

# 인공지능을 공부하는 윤도연입니다.

Master of Mathematics, Kwangwoon University  
ydy89899@gmail.com

Doyeon Yoon

# Projects



## NLP projects

1. Text Classification
2. KLUE : Relation Extraction
3. KLUE : Dialogue State Tracking(DST)

## ETC. projects

4. Sensor data Anomaly detection
5. Medical / CT-images Segmentation
6. Deep Knowledge Tracing(DKT)
7. Image Classification

그 외의 활동

# Projects



## NLP projects

1. Text Classification
2. KLUE : Relation Extraction
3. KLUE : Dialogue State Tracking(DST)

## ETC. projects

4. Sensor data Anomaly detection
5. Medical / CT-images Segmentation
6. Deep Knowledge Tracing(DKT)
7. Image Classification

그 외의 활동

# 1. Senticle : 뉴스 데이터 기반 주가 예측

포항공과대학교 인공지능연구원 교육 / 18.09 ~ 18.10 (1.5개월) / Team



## Overview

- ✓ 특정 기업의 당일 뉴스 데이터를 통해 익일 주가를 예측하는 모델 구현 & 앱 구현(Binary Text Classification)

## Dataset

- ✓ 기업별 뉴스 데이터 5년치 크롤링 후 정제를 통한 학습 데이터 구축

## Model

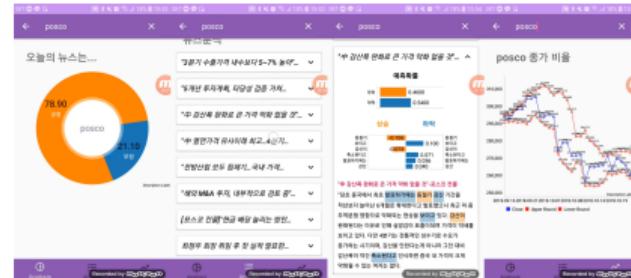
- ✓ Architecture : 1DCNN
- ✓ Tokenizer : Soynlp
- ✓ Embedding : RandomVector, Fasttext, Word2Vec
- ✓ Skills : Tensorflow V1, Android, LIME, NLP

## Role

- ✓ Preprocessing(Labeling, Tokenizing)
- ✓ Modeling

## Result

- ✓ Validation set(8:2)에 대하여 Acc : 73%
- ✓ Backtesting results were not good.
- ✓ 앱 구동 화면



## Link

- ✓ Github : <https://github.com/ydy8989/senticle-proj>
- ✓ App play : <https://youtu.be/syQfQGFAAZ0>

# 2. KLUE : Relation Extraction

네이버 부스트캠프 AI Tech / 21.04.12 ~ 21.04.23 (2주) / Solo



## Overview

- ✓ 문장 속 두 Entity의 관계를 분류하는 task

## Dataset

- ✓ Input : sentence, entity1, entity2

	sentence	entity_01	entity_02	label
0	영국에서 사용되는 스포츠 유틸리티 자동차의 브랜드로는 랜드로버(Land Rover)...	랜드로버	자동차	17
1	선거에서 민주당은 해산 전 의석인 230석에 한창 못 미치는 57석(지역구 27석...	민주당	27석	0
2	유럽 축구 연맹(UFEA) 집행위원회는 2014년 1월 24일에 열린 회의를 통해 ...	유럽 축구 연맹	UFEA	6
3	동생 공격수 차지의 부친과 시즌 초 활약한 강수 일의 형제, 시즌 중반에 영입한 세로...	강수일	공격수	2
4	왕립령 왕은 1237년에서 1247년 사이 수코타이의 왕 피룬 씨 인트라릿과 쓰양...	파룬 씨 인트라릿	8	

- ✓ Output : 42개의 relation classes 중 1개의 class

```
[{"관계_없음": 0, "인물:배우자": 1, "인물:직업/직책": 2, "단체:모회사": 3, "인물:소속단체": 4, "인물:동료": 5, "단체:별칭": 6, "인물:출신분야/국적": 7, "인물:부모님": 8, "단체:분사_국가": 9, "단체:구성원": 10, "인물:기타_전족": 11, "단체:창립자": 12, "단체:주주": 13, "인물:사망_일시": 14, "단체:상위_단체": 15, "단체:본사_주(도)": 16, "단체:제작": 17, "인물:사망_원인": 18, "인물:출생_도시": 19, "단체:분사_도시": 20, "인물:자녀": 21, "인물:제작": 22, "단체:하위_단체": 23, "인물:별칭": 24, "인물:형제/자매/남매": 25, "인물:출생_국가": 26, "인물:출생_일시": 27, "단체:구성원_수": 28, "단체:자회사": 29, "인물:기주_주(도)": 30, "단체:제산일": 31, "인물:기주_도시": 32, "단체:창립일": 33, "인물:종교": 34, "인물:기주_국가": 35, "인물:용의자": 36, "인물:사망_도시": 37, "단체:정치/종교성향": 38, "인물:종교": 39, "인물:사망_국가": 40, "인물:나이": 41}
```

- ✓ Datasets : <https://klue-benchmark.com/tasks/70/data-description>

## Model

- ✓ MLM : XLM-RoBERTa, KoELECTRA, Bert
- ✓ Preprocessing : EDA(Easy Data Augmentation), Back Translation, TEM(Typed Entity Marker)
- ✓ Skills : Pytorch, Huggingface, Tensorboard

## Link

- ✓ Github : <https://github.com/bcaitech1/p2-klue-ydy8989>
- ✓ Notion : <https://www.notion.so/whydo/KLUE-Relation-Extraction-4708eefeb1f849ac8771806898e97333>

## Rank

- ✓ ACC : 79.9% | Public LB : 46등

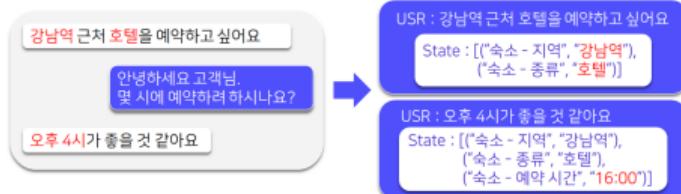
# 3. KLUE : Dialogue State Tracking(DST)

네이버 부스트캠프 AI Tech / 21.04.26 ~ 21.05.21(4주) / Team



## Overview

- ✓ 유저와의 대화에서 미리 시나리오에 의해 정의된 정보인 Slot과 매 턴마다 그에 속할 수 있는 Value의 집합인, 대화 상태 (Dialogue State)를 매 턴마다 추론하는 task



## Dataset overview

- ✓ 데이터는 Json format이며, 예측해야는 state는 "Domain-Slot-Value"의 pair로 구성되어 있다. 아래의 한국어 KLUE 벤치마크 데이터 셋의 일부로 대회를 진행하였다.
- ✓ Datasets : <https://klue-benchmark.com/tasks/73/data/description>

## Dataset I/O

- ✓ Input : Dialogue 내에서 User와 System 발화 1턴
- ✓ Output : Domain-Slot-Value로 분류되는 User 발화의 state pair

Domain-Slot                                      Value  
↓    ↓  
{ "숙소-가격대" : ["저렴", "적당", "비싼", "none", "dontcare"],  
"숙소-지역" : ["동쪽", "서쪽", "남쪽", "북쪽", "none", "dontcare"],  
"숙소-주차가능" : ["yes", "no", "none", "dontcare"],  
... }

- Domain - 5 Classes
- Slot - 45 Classes
- Value - 데이터에 따라 가변

## Metric

- ✓ Joint Goal Accuracy » Slot Accuracy » Slot F1 Score

# 3. KLUE : Dialogue State Tracking(DST)

네이버 부스트캠프 AI Tech / 21.04.26 ~ 21.05.21(4주) / Team



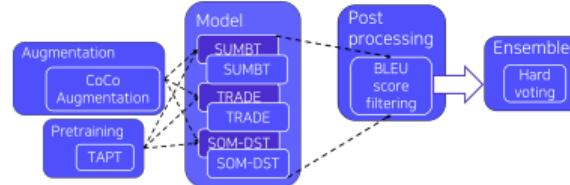
## Model

- ✓ **Architecture :**
  - ▶ Ontology based : SUMBT
  - ▶ Open Vocab based : TRADE, SOM-DST
- ✓ **Preprocessing :**
  - ▶ CoCo Augmentation
  - ▶ TAPT
- ✓ **Post-Preprocess :** BLEU score filtering(Idea)
- ✓ **앙상블 :** HardVoting
- ✓ **Skills :** Pytorch, Tensorboard, Huggingface

## Role

- ✓ SOM-DST 구현 및 fine tuning
- ✓ SOM-DST + CoCo / TAPT

## Pipeline



## Link

- ✓ **Team Github :** <https://github.com/bcaitech1/p3-dst-teamed-st>
- ✓ **Team Notion :** <https://www.notion.so/whydo/Dialogue-State-Tracking-81883b2d7c0246c7b2d3ea2cb766ba62>

## Rank

- ✓ **JGA :** 0.8344 | **Public LB :** 1등
- ✓ **JGA :** 0.7355 | **Private LB :** 1등

# Projects



## NLP projects

1. Text Classification
2. KLUE : Relation Extraction
3. KLUE : Dialogue State Tracking(DST)

## ETC. projects

4. Sensor data Anomaly detection
5. Medical / CT-images Segmentation
6. Deep Knowledge Tracing(DKT)
7. Image Classification

그 외의 활동

# 4. Samsung Semiconductor Smart Interlock

Hbee Co. AI team, Machine learning engineer / 19.09 ~ 20.03 (6개월)

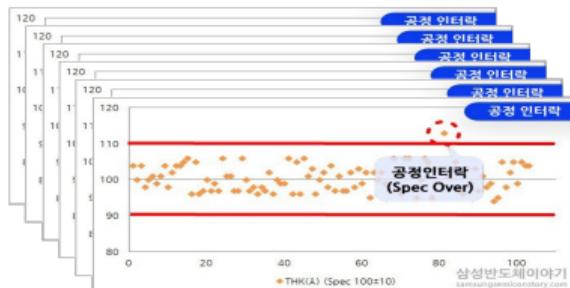


## Overview

- ✓ 반도체 공정 내 시계열 센서데이터의 이상 징후를 탐지하고 세부적으로 7개 Class로 분류하는 프로젝트

## Dataset

- ✓ 이상 징후 예후 시점 과거 30일 간의 센서 로그
- ✓ **INPUT** : Timeseries raw data 2000건
- ✓ **OUTPUT** : 진성(3개 클래스) / 가성(4개 클래스)
- ✓ **Raw data** : 원본 / **Image data** : Plotting



## Pipeline

- ✓ Two-way ensemble(image data + raw data)
  - ▶ Plotting image : SGAN
  - ▶ Raw data : Stacked Auto-encoder + Linear regression
- ✓ ensemble model  $\Leftrightarrow$  Tensorflow serving  $\Leftrightarrow$  Flask

## Model

- ✓ **Architecture** : SGAN, Auto-Encoder, Linear Regression
- ✓ **Metric importance** : Precision  $\gg$  f1-score 순서
- ✓ **Skills** : Keras, Flask, Tensorflow Serving

## Result

- ✓ **Precision** : 99% / **F1-Score** : 83%
- ✓ 삼성전자 반도체 공정 전라인 적용

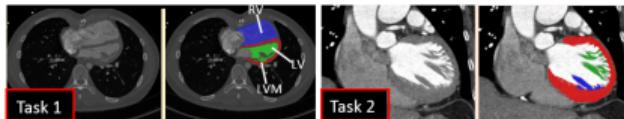
# 5. Cardiac CT images Segmentation Competition

포항공과대학교 인공지능연구원 Intern / 18.11 ~ 18.01 (2.5개월) / Team



## Overview

- ✓ 심장 Segmentation task competition



## Dataset

- ✓ 3차원 이미지(.mha format)
- ✓ Sample 1건, Train 100명분, Test 100명분  
*Train data, Test data는 블라인드(서버 내에 존재)*

## Model

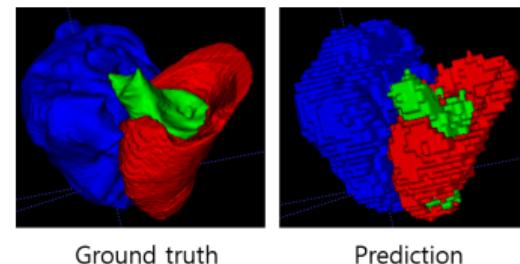
- ✓ **Architecture** : 3D-UNet
- ✓ **Preprocessing** : Patching(Resize, Crop, Augmentation), Resampling(Voxel spacing), Intensity Windowing
- ✓ **Skills** : Tensorflow, Keras, SimpleITK, Docker, Segmentation

## Role

- ✓ Preprocessing(Resampling, Intensity windowing)
- ✓ Fine tuning

## Result

- ✓ Dice Coefficient : 73%(winner:79%)



## Link

- ✓ **Github** : [https://github.com/ydy8989/Cardiac\\_Segmentation](https://github.com/ydy8989/Cardiac_Segmentation)

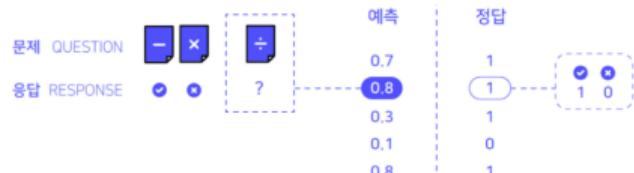
# 6. Deep Knowledge Tracing(DKT)

네이버 부스트캠프 AI Tech / 21.05.24 ~ 21.06.15(4주) / Team



## Overview

- ✓ 사용자(학생)의 문제 풀이 정보를 통해 개인화 된 지식 상태를 추적하는 Task



## Dataset

- ✓ Input : 7442명의 사용자에 대한 문제 풀이 데이터

userID	assessmentItemID	testId	answerCode	Timestamp	KnowledgeTag
0	0	A060001001	A060000001	1 2020-03-24 00:17:11	7224
1	0	A060001002	A060000001	1 2020-03-24 00:17:14	7225
2	0	A060001003	A060000001	1 2020-03-24 00:17:22	7225
...	...	...	...	...	...
2266581	7441	A030071005	A030000071	0 2020-06-05 06:50:21	438
2266582	7441	A040165001	A040000165	1 2020-06-21 01:06:39	8836

- ✓ Output : 특정 문제 조건에 대한 정답 여부

## Model

- ✓ Architecture : LGBM, SAINT, LastNQuery, GKT
- ✓ Preprocessing : Feature Engineering
- ✓ Skills : Pytorch, Tensorboard, Transformer, Bert
- ✓ Metric : AUROC

## Link

- ✓ Team Github : <https://github.com/bcaitech1/p4-dkt-decayt>
- ✓ Team Notion : <https://www.notion.so/Home-b263b1f24c3147ac9f8f2544178d66f6>

## Rank

- ✓ AUROC : 0.842 | Public LB : 2등
- ✓ AUROC : 0.845 | Private LB : 4등

# 7. Mask Image Classification

네이버 부스트캠프 AI Tech / 21.03.29~21.04.08(2주) / Solo



## Overview

- ✓ 다양한 성별 / 나이 / 마스크 착용 형태에 따른 이미지  
분류 task

## Dataset

- ✓ 2700명에 대한 이미지(한 명당 7가지 : 5종류 마스크,  
1 Normal, 1 턱스크)
- ✓ Input : 384x512 크기의 마스크를 쓴 얼굴 이미지



- ✓ Output : 18개의 클래스

- ▶ 마스크 착용 여부(no착용 / 착용 / 절반 착용)
- ▶ 성별(남 / 여)
- ▶ 나이(30이하 / 30-60 / 60이상)

## Model

- ✓ Architecture : EfficientNet b4, ResNet
- ✓ Preprocessing : Augmentation, Label filtering
- ✓ Skills : Pytorch, Tensorboard, Stratified kfold
- ✓ Metric : F1-score

## Link

- ✓ Github :  
<https://github.com/bcaitech1/p1-img-ydy8989>

## Rank

- ✓ F1-score : 0.6800 | Public LB : 158등
- ✓ F1-score : 0.6738 | Private LB : 152등

# Projects



12

## NLP projects

1. Text Classification
2. KLUE : Relation Extraction
3. KLUE : Dialogue State Tracking(DST)

## ETC. projects

4. Sensor data Anomaly detection
5. Medical / CT-images Segmentation
6. Deep Knowledge Tracing(DKT)
7. Image Classification

## 그 외의 활동

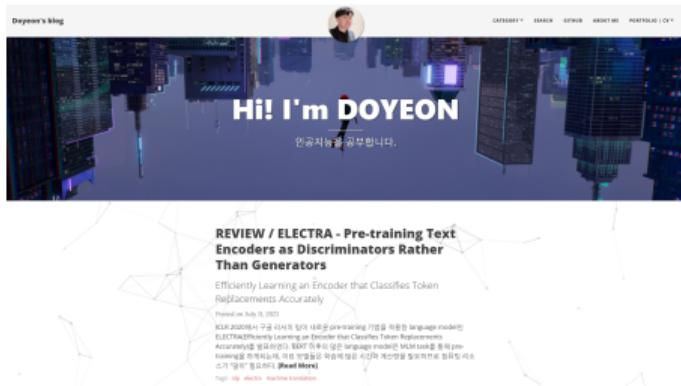
# 그 외의 활동



## 블로그

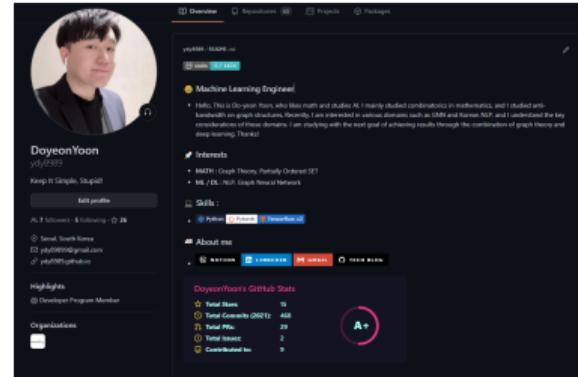
- ✓ 운영중인 블로그에 논문 리뷰 및 학습 내용을 정리하고 있습니다. 앞선 프로젝트 이외의 사항들은 아래 링크를 통해 확인하실 수 있습니다.

<https://ydy8989.github.io/>

A screenshot of a blog homepage titled "Doyeon's blog". The header features a profile picture of a man and the text "Hi! I'm DOYEON" with the Korean subtitle "안녕하세요, 소개합니다.". Below the header, there is a large image of a futuristic city skyline at night. A main article is visible with the title "REVIEW / ELECTRA - Pre-training Text Encoders as Discriminators Rather Than Generators" and a snippet of text: "Efficiently Learning an Encoder that Classifies Token Replacements Accurately". At the bottom of the page, there is a footer with the text "Posted on July 15, 2021" and a link "Read More".

## Github

- ✓ [https://github.com/ydy8989/](https://github.com/ydy8989)

A screenshot of a GitHub profile page for "ydy8989". The profile picture is the same as on the blog. The bio reads: "Hello, I'm a 20-year-old who likes math and studies AI. I mainly studied combinatorics in mathematics, and I studied algorithms on graph structures. Recently, I am interested in various domains such as DBMS and Natural NLP, and I understand the key components of these domains. I am studying with the most goal of extracting results through the combination of graph theory and deep learning." The interests section includes "MATRIX : Graph Theory, Partially Ordered SET", "ML / DS : NLP, Graph Neural Network", and "Mathematics : Combinatorics, Graph Theory". The "About me" section highlights "Developer Program Member". The "Highlights" section shows "DoyeonYoon's GitHub Stats" with the following data:

Total Stars	19
Total Commits (2021)	468
Total PRs	29
Total Issues	2
Contributed to	9

A large "A+" rating icon is displayed on the right.

## Notion

- ✓ <https://www.notion.so/whydo/Doyeon-Yoon-05603016086c4b3ca954cf2b6c64e46f>

Thank You!