DATA SCIENCE

PREDICTING EMOTIONS FROM SPEECH SIGNALS AND TEXT

2277044 LEESEUNGHYEON

2491007 KIMYEEUN

2491021 LEEBADA

2491036 JUTAEIN

PROJECT

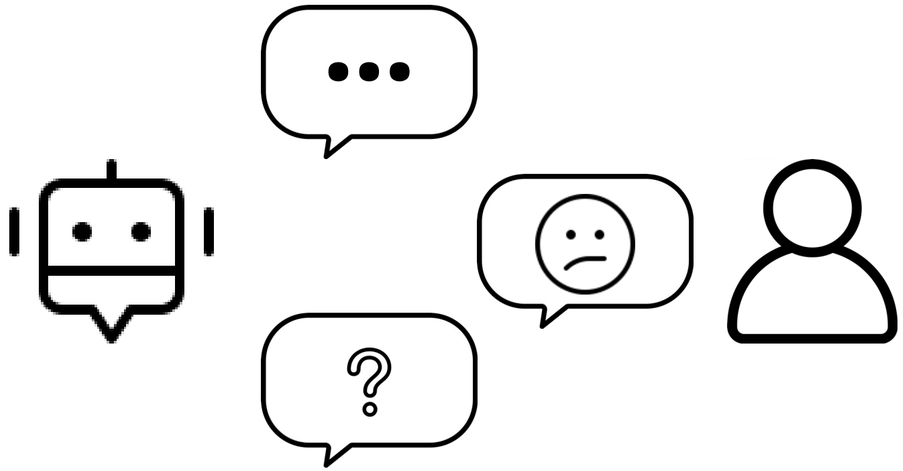
INDEX

01 Problem Definition
02 Data Review
03 Preprocessing
04 Improving Method - Data
05 Model Comparison
06 Improving Method - Modeling
07 Final Result
08 Q&A

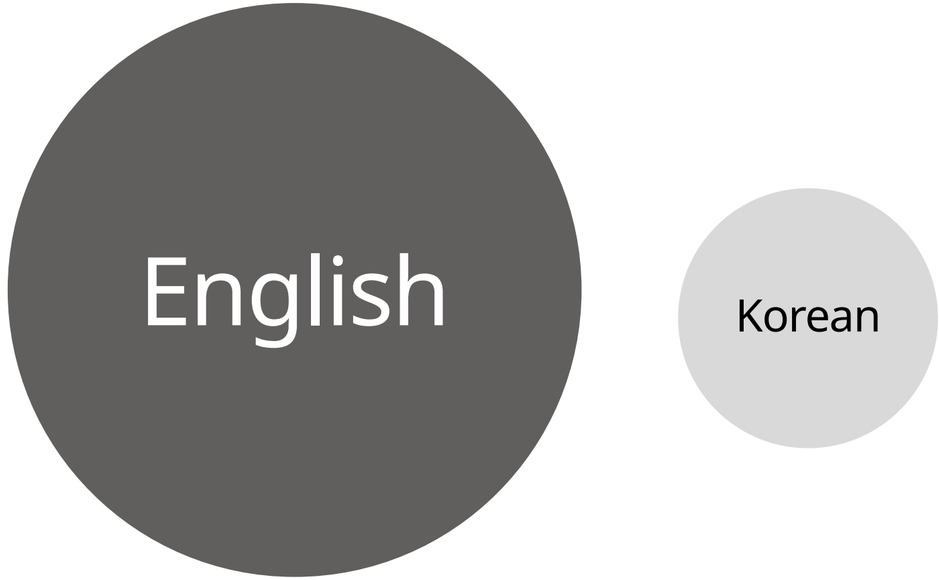
PROBLEM DEFINITION

Business Value

BACKGROUND

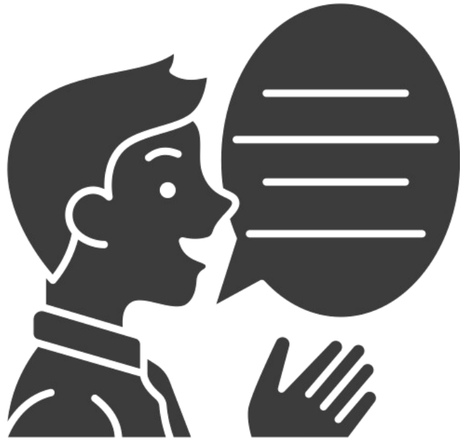


Increasing
Usage Of ChatGPT



Shortage Of
Korean Detection Model

BACKGROUND



Korean speech datasets

×

nonverbal speech signal

×

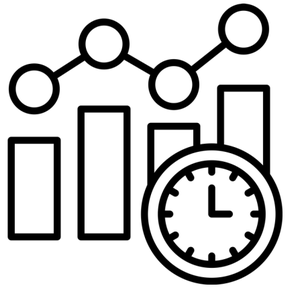
textual cues

Detecting emotinal state

PROBLEM DIFINITION

Detecting emotinal state

integration with **wearable** devices



real-time emotion prediction

personalized feedback delivery

DATA REVIEW

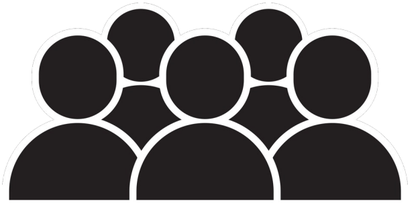
DATA

DATA ACQUISITION

by the **KAIST Artificial Intelligence Research Institute**

DATA REVIEW

19,374 Korean wav file



Labeled by 5 people

- happiness
- angry
- fear
- neutral
- disgust
- sadness
- surprise

DATA REVIEW

	wav_id	발화문	상황
0	5f4141e29dd513131eacee2f	헐! 나 이벤트에 당첨 됐어.	happiness
1	5f4141f59dd513131eacee30	내가 좋아하는 인플루언서가 이벤트를 하더라고. 그래서 그냥 신청 한번 해봤지.	happiness
2	5f4142119dd513131eacee31	한 명 뽑는 거였는데, 그게 바로 내가 된 거야.	happiness
3	5f4142279dd513131eacee32	당연히 마음에 드는 선물이니깐, 이벤트에 내가 신청 한번 해본 거지. 비싼 거야. ...	happiness
4	5f3c9ed98a3c1005aa97c4bd	에피타이저 정말 좋아해. 그 것도 괜찮은 생각인 것 같애.	neutral

1번 감정	1번 감정세기	2번 감정	2번 감정세기	3번 감정	3번 감정세기	4번 감정	4번감정세기	5번 감정	5번 감정세기	나이	성별
angry	2	surprise	2	happiness	2	happiness	2	happiness	2	48	female
neutral	0	happiness	2	happiness	2	happiness	2	happiness	2	48	female
angry	2	happiness	2	happiness	2	happiness	2	happiness	2	48	female
angry	2	happiness	2	happiness	2	happiness	2	happiness	1	48	female
happiness	2	happiness	1	happiness	2	happiness	1	happiness	1	48	female

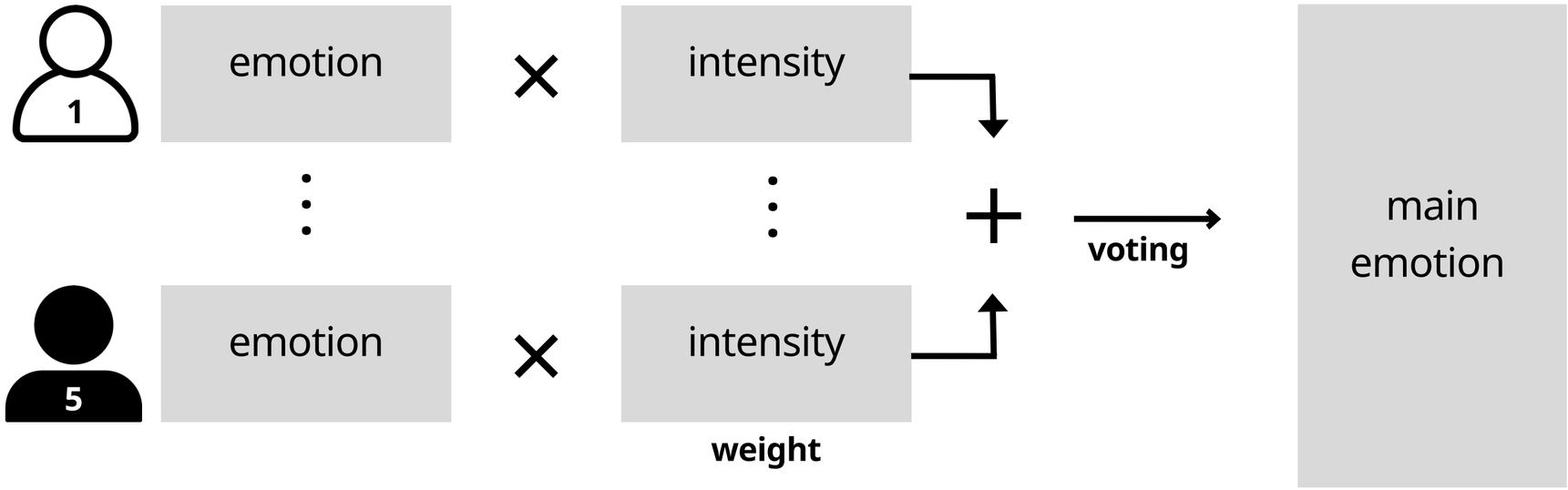
	emotion	count
0	sadness	4245
1	happiness	3390
2	disgust	3370
3	surprise	3023
4	angry	2747
5	fear	2173

PREPROCESSING

DATA

LABELING

weighted voting



main emotion ∈ {happiness, anger, disgust, fear, sadness, surprise}

OPEN SMILE

eGeMAPS (88 features)

example

feature	mean	
MFCC	timbre, 13 related features	positive/negative classification
F bandwidth	vocal tract openness	larger → exaggerated
Spectral Flux	rate of spectral change	higher → anxiety, fear
HNR	Harmonic-to-Noise Ratio	higher → calmness and positive

more **applicapable** to physical properties;
more **interpretable** compared to other method

IMPROVING METHOD

DATA

IMPROVING METHOD

01

balanced acc across label

02

feature selection

BALALANCED ACC

balanced data set

extra data

5.2th + new 4th

samplin g

smote down

modified down

emotion	count
angry	4218
disgust	4218
sadness	4218
happiness	3556
fear	3454
surprise	3300

17,531 > 22,964

BALALANCED ACC

balanced data set

angry	0.43
disgust	0.60
fear	0.70
happiness	0.75
sadness	0.62
surprise	0.68

angry	0.55
disgust	0.58
fear	0.64
happiness	0.76
sadness	0.61
surprise	0.66

FEATURE SELECTION

feature importance < 0.5

feature	mean	
`equivalentSoundLevel_dBp`	Average energy (in dB)	
`loudness_sma3_pctlrange0-2`	Loudness in 0–2%	whispering/unstable
`F2amplitudeLogRelF0`	F2 amplitude (normalized)	F2; tongue front–back position and tension
`loudness_sma3_percentile80.0`	Loudness in 80–100%	
`F3amplitudeLogRelF0`	F3 amplitude (normalized)	F3; subtle articulatory (tongue shape and nasality)
`F1amplitudeLogRelF0`	F1 amplitude (normalized)	F1; degree of mouth openness
`loudness_sma3_amean`	Average loudness	

overall

/

feature	mean
`loudness_sma3_amean`	Average loudness

intersection of emotion-specific

MODEL COMPARISON

MODELING

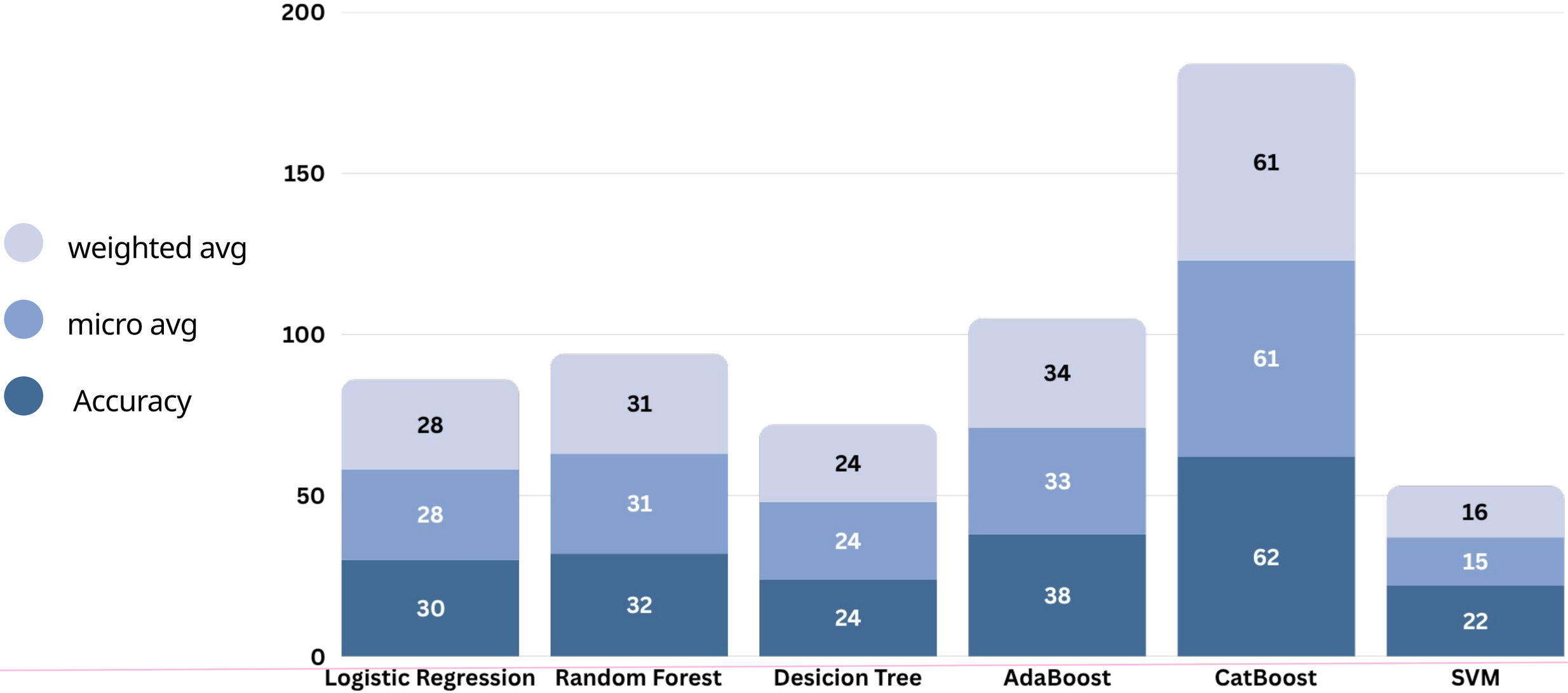
MODEL COMPARISON

Precision

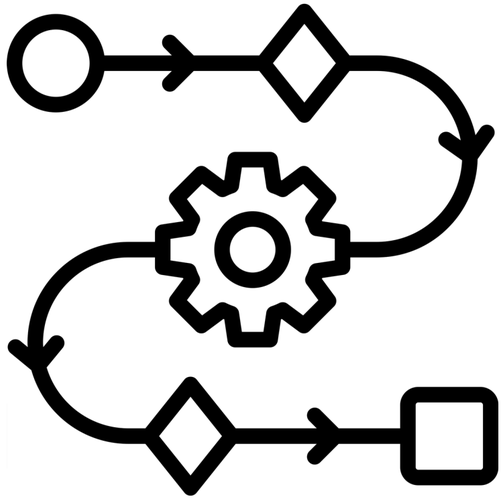
	angry	disgust	fear	happiness	sadness	surprise
Logistic Regression	0.23	0.27	0.27	0.31	0.32	0.31
Random Forest	0.25	0.28	0.42	0.37	0.33	0.31
Decision Tree	0.19	0.25	0.21	0.25	0.27	0.24
SVM	0	0.19	0	0.21	0.25	0.29
AdaBoost	0	0.67	0.93	0.26	0.39	0.6
CatBoost ✅	0.51	0.56	0.76	0.69	0.55	0.67

MODEL COMPARISON

F1-score

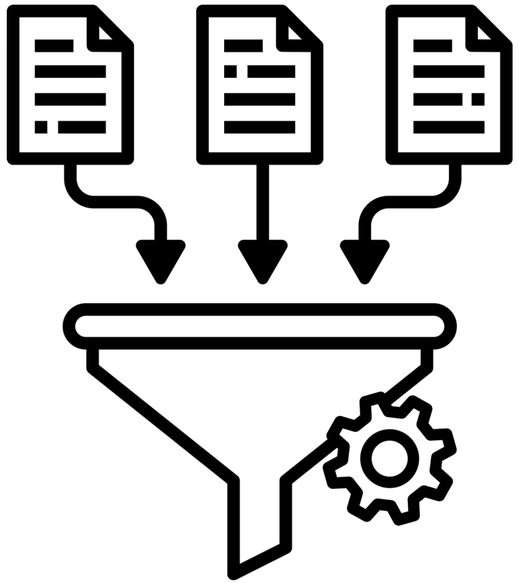


CAT BOOST



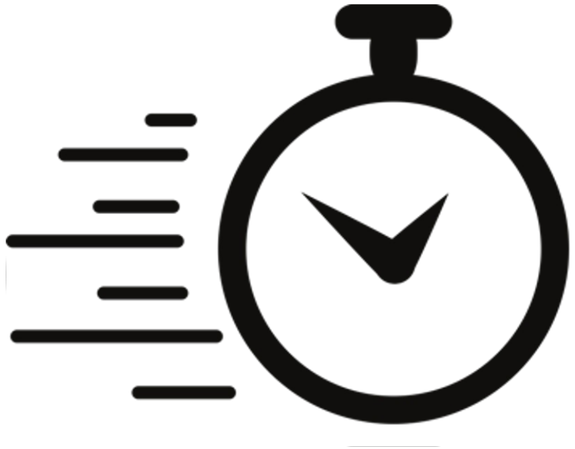
Automatic Processing

no need to encoding



Built-in Regularization

prevent overfitting



Fast speed

GPU + parallel

IMPROVING METHOD

MODELING

IMPROVING METHOD

01

Okt vs Mecab

02

Early Fusion vs Late Fusion

OKT VS MECAB

	Okt	**MeCab (mecab-ko)** ✅
Speed	Moderate	**Very fast**
Tokenization Unit	Word-based + whitespace handling	**Fine-grained morpheme segmentation**
POS tag	about 15 tags	**over 50 tags**
Korean OOV	Simple whitespace-based	**Dictionary-based (ipa-ko) + custom dictionaries**

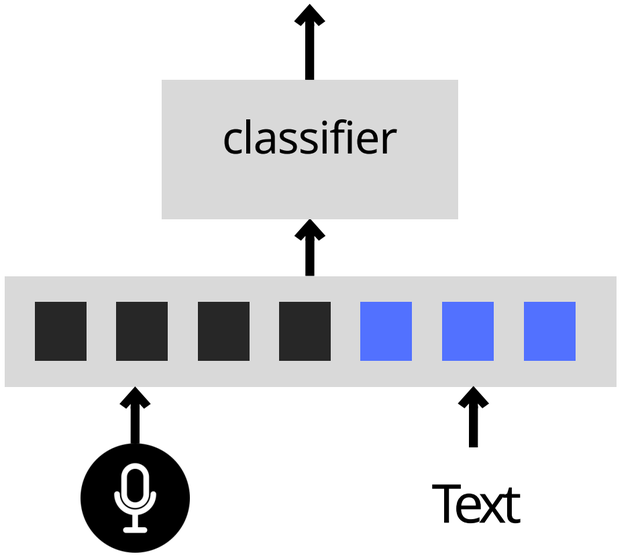
OKT VS MECAB

인공지능기반음성감정분석모델이 구현되었다.

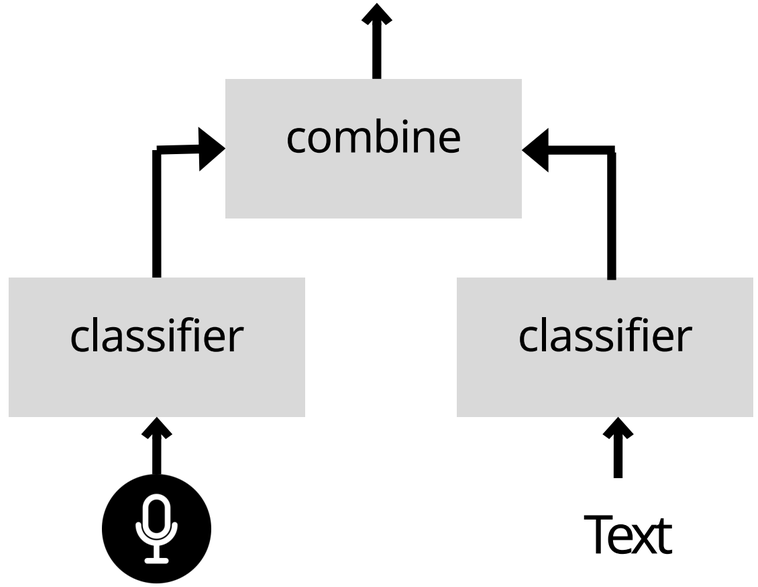
⬇

Okt	['인공지능기반음성감정분석모델', '이', '구현', '되', '었', '다', '.']
MeCab	['인공', '지능', '기반', '음성', '감정', '분석', '모델', '이', '구현', '되', '었', '다', '.']

EARLY FUSION VS LATE FUSION



Early Fusion



Late Fusion

EARLY FUSION VS LATE FUSION

	precision	recall	f1-score	support
angry	0.56	0.54	0.55	843
disgust	0.61	0.56	0.58	844
fear	0.71	0.59	0.64	691
happiness	0.74	0.78	0.76	711
sadness	0.54	0.68	0.61	844
surprise	0.68	0.64	0.66	660
accuracy			0.63	4593
macro avg	0.64	0.63	0.63	4593
weighted avg	0.63	0.63	0.63	4593

✅ **Early Fusion**

acc : 63 %

	precision	recall	f1-score	support
angry	0.57	0.51	0.54	843
disgust	0.60	0.57	0.59	844
fear	0.68	0.57	0.62	691
happiness	0.73	0.77	0.75	711
sadness	0.54	0.70	0.61	844
surprise	0.64	0.59	0.61	660
accuracy			0.62	4593
macro avg	0.63	0.62	0.62	4593
weighted avg	0.62	0.62	0.61	4593

Late Fusion

acc : 62%

FINAL RESULTS

FINAL RESULTS

```
TF-IDF shape: (22964, 500)
=== CatBoost Classification Report ===
              precision    recall  f1-score   support

       angry       0.56      0.54      0.55       843
     disgust       0.62      0.55      0.59       844
        fear       0.70      0.61      0.65       691
   happiness       0.74      0.78      0.76       711
     sadness       0.55      0.69      0.61       844
    surprise       0.68      0.65      0.67       660

    accuracy                           0.63      4593
   macro avg       0.64      0.64      0.64      4593
weighted avg       0.64      0.63      0.63      4593
```

FINAL RESULTS

-PREDICTION

UTTERANCE

“도둥리서 배 탈 생각도 말아라! 저걸, 확 씨."

TRUTH LABEL

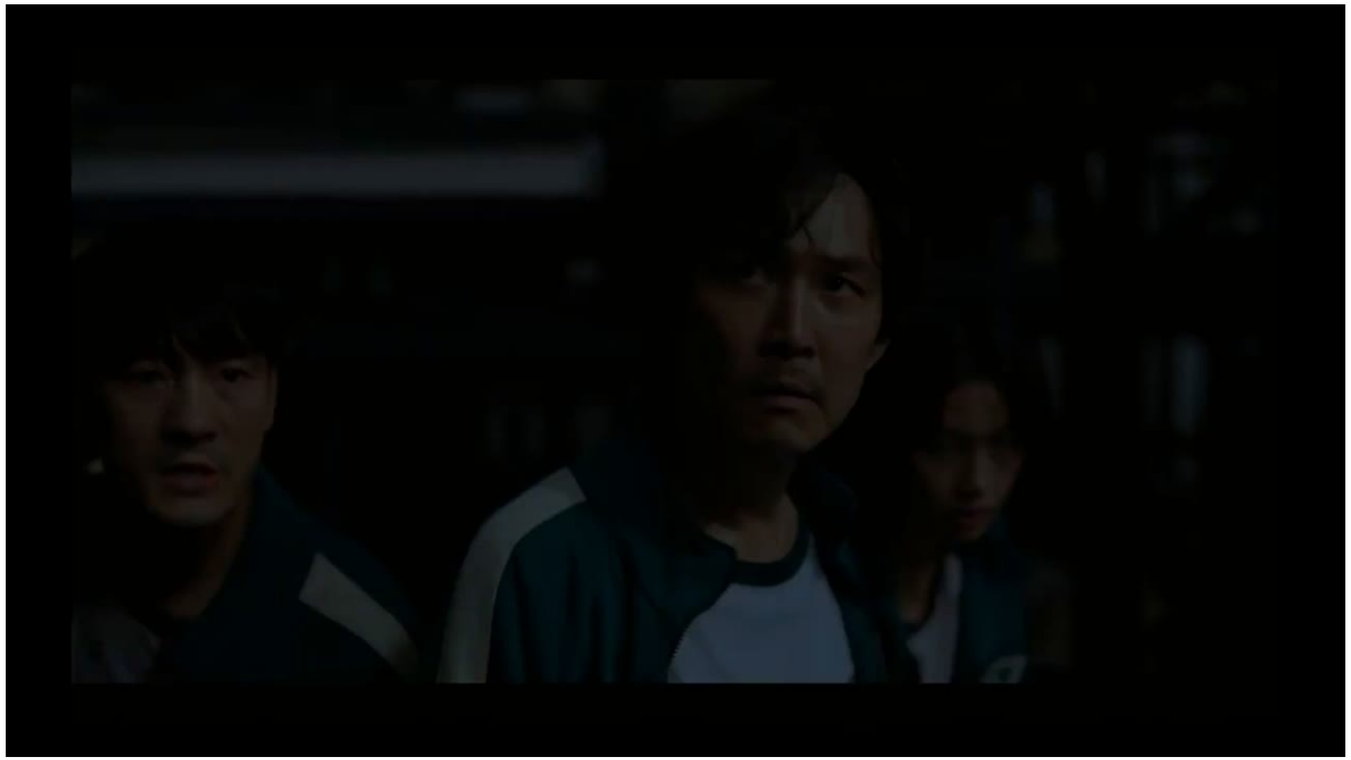
“ANGRY”

PREDICTED LABEL

“ANGRY”

FINAL RESULTS

-PREDICTION



UTTERANCE

“제발 그만해, 나, 나
무서워.
이러다가는 다 죽어!"

TRUTH LABEL

“FEAR”

PREDICTED LABEL

“FEAR”

FUTUTE DIRECTION

FUTUTE DIRECTION

01

Limitations of Machine Learning

02

OpenSMILE; Increased Feature Dimensionality

THANK YOU


Q&A