팀 2

# 음성 감정 분석을 통한 자살 예방 프로젝트 **최종 발표**

팀2 장서윤 권유진 송창용 윤성식

# Contents

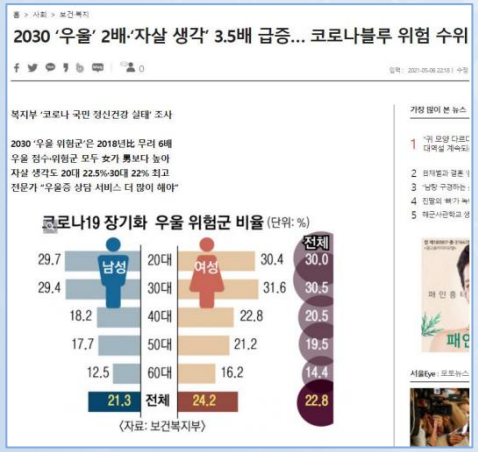


# 01 주제 설명 및 연구 가치

## 연구 동기

### 우울증 환자 증가

- 최근 코로나로 인한 우울증 지수 및 자살 징조 증가
- 우울감이 높아지는 사람들은 다른 사람에게 이를 표현하지 않으려는 경향이 있음
- 표정이나 행동으로는 캐치하기 힘든 우울감을 상징하는 특징들을 음성을 통해 잡아낼 수 있음


2030 '우울' 2배·'자살 생각' 3.5배 급증... 코로나블루 위험 수위

복지부 '코로나 국민 정신건강 실태' 조사

2030 '우울 위험군'은 2018년比 무려 6배
우울 점수·위험군 모두 女가 男보다 높아
자살 생각도 20대 22.5%·30대 22% 최고
전문가 "우울증 상담 서비스 더 많이 해야"

코로나19 장기화 우울 위험군 비율 (단위: %)

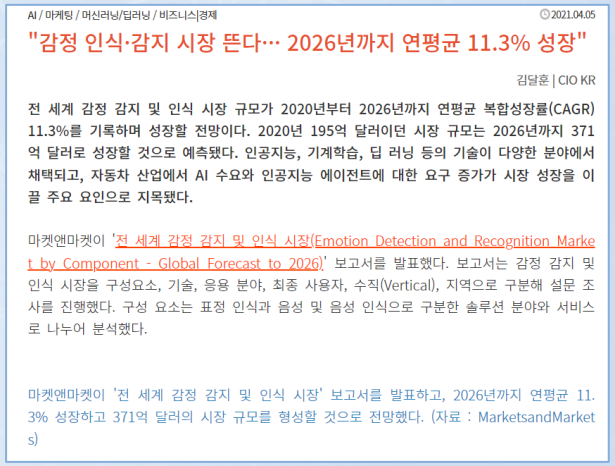
| 남성 | | 여성 | | 전체 |
|---|---|---|---|---|
| 29.7 | 20대 | 30.4 | | 30.0 |
| 29.4 | 30대 | 31.6 | | 30.5 |
| 18.2 | 40대 | 22.8 | | 20.5 |
| 17.7 | 50대 | 21.2 | | 19.5 |
| 12.5 | 60대 | 16.2 | | 14.4 |
| 21.3 | 전체 | 24.2 | | 22.8 |

(자료: 보건복지부)

<코로나 블루로 인한 우울증 증가>

### 감정 인식/감지 시장 증가 및 연구 확대

- 마켓 앤 마켓의 보고서에 따르면 2020년 부터 2026년까지 감정 감지 및 인식 시장 규모가 195억 달러에서 371억 달러로 성장할 것으로 예측
- 음성 감정인식 분야에서 연구가 활발히 이루어지고 있으나 한국어 데이터셋을 기반으로 한 연구는 비교적 많지 않음.[1)]


AI / 마케팅 / 머신러닝/딥러닝 / 비즈니스|경제 2021.04.05

"감정 인식·감지 시장 뜬다… 2026년까지 연평균 11.3% 성장"

김달훈 | CIO KR

전 세계 감정 감지 및 인식 시장 규모가 2020년부터 2026년까지 연평균 복합성장률(CAGR) 11.3%를 기록하며 성장할 전망이다. 2020년 195억 달러이던 시장 규모는 2026년까지 371억 달러로 성장할 것으로 예측됐다. 인공지능, 기계학습, 딥 러닝 등의 기술이 다양한 분야에서 채택되고, 자동차 산업에서 AI 수요와 인공지능 에이전트에 대한 요구 증가가 시장 성장을 이끌 주요 요인으로 지목됐다.

마켓앤마켓이 '전 세계 감정 감지 및 인식 시장(Emotion Detection and Recognition Market by Component - Global Forecast to 2026)' 보고서를 발표했다. 보고서는 감정 감지 및 인식 시장을 구성요소, 기술, 응용 분야, 최종 사용자, 수직(Vertical), 지역으로 구분해 설문 조사를 진행했다. 구성 요소는 표정 인식과 음성 및 음성 인식으로 구분한 솔루션 분야와 서비스로 나누어 분석했다.

마켓앤마켓이 '전 세계 감정 감지 및 인식 시장' 보고서를 발표하고, 2026년까지 연평균 11.3% 성장하고 371억 달러의 시장 규모를 형성할 것으로 전망했다. (자료 : MarketsandMarkets)

<감정 인식 / 감지 시장 정보>

# 01 주제 설명 및 연구 가치

## 주제, 연구가치, 방향성

### 주제 설명

- **연구 주제 :** 감정 인식을 통해 우울증을 진단하거나, 자살 전조 증상을 탐지해 내는 것

### 방향성

- 단일 신호에만 의존하지 않고 음성 데이터와 텍스트 데이터를 동시에 활용하는 방안을 고려할 것[2)]
- CNN, LSTM, Attention 계열의 딥러닝 모델들을 활용

### 연구 가치 및 활용도

- 의학적 진단을 내리는 도구로 사용될 수 있음
- 우울 관련 증상으로 인해 상담이나 치료를 받는 사람에게는 모니터링 도구로 기능할 수 있을 것
- 궁극적으로 한국인의 우울감을 낮추고, 자살률을 떨어뜨리는 데 기여할 것

# 02 데이터셋 및 전처리

## 감성 대화 말뭉치

AI Hub

### 60가지 감정 Category

< 데이터셋 감정 분포>

### 이 데이터셋 사용의 적합성

1. 다중분류 감정에 대한 대화 음성 데이터는 IEMOCAP 등 영어 데이터가 많고, 한국어 데이터는 부족한데 **한국어로 구성된 데이터임**
2. **텍스트 형태의 데이터와 음성 형태의 데이터를 모두 제공**
3. **60가지의 감정이 소분류로 분류되어 우울과 관련한 정교한 감정 분석이 가능**

### 데이터 구조(텍스트 csv파일 + 음성 zip파일)

한 행당 한사람에 대한 데이터로,
성별, 연령 정보와 감정 분류, 인당 4개의 발화문을 포함

| B | C | D | E | F | G | H |
|---|---|---|---|---|---|---|
| 연령 | 성별 | 상황키워드 | 신체질환 | 감정_대분류 | 감정_소분류 | 사람문장1 |
| 청년 | 남성 | 연애, 결혼, 출산 | 해당없음 | 기쁨 | 신이 난 | 아내가 드디어 출산하게 되어서 정말 신이 나. |
| 노년 | 남성 | 건강, 죽음 | 만성질환 | 불안 | 스트레스 받 | 당뇨랑 합병증 때문에 먹어야 할 약이 열 가지가 넘어가니까 스트레스 |
| 청소년 | 여성 | 학업 및 진로 | 해당없음 | 당황 | 당황 | 고등학교에 올라오니 중학교 때보다 수업이 갑자기 어려워져서 당황스 |
| 노년 | 남성 | 재정 | 만성질환 | 기쁨 | 신이 난 | 재취업이 돼서 받게 된 첫 월급으로 온 가족이 외식을 할 예정이야. |
| 노년 | 여성 | 재정 | 만성질환 | 기쁨 | 안도 | 빚을 드디어 다 갚게 되어서 이제야 안도감이 들어. |
| 중년 | 여성 | 재정, 은퇴, 노후준비 | 해당없음 | 불안 | 취약한 | 이제 곧 은퇴할 시기가 되었어. 내가 먼저 은퇴를 하고 육 개월 후에 |
| 중년 | 남성 | 건강 | 해당없음 | 슬픔 | 우울한 | 사십 대에 접어들면서 머리카락이 많이 빠져 고민이야. |
| 노년 | 남성 | 재정 | 만성질환 | 분노 | 구역질 나는 | 이제 돈이라면 지긋지긋해. |

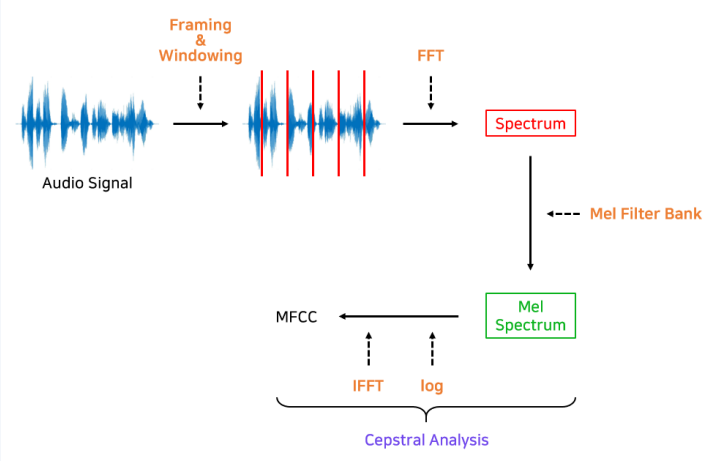
< 실제 데이터셋 구성>

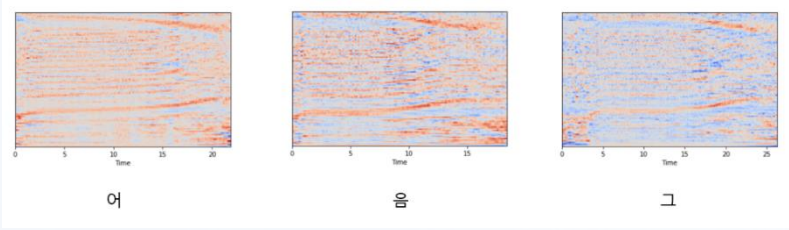
# 02 데이터셋 및 전처리

## Audio Data 전처리

### MFCC (Mel-Frequency cepstral coefficients)

- 음성 특징 추출 방법 중 최근 많이 사용되고 있는 방법 중 하나
- 음성 인식을 위해 가장 먼저 입력 신호에서 노이즈 및 배경 소리로부터 유효한 소리의 특징을 추출 해야함
- 입력 소리 전체를 대상으로 하는 것이 아니라 일정 구간 씩 나누어, 이 구간에 대한 스펙트럼의 특징 추출




<오디오 데이터 전처리: MFCC>

### MFCC 전처리 과정 요약

1. 오디오 신호를 프레임별(보통 20ms - 40ms)로 나누어 FFT(고속 푸리에 변환)를 적용해 Spectrum 산출
2. Spectrum에 Mel Filter Bank를 적용해 Mel Spectrum 산출
3. Mel Spectrum에 Cepstral 분석을 적용해 MFCC 계산

Cepstral 분석 방법은 포먼트들을 연결한 곡선을 Spectral Envelope라고 하는데 이 곡선과 Spectrum을 분리하는 과정에서 MFCC가 도출됨. 이 때 사용되는 수학 알고리즘은 log와 역 고속 푸리에 변환

# 02 데이터셋 및 전처리

## Text Data 전처리

### 사람문장1, 2, 3, 4 병합
- 사람문장 / 시스템문장 쌍으로 이루어진 데이터에서 시스템 문장 제거
- 사람문장 만을 발화 데이터로 인식하기 위해 하나의 인덱스에 대해 문장 1 ~ 4를 병합

### 길이가 짧은 문장 제거
- 발화문장 중 추임새 등이 들어갈 수 있으므로 문장 안 단어의 개수가 2개 이하인 것은 제거

### KoBERT Tokenizer 활용
- 문장을 임의의 Token단위로 구분하기 위해 한국어 데이터로 Pretrained된 KoBERT의 Tokenizer를 활용

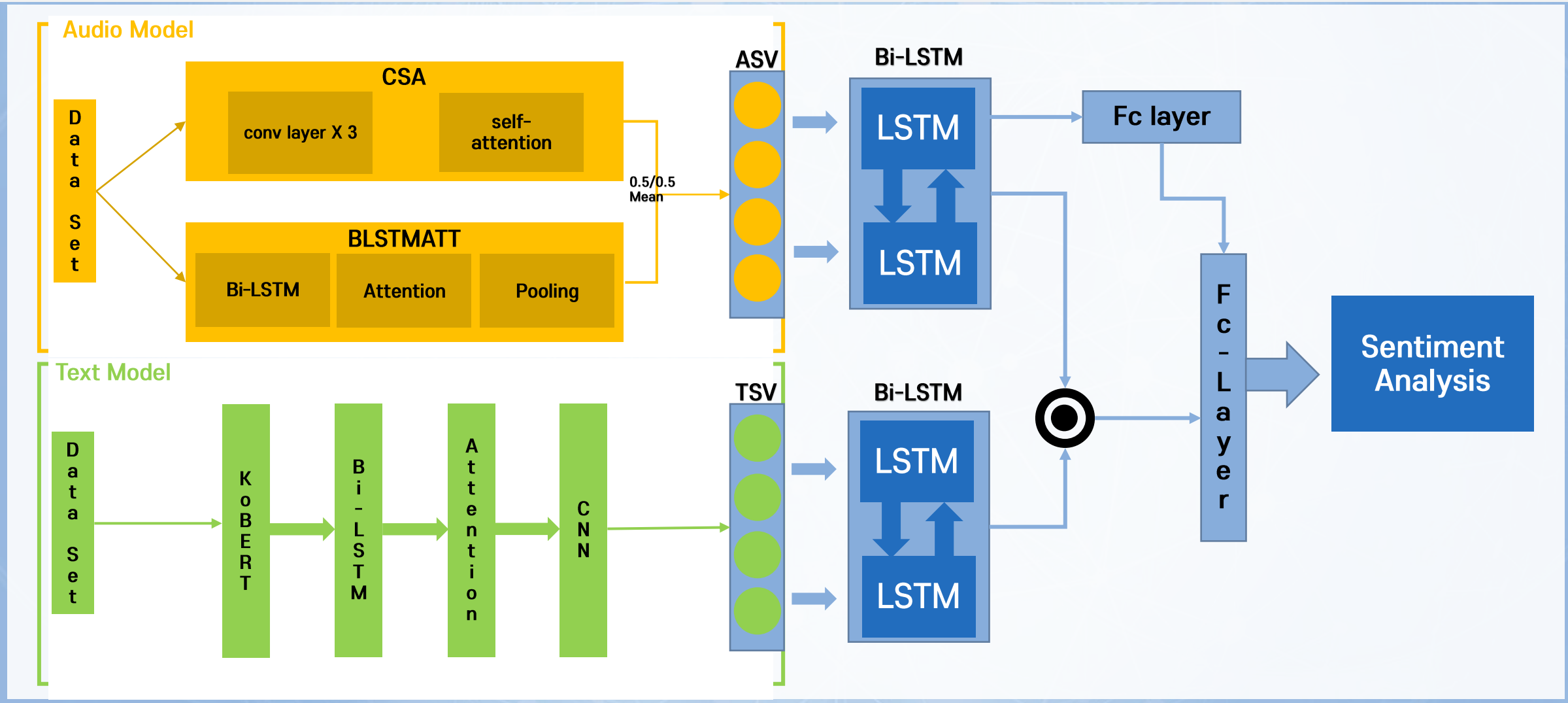
### 불용어 처리
- 조사 접속사 등 Token화 후에 제거하고 싶은 단어의 사전을 임의로 설정하여 제거

| 사람문장1 | 시스템응 | 사람문장2 | 시스템응 | 사람문장3 | 시스템응답 | 사람문장4 | 시스템응답4 |
|---|---|---|---|---|---|---|---|
| 요즘 부모님 | 어떤 일로 | 난 하고 싶 | 부모님과 | 응. 그래도 | 자신이 하고 싶은 일을 목표로 하기로 하셨군요. | | |
| 엄마가 결 | 어머니께 | 엄마도 엄마 | 정말 많 | 일단 좀 자 | 일단 휴식을 가질 생각이시군요. | | |
| 학교에서 | 용기 있는 | 그런데 말 | 그랬군요 | 나는 옳은 | 지금의 상황이 나의 방식으로써 잘 해결될 수 있기를 바라요. | | |
| 이번에 팀 | 팀장님이 | 곧 있으면 | 마음이 | 우선 잘못된 | 잘못된 부분을 잘 수정해서 좋은 결과가 있었으면 좋겠어요. | | |
| 남편이 이 | 많이 화가 | 어떻게 그 | 어떻게 | 나도 변호사 | 자녀분들이랑 함께 하며 슬픔을 극복하시길 바라요. | | |
| 친구들과 | 친구분들 | 이렇게 노 | 지금의 | 우선 다른 | 다른 친구들과 이야기해보려고 하시는군요. | | |
| 직장에서 | 직장에서 | 나는 손도 | 정말 억울하고 마음이 힘드셨을 것 같아요. 오해가 잘 풀리기를 바라요. | | | | |
| 요즘 딸에 | 딸에게 화 | 딸과 더 친 | 딸이 신 | 딸과 좀 더 | 딸과 대화함으로써 관계가 좋아지기를 바라요. | | |
| 언제까지 | 대출금을 | 이제 중년 | 허무하고 | 조용한 카페 | 조용한 카페에 가서 차를 마시면서 여유를 가지고 머리를 식히려고 하시는군요. | | |
| 이 업무는 | 업무를 하 | 이번 주는 | 업무를 | 그동안 너무 | 일을 끝내고 좋은 휴식을 보내길 바라요. | | |
| 나 대학에 | 학원 선생 | 내가 잘 모 | 알기 쉽도록 설명해 주시는 게 학원의 좋은 점이군요. | | | | |
| 작년에 왔 | 이직 제안 | 맞아. 지금 | 너무 속상하시겠어요. 어떻게 하면 기분이 조금 나아질까요? | | | | |
| 나이는 먹 | 경제적인 | 얼마 있으 | 남편분의 | 지금은 여유 | 원활한 생활을 위해서는 경제적 여유가 우선이군요. | | |
| 이번에 약 | 약이 많이 | 의사에게 | 좋아요. 부디 당신에게 좋은 결과가 있기를 바랄게요. | | | | |

<실제 데이터 구성>

# 03 모델링

## 모델 전체 구조



# 03 모델링

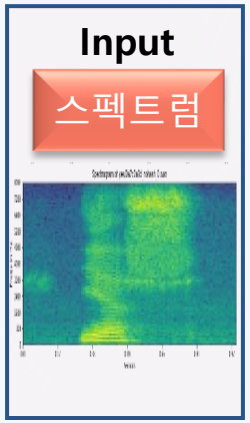
## Audio Model 설명

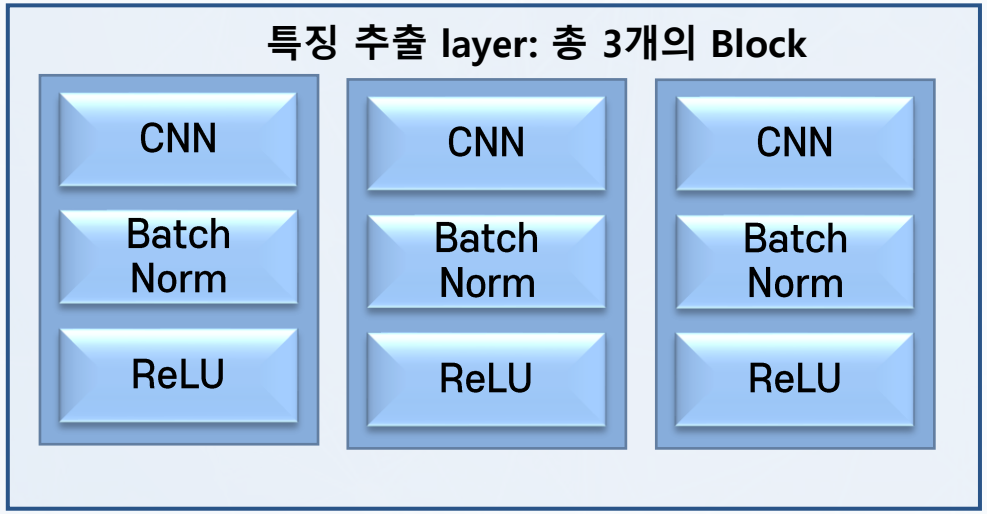
**음성 데이터 분석 모델로 아래 SOTA논문으로부터 모델 구조를 차용**
논문 : Empirical Interpretation of Speech Emotion Perception with Attention Based Model for Speech Emotion Recognition
저자: Md Asif Jalal, Rosanna Milner, Thomas Hain

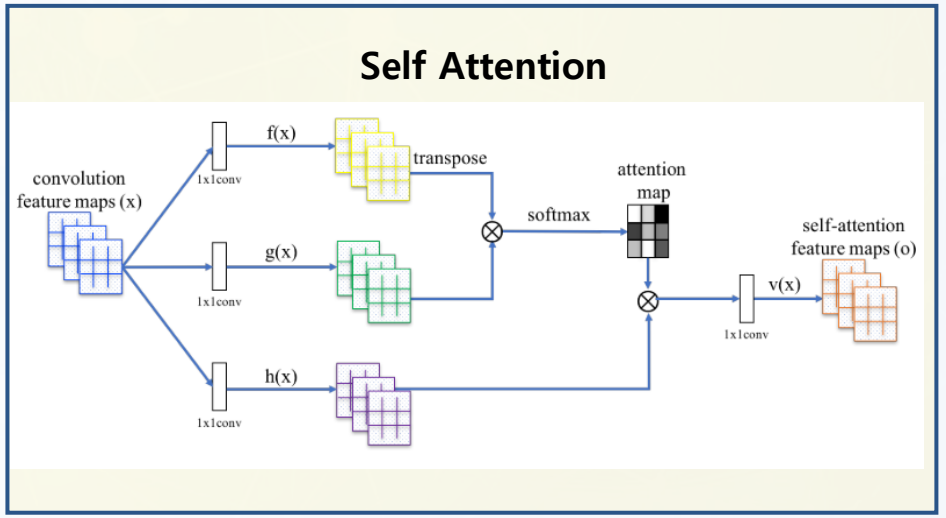
**CSA, BLSTMATT 두 개의 모델을 사용해서 각 모델 구조의 이점을 활용하여 음성데이터를 분석**
**각각의 모델로 분석을 거친 뒤 이 두 모델의 결과 벡터가 0.5/0.5 가중치로 가중 합 되어 텍스트 데이터 분석 모델과 합쳐질 예정**

### 1. CSA Convolutional Self-Attention

공간적 feature를 추출하고 attention을 이용해 feature을 고차원적으로 확장하기 위함



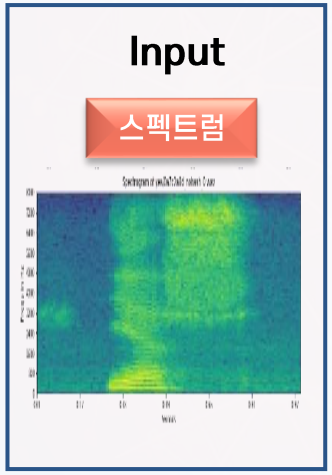



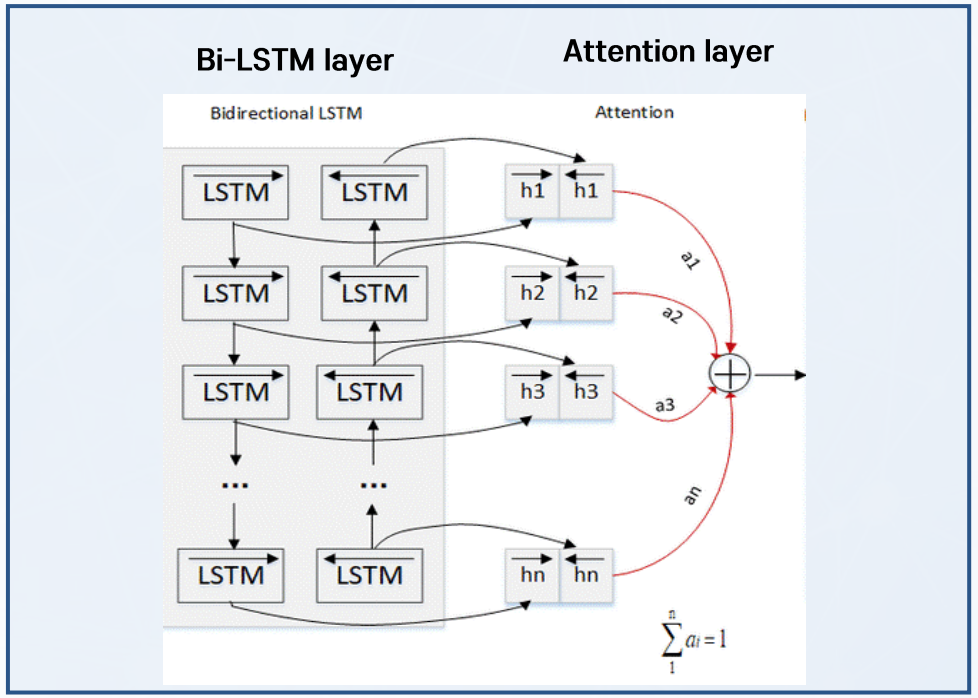
# 모델링

## Audio Model 설명

### 2. BLSTMATT(BLSTM with attention)

이 방법론을 통해 BiLSTM층에서 hidden state를 2개 두어 양 쪽에서 학습함으로 순서에 의한 temporal feature distribution 얻을 수 있고, Attention으로 global mean을 계산함으로써 global 정보를 포착 가능




→ 이후 CSA에서 나온 vector와 BLSTMATT에서 나온 vector를 0.5/0.5 비율로 sum하여 하나의 TSV를 생성

# 03 모델링

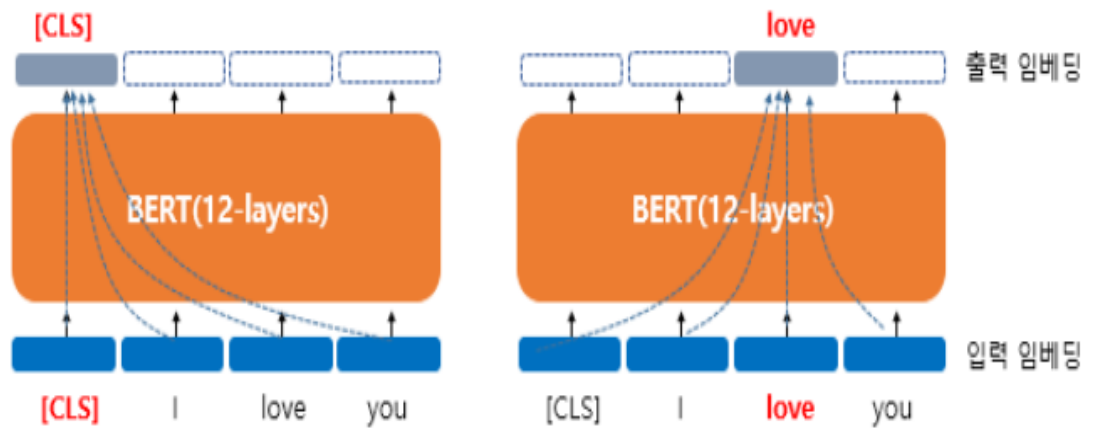
## Text Model 설명

### 1. KoBERT

Transformer의 Encoder를 활용한 BERT 모델을 사용함으로써 문맥을 반영한 임베딩을 추출

다만, BERT를 그대로 사용하는 것이 아니라 fine tuning은 거치지 않고 BERT에서 생성된 임베딩벡터 만을 사용

**일반 BERT모델은 영어를 사용해 학습된 것 이므로, 한국어 위키를 통해 학습된 KoBERT모델을 사용**




<BERT 입력과 출력 임베딩>

### Tokenizer

문장을 적절한 Token으로 분할하기 위해 Tokenizer 사용
BERT의 경우 단어보다 더 작은 단위로 Token화를 하는 WordPiece Tokenizer를 사용
**이 경우에도 한국어로 학습된 KoBERT Tokenizer를 사용**

| Word | Token(s) |
|---|---|
| surf | ['surf'] |
| surfing | ['surf', '##ing'] |
| surfboarding | ['surf', '##board', '##ing'] |
| surfboard | ['surf', '##board'] |

# 모델링

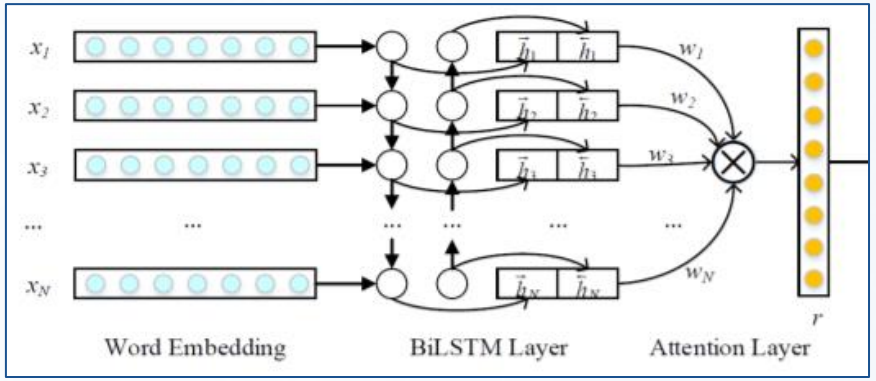
## Text Model 설명

### 2. BiLSTM + Attention

Audio Model에서 처럼 Bidirectional LSTM과 Attention weight를 사용해 temporal feature distribution 얻고,
Attention으로 global mean을 계산함으로써 global 정보를 포착할 수 있음
하지만 Word Embedding 값을 Input으로 사용하는 점이 다름

### 3. CNN

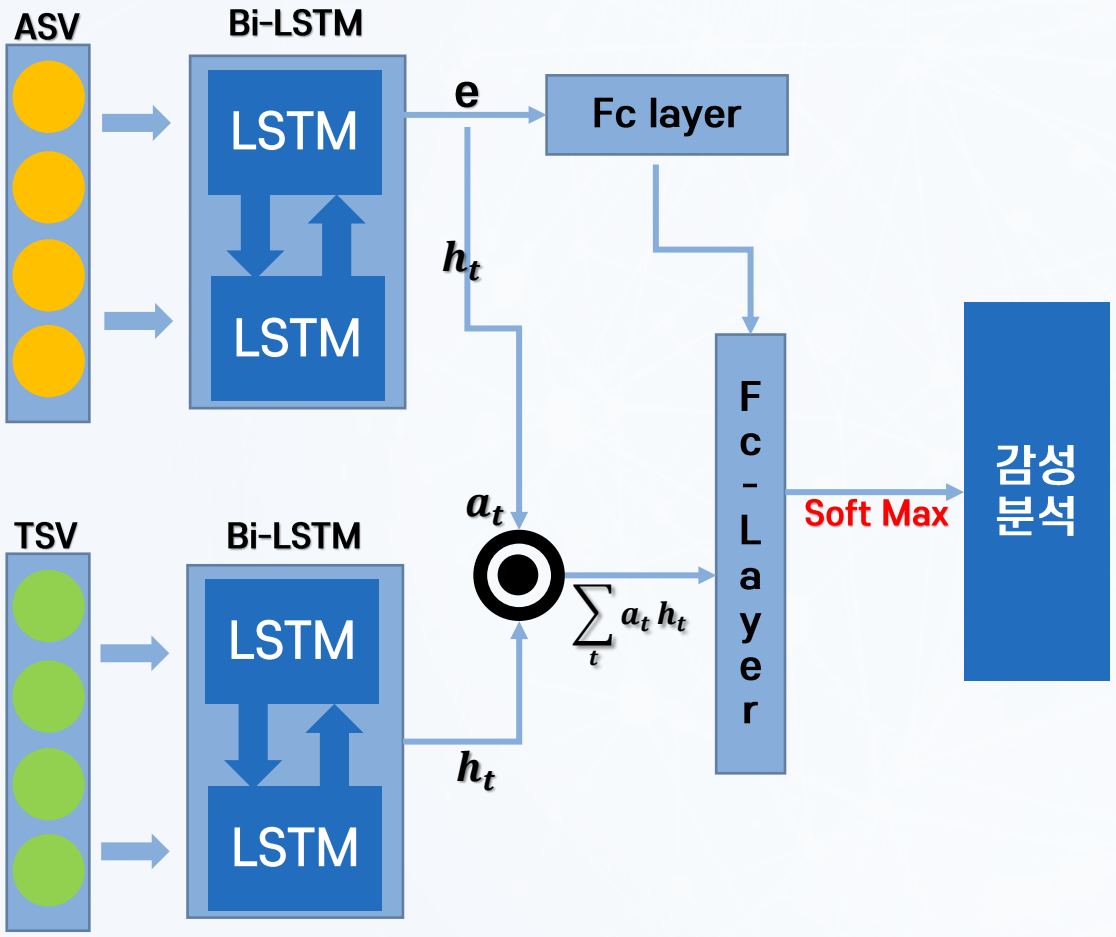
2번에서 산출된 representation 1d vector에서 feature map(다양한 filter)을 뽑기 위해 Convolution layer를 사용한다.
이렇게 나온 다양한 feature map들은 concat되어 TSV를 구성한다.




<BiLSTM + Attention>

# 03 모델링

## Model 병합 및 평가




### ASV와 TSV의 결합 및 Output 도출

**BiLSTM**

ASV와 TSV 시퀀스의 중간 출력 층을 보존 하기 위해 사용되었고, output 값과 hidden state등을 산출하는 역할

**산출된 hidden state들의 계산**

- ASV의 마지막 encoding vector는 $e$ (context vector)로 dot product를 위한 유사도 계산식인 $a_t$를 구하기 위해 사용됨 또한 그 자체로 FC를 통과해 $A$ vector를 생성

$$a_t = \frac{\exp(e^T h_t)}{\sum_t exp(e^T h_t)} \qquad Z = \sum_t a_t h_t$$

- TSV의 hidden state는 $e$ 와 유사도를 계산하여 $a_t$를 만들고 이를 weight로 하여 weighted sum 값인 $Z$를 생성 이후 $A$와 $Z$를 concat 하여 softmax를 사용해 확률값을 계산함

$$\widehat{y_{i,j}} = softmax(concat(Z, A)^T M + b)$$

- 요약: attention mechanism에 영감을 받아 Audio와 Text modal 데이터 에서 어느 부분이 강한 감정정보를 가지고 있는지 계산 후 softmax를 취하는 과정

# 03 모델링

## Model 병합 및 평가

### Hyperparmeter & Loss function

기준 Epochs: **50**

Batch-size: **32**

FC-layer의 activation function: **ReLU**

Optimizer: **Adam**

Loss function: **Cross Entropy(NLL Loss)**

BiLSTM neuron의 수: **200**

하이퍼 파라미터 설정의 기준을 잡기 위해
앞선 논문의 하이퍼 파라미터 참고

### 평가지표

1. MACRO F1

F1 score가 2 class의 classification을 위한 것이기 때문에 사용하지 못한다는 것을 고려하여 각 레이블에 동일한 가중치를 부여해 평가할 수 있는 MACRO F1지수를 사용

$$Macro\ F1 = \frac{\sum_{1}^{n} F_{1n}}{n}$$

2. WA, UA

Prediction을 평가하는 기본적인 평가지표인 Accuracy를 사용하는데 있어 각 class에 대한 가중치를 고려한 경우와 그렇지 않은 경우 2가지 다 채택

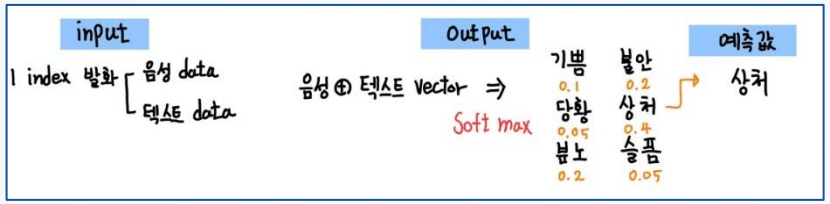
$$UA = \frac{TP + TN}{P + N}, \qquad WA = \frac{1}{2}\left(\frac{TP}{P} + \frac{TN}{N}\right)$$

동일한 데이터를 사용한 모델이 없기 때문에 객관적인 성능 비교는 불가하지만 다른 Emotion Recognition 모델에서 빈번히 사용되는 평가지표들을 채택

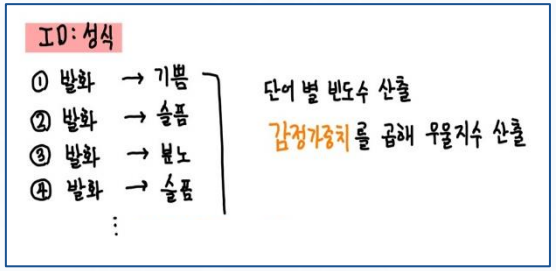
# 04 결론

## 우울증 판별 및 모델 활용

### 최종 output




<Input 부터 pred값 까지의 Flow>




<하나의 ID로 부터 우울 여부 판단을 위한 예측 값 누적 과정>

구축한 모델은 한 index(하나의 긴 발화 문장) 당 1개의 감성분석 결과를 내보냄
하지만 한 번의 발화만으로 우울증을 판별(우울 지수를 산출)하는 것은 적합하지 않다고 판단
따라서 여러 번의 대화를 통해 해당 사용자의 감정을 저장해 놓음

### 감정 가중치 계산

- 대부분의 우울 판별 지표는 설문조사를 통한 점수로 표현되고, 그 점수에 대한 합이 특정 기준 이상일 때 조금 우울함, 많이 우울함 등 우울 정도를 나타낼 수 있음

→ 따라서 우울함을 정량화 한 우울지수를 구해야 할 필요성이 있음

- 감정에 따라 우울함에 기여하는 바가 상이하기 때문에 각각의 가중치를 달리하여 우울 지수를 산출해야 함[3)]

| 분노 | 기대 | 역겨움 | 공포 | 즐거움 | 슬픔 | 놀라움 | 신뢰 |
|---|---|---|---|---|---|---|---|
| 24.28 | 5.71 | 15 | 9.28 | 0.71 | 39.28 | 3.57 | 2.14 |

<가중치 설정 예시>

# 04 결론

## 우울증 판별 및 모델 활용

### 가중치 계산을 위한 데이터셋 가정

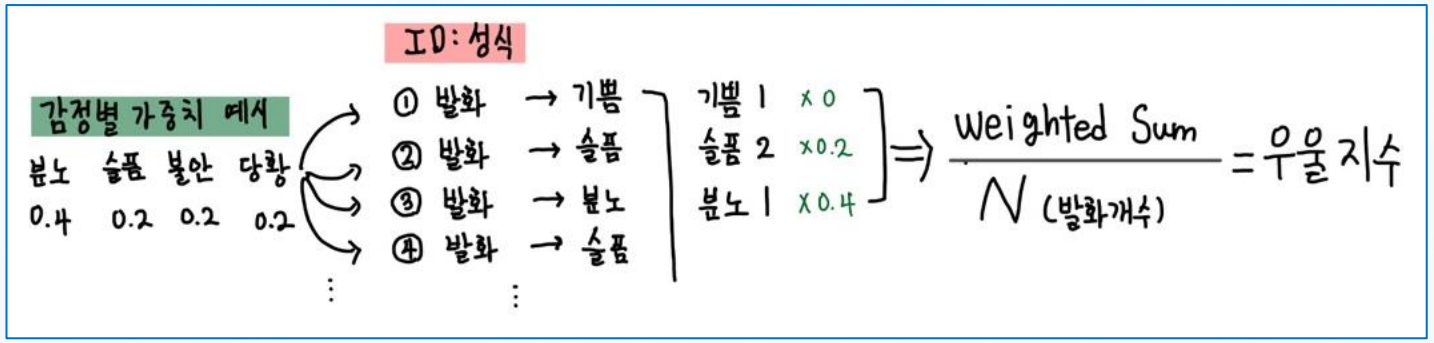
앞의 논문에서 나온 바와 같이 우울증을 판별하기 위한 우울 지수를 구하는 데에 가중치를 구하기위한 근거 필요
수집한 AIHub data와 같은 형태를 가진 발화문 및 labeling data를 우울증 환자들에게서 만 특정해 구한다고 가정

| 번호 | 연령 | 성별 | 상황키워드 | 신체질환 | 감정_대 | 감정_소분류 | 사람문장1 | 시스템응 | 사람문장2 | 시스템응 | 사람문장3 | 시스템응답 | 우울증환자여부 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 23093 | 청년 | 여성 | 진로, 취업, 직장 | 해당없음 | 분노 | 툴툴대는 | 요즘 부모님과 많( | 어떤 일로 | 난 하고 싶은 일이 | 부모님과 | 응. 그래도 ( | 자신이 하고 | o |
| 32848 | 청소년 | 남성 | 가족관계 | 해당없음 | 슬픔 | 비통한 | 엄마가 결국 집을 | 어머니께서 | 엄마도 엄마만의 | 정말 많( | 일단 좀 자이 | 일단 휴식을 | o |
| 35590 | 청소년 | 남성 | 학교폭력/따돌림 | 해당없음 | 불안 | 조심스러운 | 학교에서 한 친구( | 용기 있는 | 그런데 말을 하고 | 그랬군요 | 나는 옳은 일 | 지금의 상황 | o |
| 169 | 청년 | 남성 | 진로, 취업, 직장 | 해당없음 | 당황 | 죄책감의 | 이번에 팀장님이 ( | 팀장님이 | 곧 있으면 인턴에 | 마음이 ( | 우선 잘못된 | 잘못된 부분 | o |
| 38435 | 중년 | 여성 | 재정, 은퇴, 노후준비 | 해당없음 | 분노 | 노여워하는 | 남편이 이혼할 때 | 많이 화가 | 어떻게 그럴 수가 | 어떻게 ( | 나도 변호사 | 자녀분들이 | o |

<우울증 환자들에게서 수집한 데이터셋(가정)>

해당 Dataset에서 각 감정의 빈도 수를 파악한 후 빈도의 비율에 따라 1을 총합으로 해 가중치 생성
→ 실제 우울증 환자에게서 나온 감정 비율이기 때문에 우울 지수 판별에 유용한 가중치가 될 것이라 판단

### 우울지수 산출




<우울지수 계산 Flow)>

# 04 결론

## 우울증 판별 및 모델 활용

### 우울 위험 정도를 파악하기 위한 Threshold 설정

마인드디텍터는 웹 어플리케이션 형태의
'AI 기반 우울감 자가진단 서비스'이며
사용법은 다음과 같습니다.

1) 유저가 문항을 읽고 생각과 감정을 적는다.

2) AI 자연어처리 기술을 통해 답변을 다섯 가지 감정으로 분류

3) 분류된 감정을 통해 유저의 우울 지수를 측정

4) 유저의 특성과 부합하는 자살률 공공데이터를
바탕으로 우울 위험 가중치를 부여

5) 최종 집계된 유저의 우울 지수 결과와
서비스 전체 유저의 평균 우울 지수를 제공

**AI 기반 우울 자가진단 서비스를 제공하는 "마인드디텍터"**

- 마인드 디텍터에서는 실제로 AI기반 자가진단 서비스를 운영하고 있음

- 왼쪽과 같은 사용법들 중 최종 집계된 유저의 우울 지수 결과와 서비스 전체 유저의 평균 우울 지수를 제공하는 방법을 차용

- 현재 사용자(ID)에 대한 우울 지수 결과와 우울증 환자들의 평균 우울 지수를 함께 제공해서 이 평균보다 우울 지수가 높다면 우울증을 의심할 수 있음

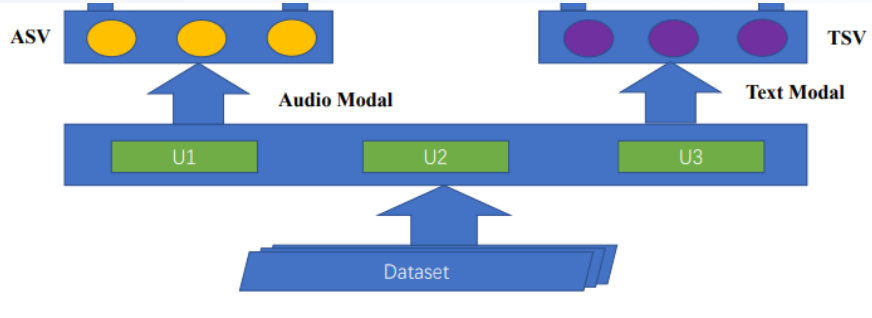
즉, 우울증 환자들의 평균 우울지수를 Threshold로 설정

# 04 결론

## 프로젝트 차별성

### Multi Modal Emotion Recognition

텍스트나 오디오 중 한 가지 모델만을 사용하는 것이 아니라 더 확실히 문맥을 이해하고 감정을 구분할 수 있도록 두 가지 모델을 사용한 점




### 한국어 기반 모델

단순 영어 Dataset을 사용한 것이 아니라 학습 목적과 부합하는 최신 AI Hub Data를 통해 한국어 모델을 학습 시켰기 때문에 모델을 활용하는데 있어 한국어의 특성을 더 잘 이해할 수 있을 것

# 04 결론

## 프로젝트 차별성

★★★

**우울증 판단과**
**자체적인 우울지수 설정**

1. 실제 환자들의 감정 분류 결과를 데이터 셋으로 가정하여 가중치로 활용함으로써 우울증 판단의 당위성을 확보

2. 같은 방식을 감정 소분류에 적용하면서 보다 정교하게 분석이 가능할 것

3. 기존 우울 진단에 검사지나 심리적 연구 이론들이 바탕이 된 데에 비해 추상적 개념을 정량화된 결과로 확인 가능함


Thank You
감사합니다

# APPENDIX

## 인용논문

3p. 제목: **DNN 을 이용한 End-to-End 한국어 음성 감정 인식,**
**저자:** 1이정필, 2류휘정, 2장두성, 1구명완, 1서강대학교 컴퓨터공학과, 2 KT 융합기술원

4p. 제목: 멀티모달 접근을 통한 딥러닝 기반 감정인식 알고리즘
저자: 김석민, 조원익, 김형주, 김남수

15p. 제목: Detecting the magnitude of depression in Twitter users using sentiment analysis
저자: Jini Jojo Stephen, Prabu P Department of Computer Science, Christ (Deemed University), India

## 모델링에서 참조한 논문

<모델 전반 구조>
제목:Audio-Text Sentiment Analysis using Deep Robust Complementary Fusion of Multi-Features and Multi-Modalities
저자: Feiyang Chen , Ziqian Luo

<Audio model part>
제목: Empirical Interpretation of Speech Emotion Perception with Attention Based Model for Speech Emotion Recognition
저자: Md Asif Jalal, Rosanna Milner, Thomas Hain Speech and Hearing Group (SPandH), The University of Sheffield

# APPENDIX

## 이미지 출처 및 활용출처

3p. 감정인식 시장
이미지 출처https://www.ciokorea.com/news/188928

6p. (이 1은 슬라이드 번호 임의로 해뒀어요) 음성 데이터 전처리
이미지 출처: https://youdaengcom.tistory.com/5
이미지 출처: https://hyunlee103.tistory.com/48

9p. 음성 데이터 구조
이미지 출처: https://medium.com/mlearning-ai/self-attention-in-convolutional-neural-networks-172d947afc00

10p. 음성 데이터 구조
이미지 출처: https://www.researchgate.net/publication/329512919_NLP_at_IEST_2018_BiLSTM-Attention_and_LSTM-Attention_via_Soft_Voting_in_Emotion_Classification/figures?lo=1

11p. KoBERT
이미지 출처: https://wikidocs.net/115055
word piece tokenizer 이미지 출처: https://towardsdatascience.com/how-to-build-a-wordpiece-tokenizer-for-bert-f505d97dddbb

12p. BiLSTM
이미지 출처: https://www.researchgate.net/figure/Attention-mechanism-used-in-our-Bi-LSTM-neural-network_fig3_328054871

17p. 마인드디텍터 활용
출처: https://wansook0316.github.io/cv/projects/2020/12/21/Mind-detector.html