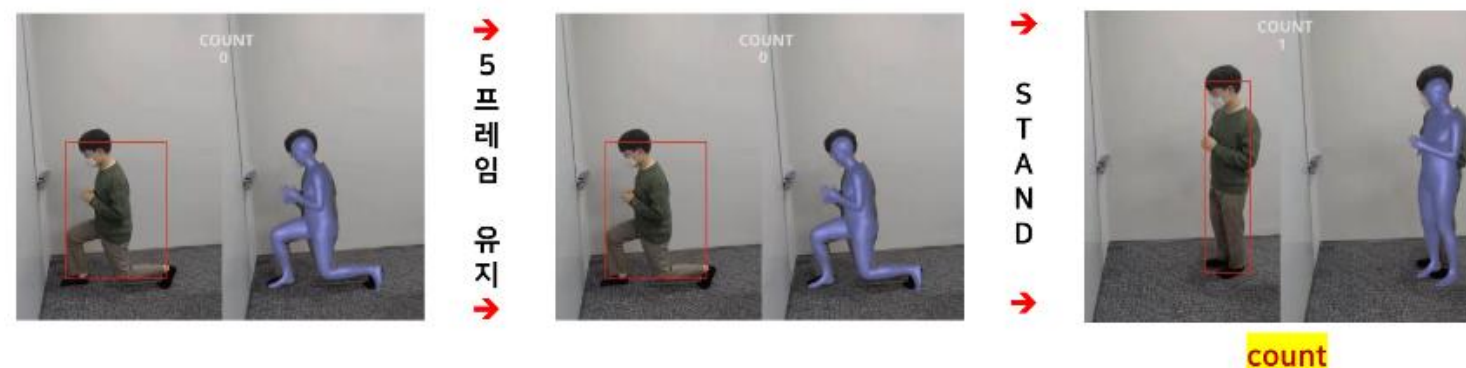
- Realtime 에 가까운 빠른 속도로 실시간 자세와 개선점 출력에 적합
- 3D mesh를 통해 실제 인간의 움직임을 고려하고 손동작까지 capture

Architecture



운동 자세의 Ground Truth를 정의하고 frame 별로 angle vector를 출력해 Absolute Error를 활용해 keypoint 별로 error를 측정

Results



- Error가 10보다 낮아진 상태에서 5 frame간 유지되면 Stand 상태가 될 때 count
- 만약 바르지 못한 자세라면, Error가 가장 큰 부분에 대해 코칭

2. Multimodal modeling을 통한 Sleep stage prediction

기여도 100% ●●●●●

- ✓ 2021.06 ~ 2021.11
- ✓ NYX 인턴십에서 단독 프로젝트로 진행

ML

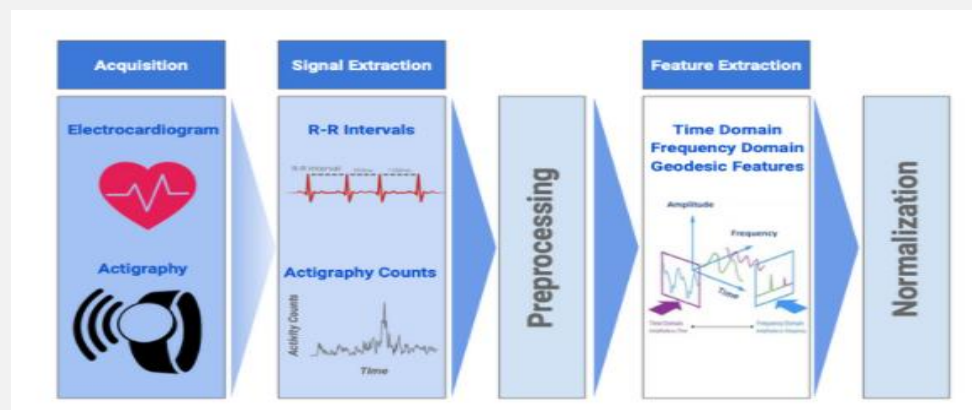
Medical data

Problem

임상시험 피험자의 심전도와 뇌파로 수면 단계 분류

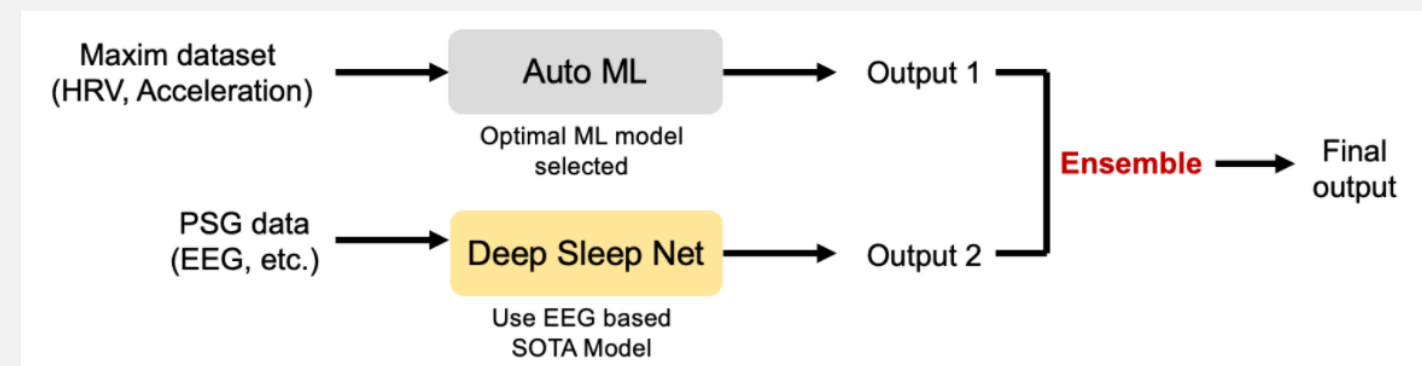
Main idea

HRV + Actigraphy feature fusion



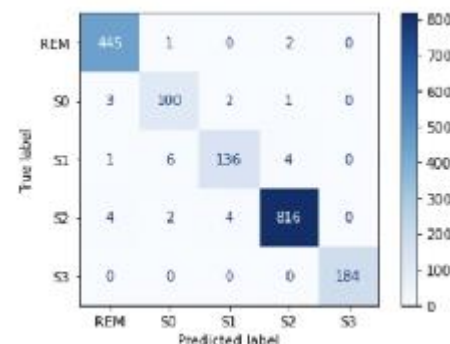
임상시험 피험자의 HRV(Heart Rate Variability)와 Actigraphy(가속도계) data를 fusion 해 multimodal sensor data 생성

Model output ensemble



임상시험 피험자의 HRV data와 EEG data(뇌파)를 각각 분류기에 넣어 학습하고 fusion strategy를 통해 concat

Results



- 각 임상시험 피험자 별 Sleep stage (Wake, REM, S1, S2, Deep sleep) 다중분류 수행

3. Object Detection을 이용한 **ADAS system** 개발

기여도 70%



- ✓ 2021.06 ~ 2021.11
- ✓ 2021 한이음 프로젝트로 진행 - 장려상 수상

CV

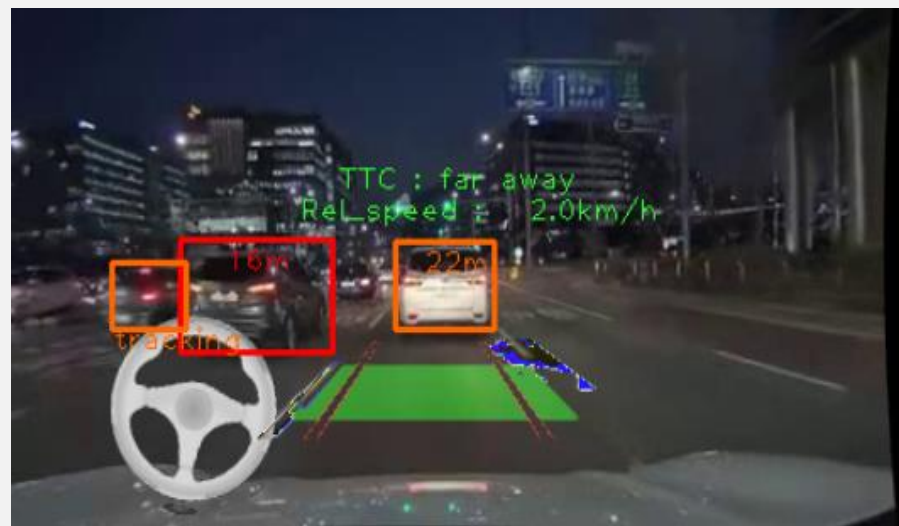
Self-driving

Problem

Object detection을 이용한 차선 이탈 정보, 앞 차 충돌 경보 등의 ADAS system을 출력

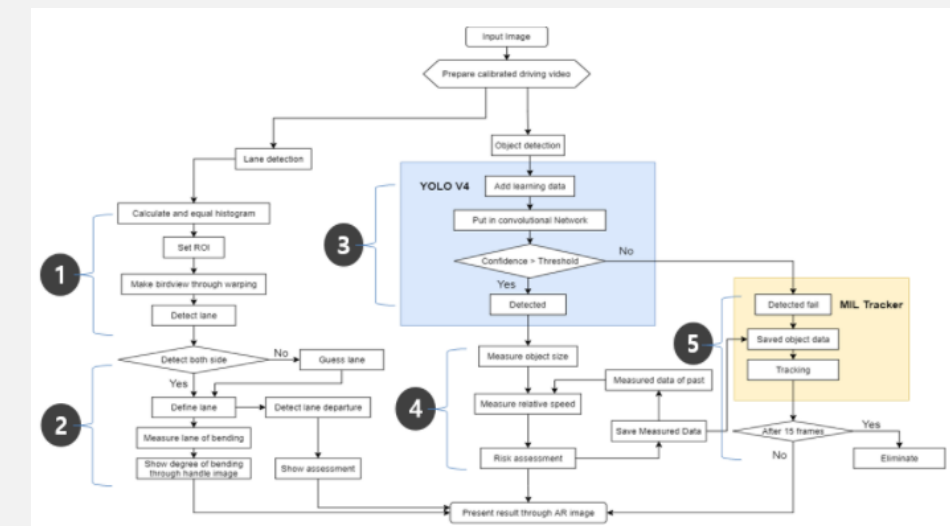
Main idea

차선 이탈 정보 & 앞 차량 Detection



- 거리에 따른 위험도 순서로 Detection line 색 지정
- 차선 검출에는 canny detection을 주로 사용하고, 앞선 차량 detection에는 YOLO-v4 model을 사용

Architecture



- 전통 detection 차선검출 방법과 YOLO계열 딥러닝 모델을 동시에 사용
- Object line을 이용해 거리, 상대속도, TTC를 측정하고 위험도를 알림

Thank You.



Daeun Lee