

# 우울증 종합 진단 플랫폼

딥러닝팀

김예찬

윤지영

채소연

한지원

홍지우



# 목차



## 1. 주제 소개

주제 선정, 계획/모델 전개

CON  
TENT

10%

CONTENT

30%

## 2. EDA

2-1 한국어 데이터 EDA

2-2 영어 데이터 EDA

## 3. 데이터 전처리

음성, 텍스트 전처리 유형별 소개

CONTENT

15%

CONTENT

30%

## 4. 감정 분류

4-1 음성 감정 분류

4-2 텍스트 감정 분류

## 5. Multi-Modal 감정 분류 모델

모델 고안, Input Data

CONTENT

15%

Fin.

3주차 계획





## 1. 주제 소개



## 1. 주제 소개



### 딥러닝을 이용한 우울증 진단 모델



마음의 감기 우울증

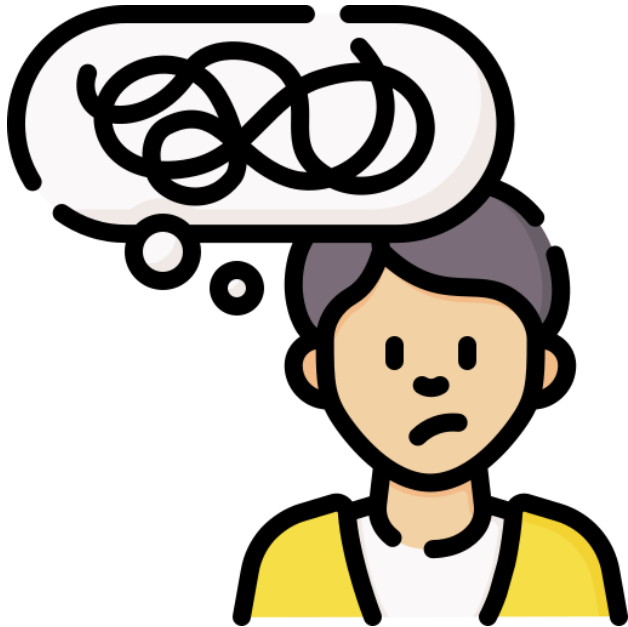




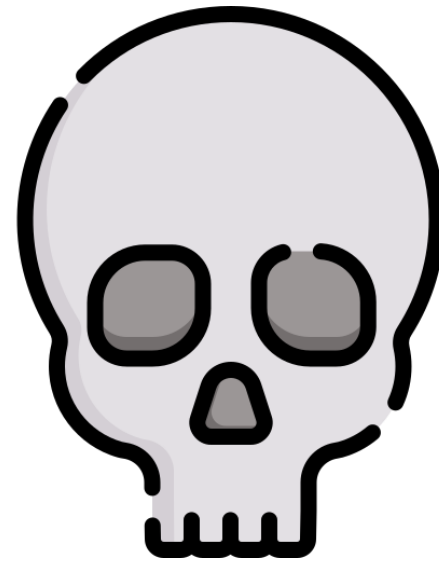
## 1. 주제 소개



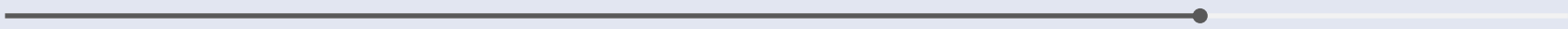
### 딥러닝을 이용한 우울증 진단 모델



마음의 감기 우울증



생명까지 위협





## 1. 주제 소개



### 딥러닝을 이용한 우울증 진단 모델



마음의 감기 우울증

집에서 쉽게 우울증을  
진단할 수 있다면 어떨까?



생명까지 위협





## 1. 주제 소개



# 딥러닝을 이용한 우울증 진단 모델

Original Article | [Open Access](#) | [Published: 27 August 2021](#)

## Deep learning for prediction of depressive symptoms in a large textual dataset

[Md Zia Uddin](#) [Kim Kristoffer Dvsthe](#), [Asbjørn Følstad](#) & [Petter Bae Brandtzaeg](#)

[Neural Computin](#)

Conference Paper

PDF Available

5263 Accesses

## Early Depression Detection from Social Network Using Deep Learning Techniques

June 2020

DOI: [10.1109/TENSYMP50017.2020.9231008](#)

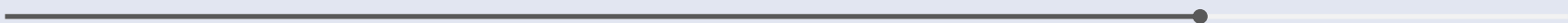
Conference: Tensymp 2020 · At: Dhaka, Bangladesh

Projects: [Text Mining](#) · [Text mining](#)

우울증 진단 모델 - 해외에는 이미 존재

## Deep Learning for Depression Detection of Twitter Users

[Ahmed Hussein Orabi](#), [Prasadith Buddhitha](#), [Mahmoud Hussein Orabi](#), [Diana Inkpen](#)







## 1. 주제 소개



# 딥러닝을 이용한 우울증 진단 모델

Original Article | Open Access | Published: 27 August 2021

Deep learning for prediction of depressive symptoms in a large textual dataset

Md Zia Uddin  Kim Kristoff  et al.

Neural Computir

Conference Paper

PDF Available

5263 Accesses

Early Depression Detection from Social Network Using Deep Learning Techniques

June 2020

DOI: [10.1109/TENSYMP50017.2020.9231008](https://doi.org/10.1109/TENSYMP50017.2020.9231008)

Conference: Tensymp 2020 · At: Dhaka, Bangladesh

Projects: [Text Mining](#) · [Text mining](#)

하지만 한국에는 우울증 라벨링 데이터 없음

우울증 진단 모델 - 해외에는 이미 존재

+

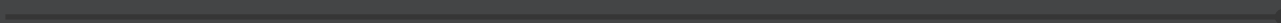
관련된 연구 부족!

## Deep Learning for Depression Detection of Twitter Users

Ahmed Hussein Orabi, Prasadith Buddhitha, Mahmoud Hussein Orabi, Diana Inkpen



Deep Learning  
-PSAT



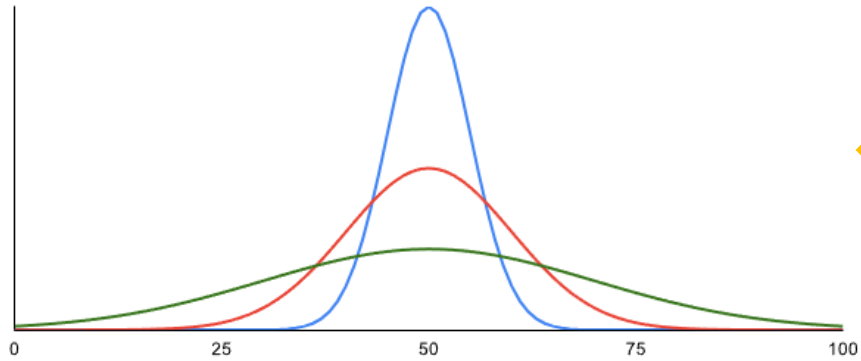




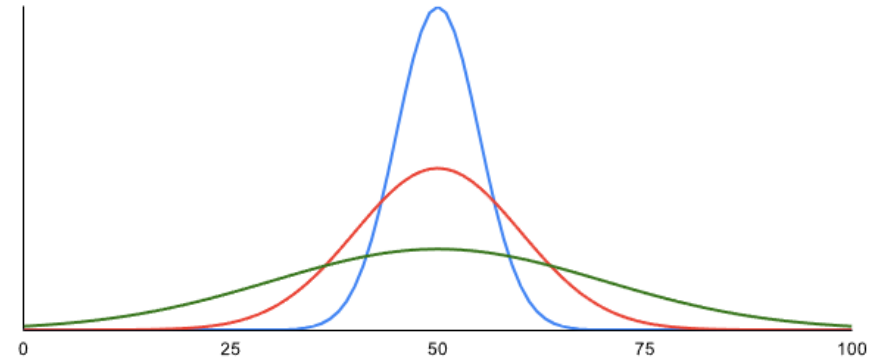
## 1. 주제 소개



### 딥러닝을 이용한 우울증 진단 모델



우울증 진단 여부가 라벨링 된 영어 Corpus



감정 분류가 라벨링 된 영어 Corpus

우울증 환자의 감정 분포 유형화

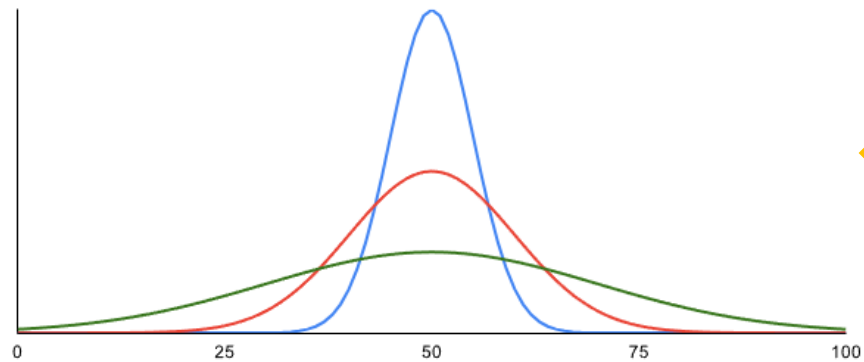




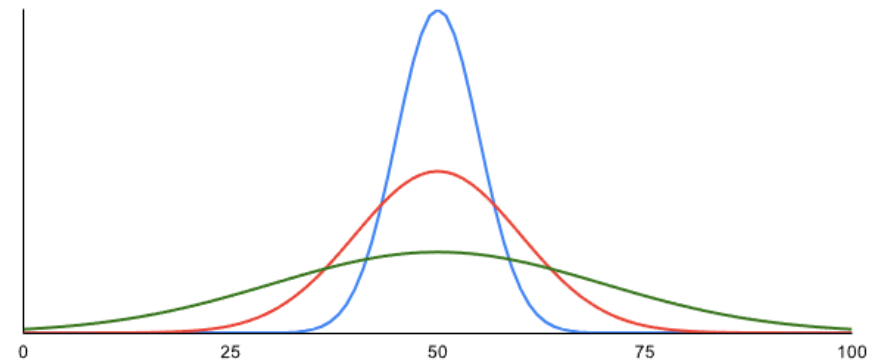
## 1. 주제 소개



### 딥러닝을 이용한 우울증 진단 모델



우울증 진단 여부가 라벨링 된 영어 Corpus



감정 분류가 라벨링 된 영어 Corpus

영어 Corpus 속 우울증 환자들과 한국어 Corpus 속 화자들의 감정 분포 유사도 바탕 라벨링





## 1. 주제 소개



### 딥러닝을 이용한 우울증 진단 모델

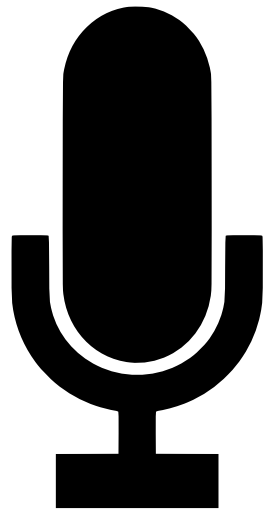




## 1. 주제 소개



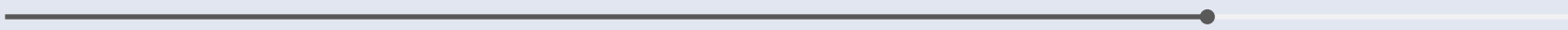
### 딥러닝을 이용한 우울증 진단 모델



음성



텍스트





## 2-1. 한국어 데이터 EDA



## 2-1. 한국어 데이터 EDA

출처 : AIHub 감성 대화 말뭉치 <https://aihub.or.kr/aidata/7978>



### 한국어 데이터 설명

우울증 관련 언어 의미 구조화

+

응답 시나리오 동반 감성 텍스트 언어 수집

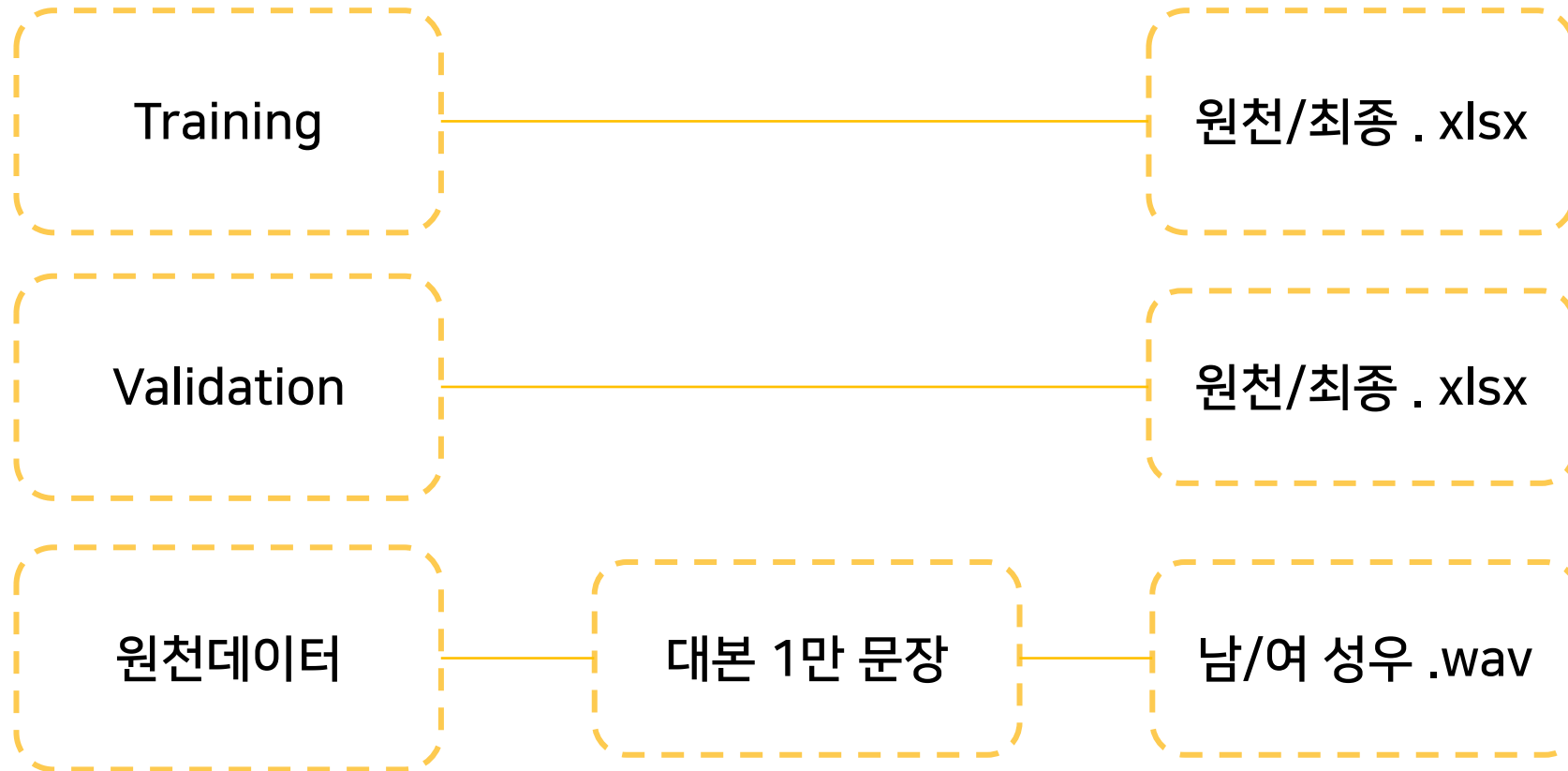
60가지의 세부 감정에 대한  
자연어 처리 말뭉치

클라우드 소싱 수행으로 일반인  
1,500명 대상으로  
음성 10,000 문장 및  
코퍼스 27만 문장 구축





### 데이터 구성





### 데이터 구성



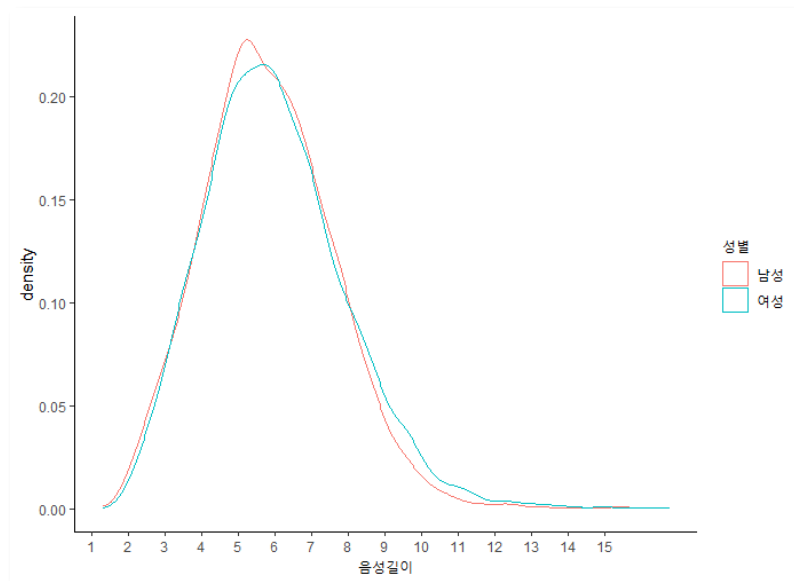




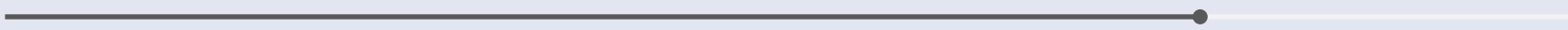
## 2-1. 한국어 데이터 EDA



### 음성 데이터 길이



남/여 모두  
대체로 5초 ~ 6초  
사이인 것을 확인할 수 있음



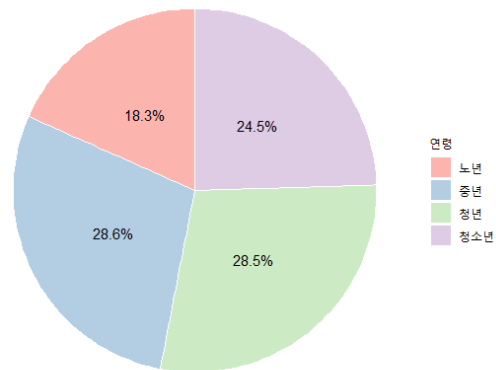


## 2-1. 한국어 데이터 EDA



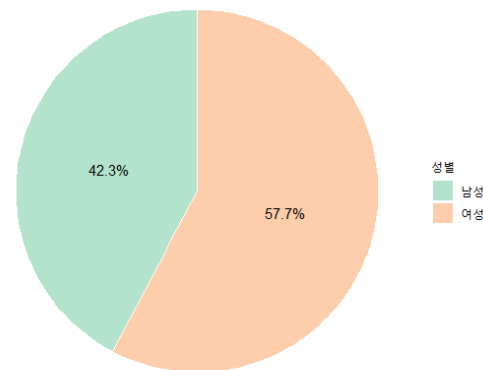
### Training Data

연령



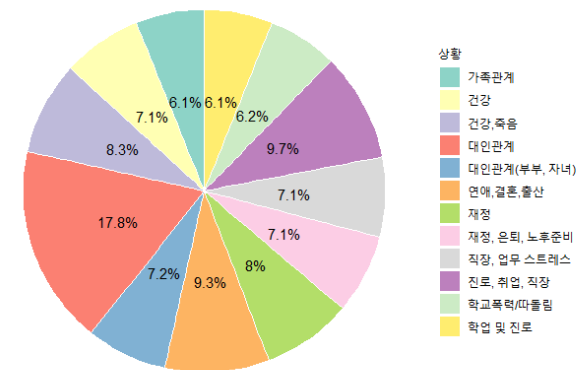
청년이 가장 많고  
노년이 가장 적음

성별



여성이 남성보다  
더 자료가 많음

상황별



대인관계 항목이 가장 자료가 많고  
그 외에는 분포의 차이가 크지 않음



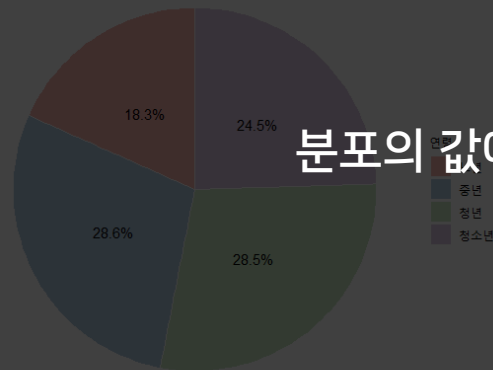


## 2-1. 한국어 데이터 EDA



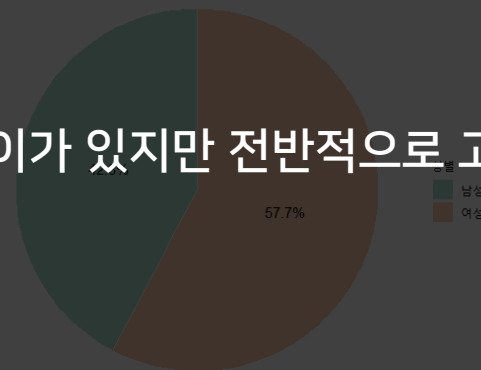
### Training Data

연령



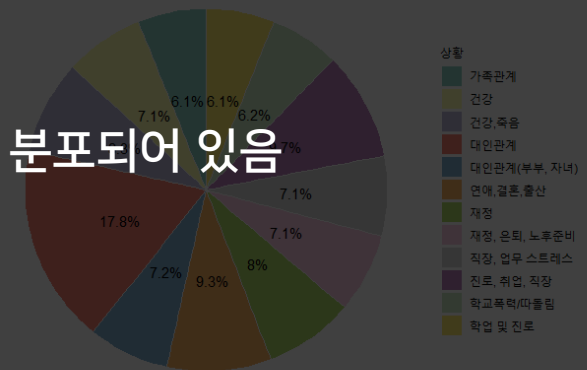
청년이 가장 많고  
노년이 가장 적음

성별



여성이 남성보다  
더 자료가 많음

상황별



대인관계 항목이 가장 자료가 많고  
그 외에는 분포의 차이가 크지 않음

분포의 값에는 차이가 있지만 전반적으로 고르게 분포되어 있음



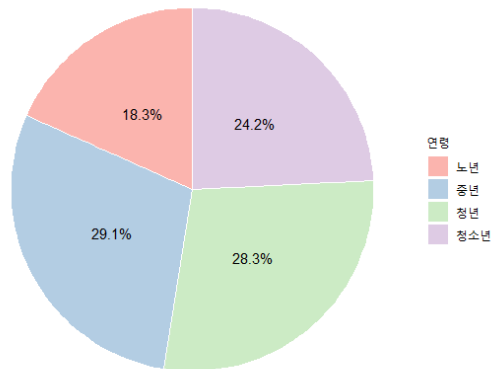


## 2-1. 한국어 데이터 EDA

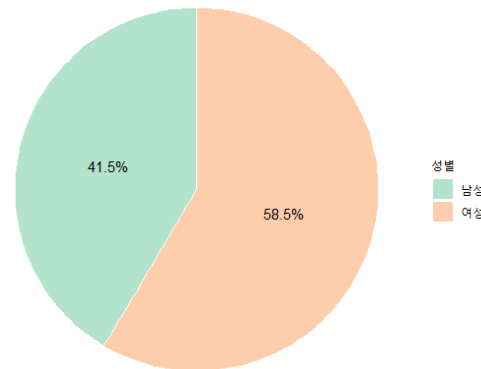


### Validation Data

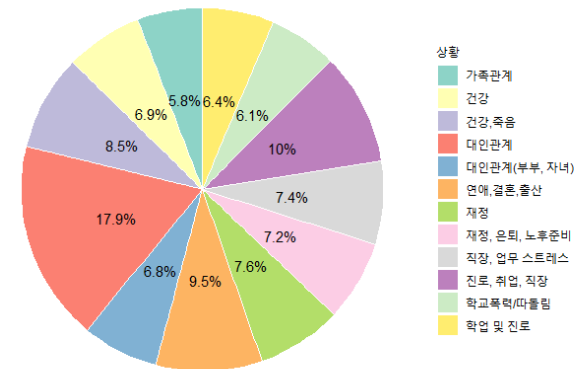
연령



성별



상황별



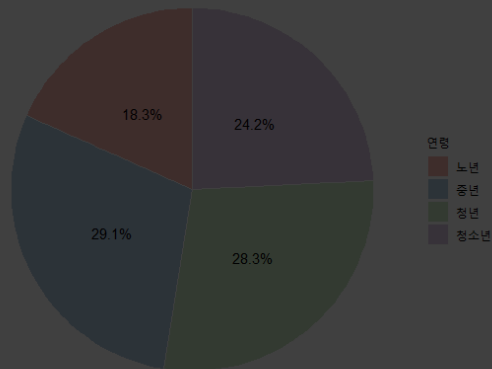


## 2-1. 한국어 데이터 EDA

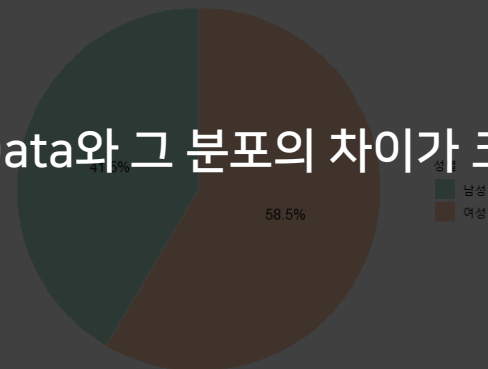


### Validation Data

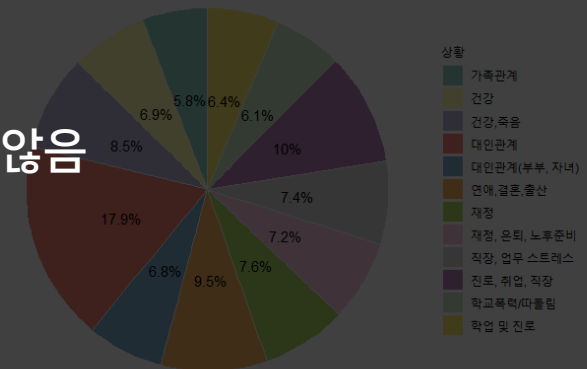
연령



성별



상황별



Train Data와 그 분포의 차이가 크지 않음



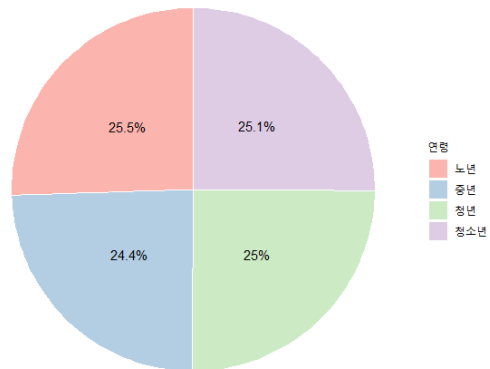


## 2-1. 한국어 데이터 EDA

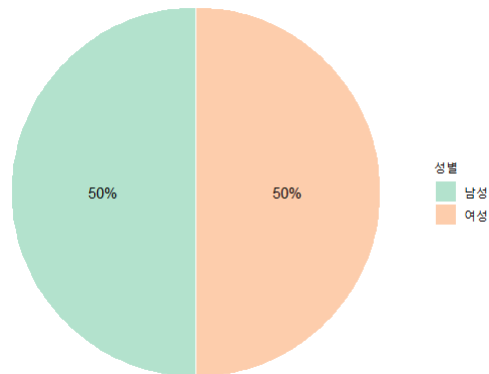


### Audio Data

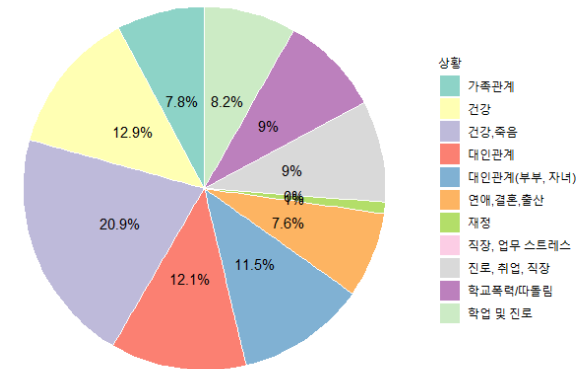
연령



성별



상황별

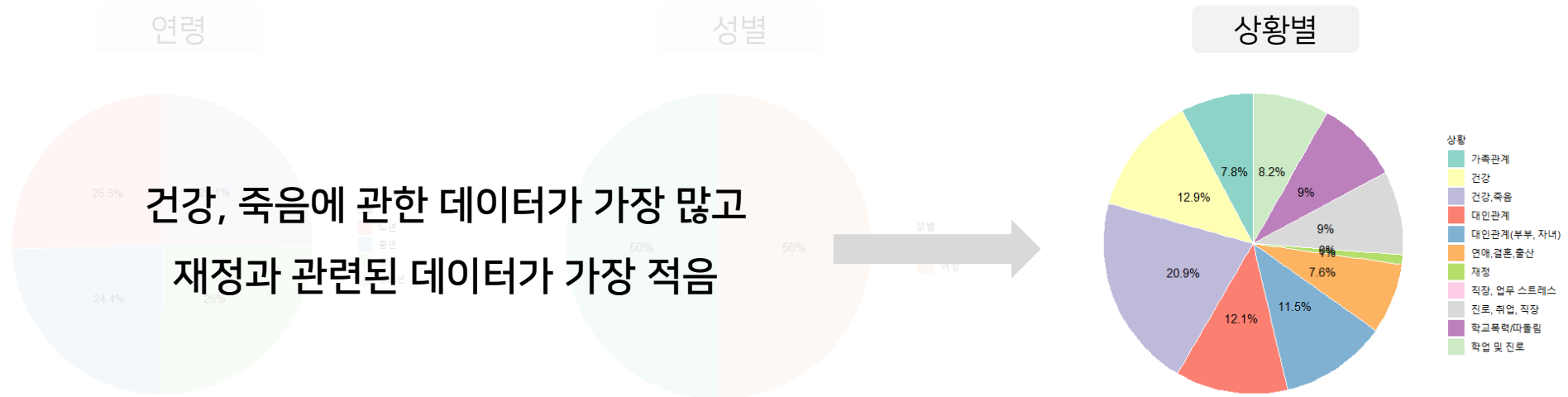




## 2-1. 한국어 데이터 EDA



### Audio Data



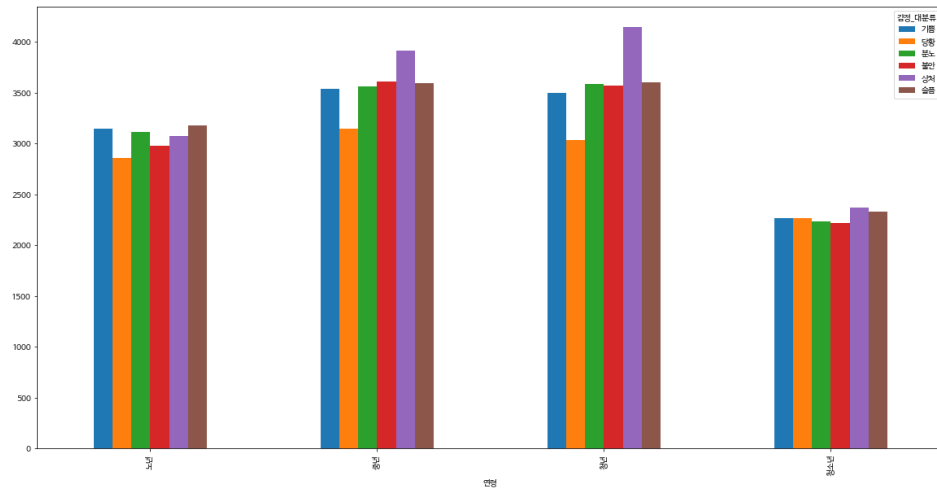


## 2-1. 한국어 데이터 EDA

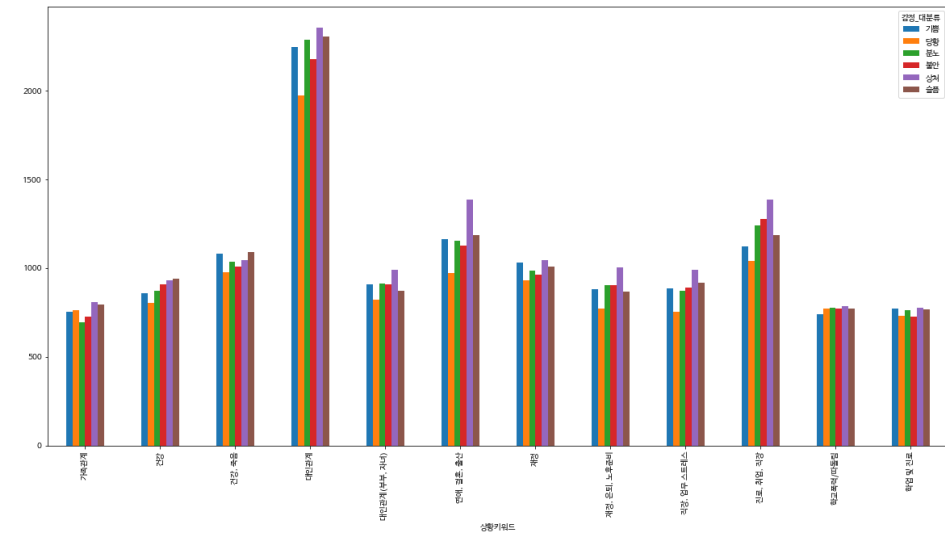


### Train vs 감정\_대분류

연령별



상황분류별





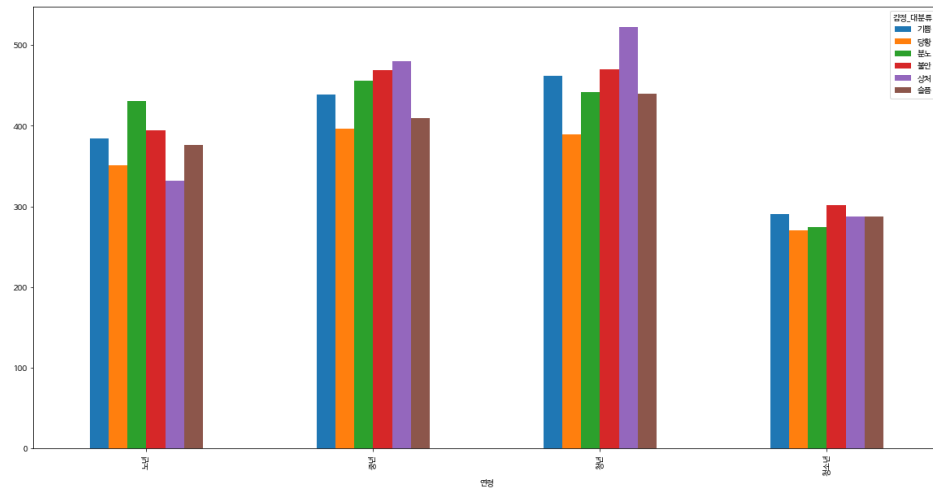


## 2-1. 한국어 데이터 EDA

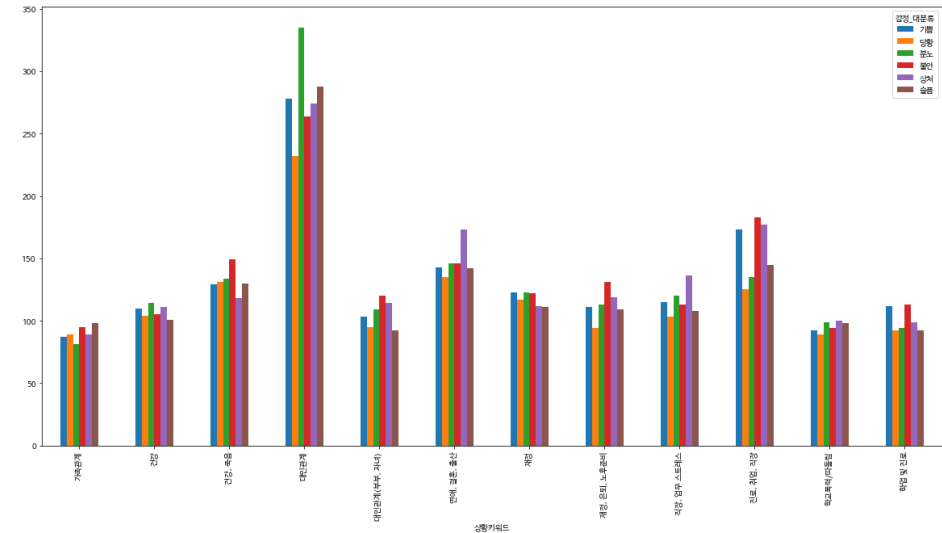


### Valid vs 감정\_대분류

연령별



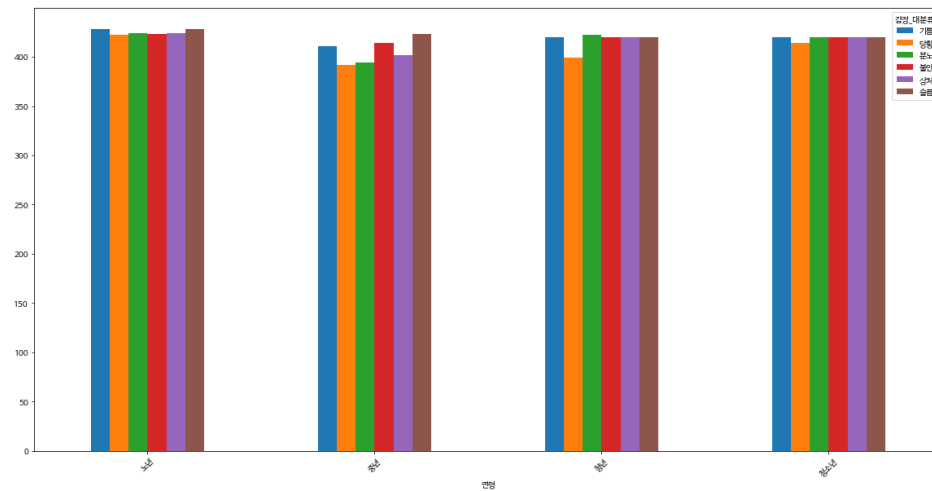
상황분류별



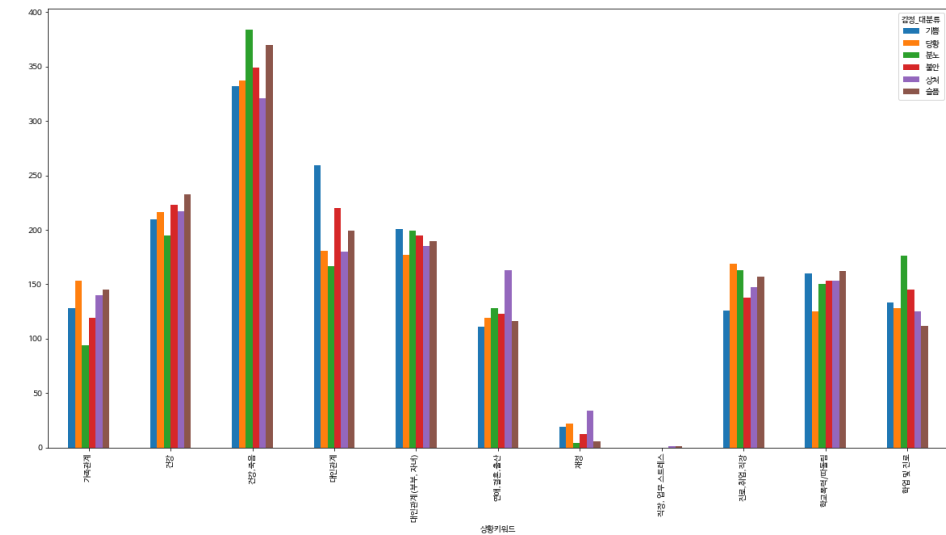


## Audio vs 감정\_대분류

연령별



## 상황분류별





### 한국어 데이터의 특징

- 3가지 데이터 모두 연령, 상황별, 성별에 분포 차이가 크지 않음.
- 따라서 불균형 요소는 고려할 필요 X
- 각 카테고리별 감정의 분포를 확인해보면 그 분포 역시 차이가 크지 않음을 확인할 수 있음.
- 따라서 그 불균형 요소는 고려 X





## 2-2. 영어 데이터 EDA



## 2-2. 영어 데이터 EDA



### 데이터의 구성

출처 : Emotional Voice Conversion:  
Theory, Databases and ESD  
<https://arxiv.org/abs/2105.14762>

어조에 따른  
감정을 나타내는  
데이터

중국 원어민 10명  
영어 원어민  
10명으로 구성

이중 영어 원어민 10명의 데이터 사용





### 데이터의 구성

WAV  
17500개

TXT  
350개

총 350개의 문장을 5가지 감정×10명 = 총 17500개

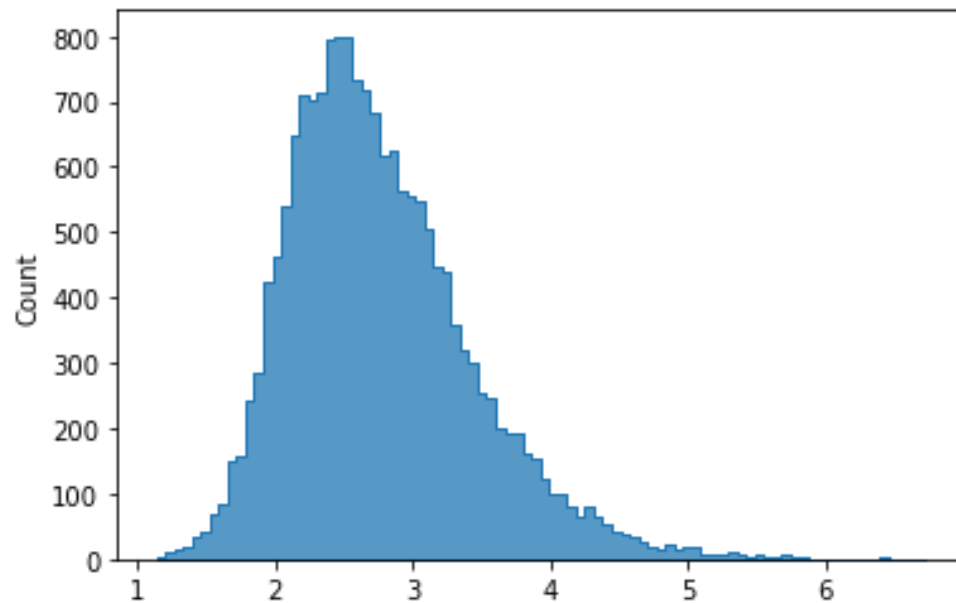




## 2-2. 영어 데이터 EDA



### 음성 데이터 길이



음성 데이터의 길이가 대체로  
2초~3초 정도 사이인 것을  
확인할 수 있음

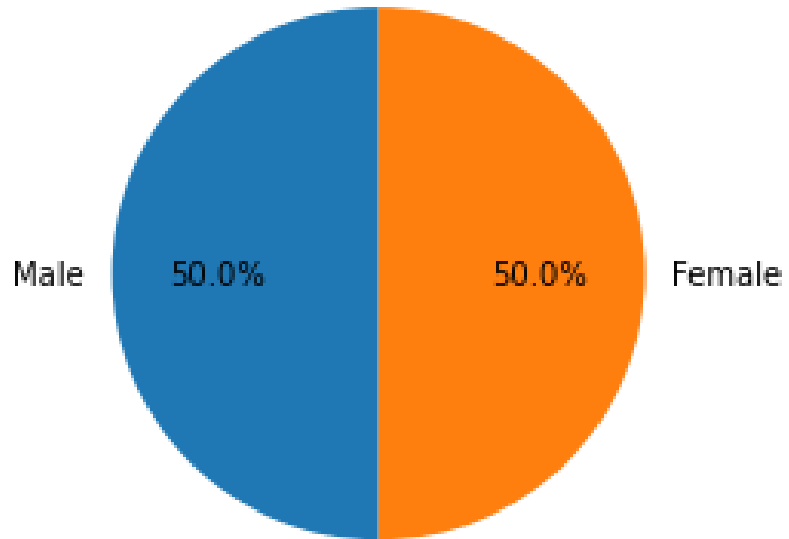




## 2-2. 영어 데이터 EDA



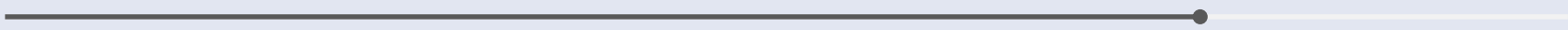
### 성별 분포



정제된 데이터로

남자와 여자의 비율이 정확히

50:50인 것을 확인할 수 있음



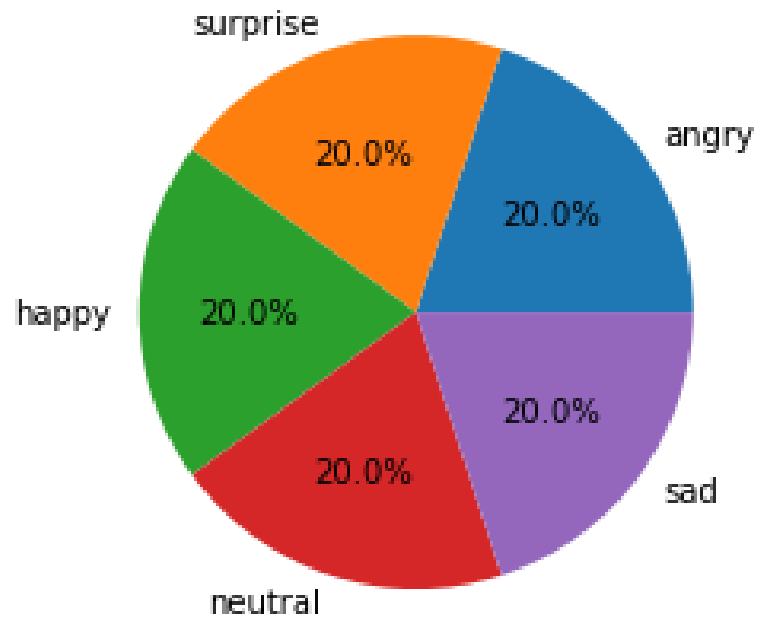




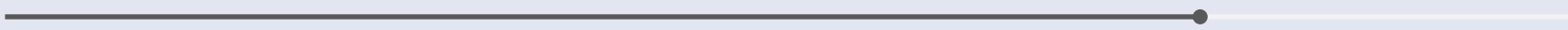
## 2-2. 영어 데이터 EDA



### 감정 분포



마찬가지로  
각 감정의 분포가 정확히  
20퍼센트인 것을  
확인할 수 있음

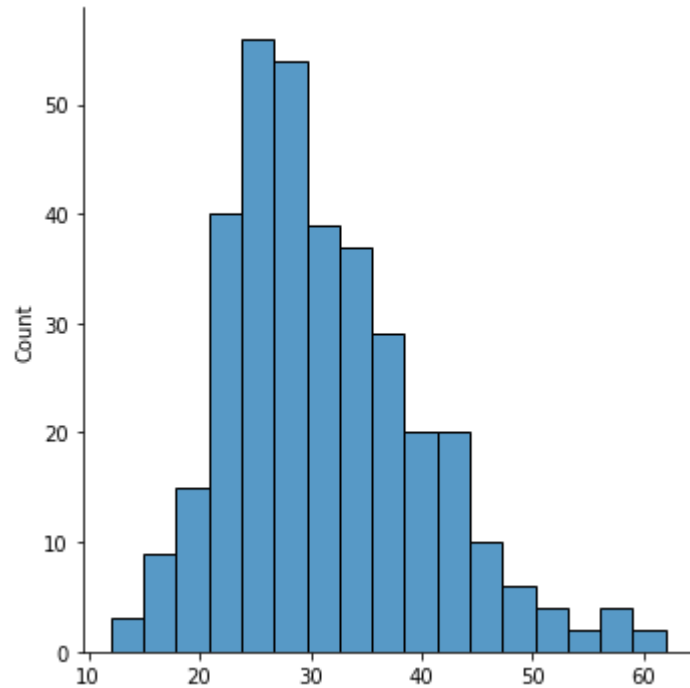




## 2-2. 영어 데이터 EDA



### 텍스트 데이터 길이



텍스트의 길이를 전처리한 결과  
대략 30개의 알파벳으로 구성되어  
있는 것을 확인할 수 있음





### 영어데이터의 특징

#### 장점

한국어 데이터와는 다르게  
매우 정제되어 있는 데이터



성별이나 감정에 대한  
불균형 요소는 고려할 필요 X





### 영어데이터의 특징

#### 장점

한국어 데이터와는 다르게  
매우 정제되어 있는 데이터



성별이나 감정에 대한  
불균형 요소는 고려할 필요 X

#### 단점

상황이나 연령에 대한 요소들  
전혀 포함되어 있지 않음



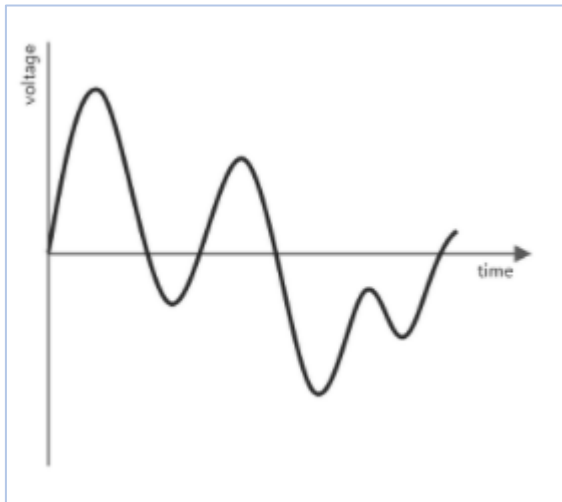
### 3. 데이터 전처리



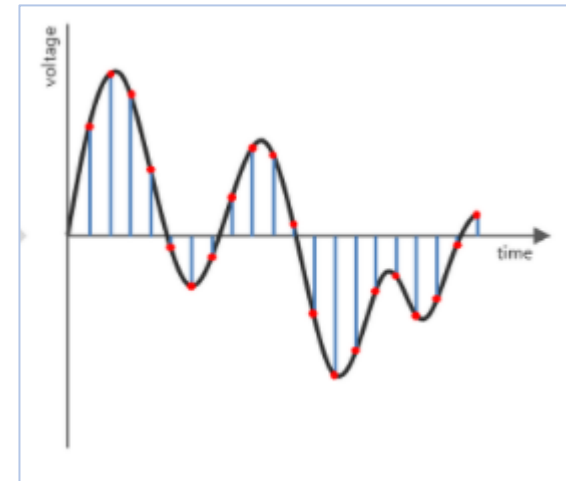
### 3. 데이터 전처리



## 음성 전처리



디지털화



연속적인 음성을 특정 주기당 Sampling을 통해 디지털화





### 3. 데이터 전처리



## 음성 전처리 - 라이브러리

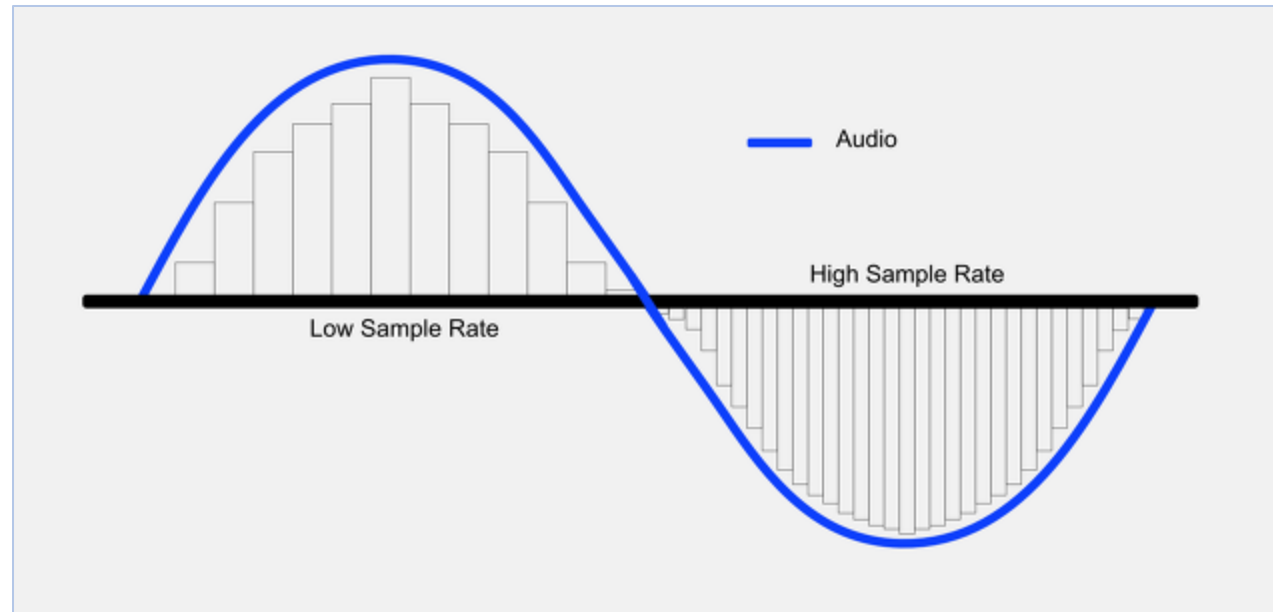




### 3. 데이터 전처리



## 음성 전처리 - Sampling Rate



Sampling Rate : 초당 Sampling 횟수 (보통 44100)



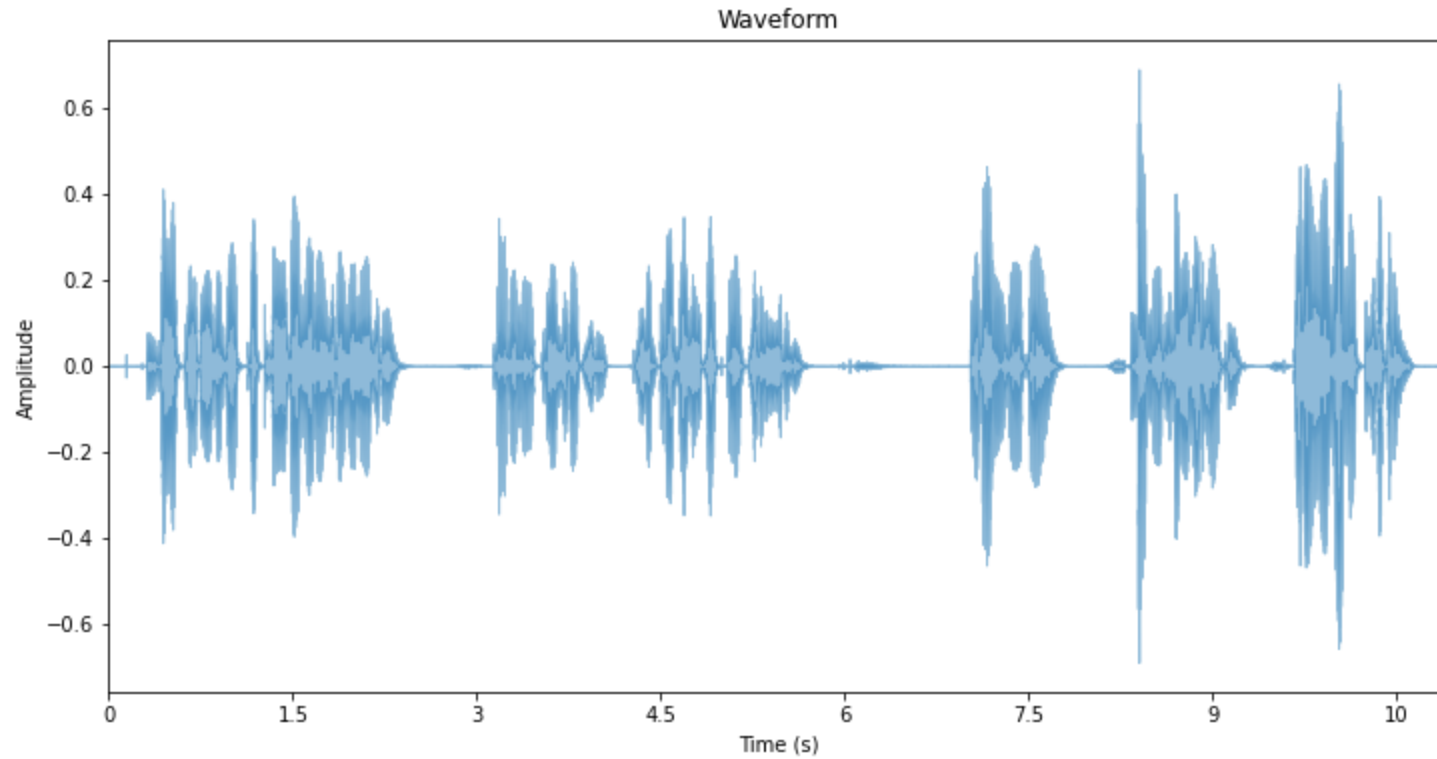




### 3. 데이터 전처리



## 음성 전처리 - Waveform



시간대별 진폭을 2차원 그래프의 형태로 출력

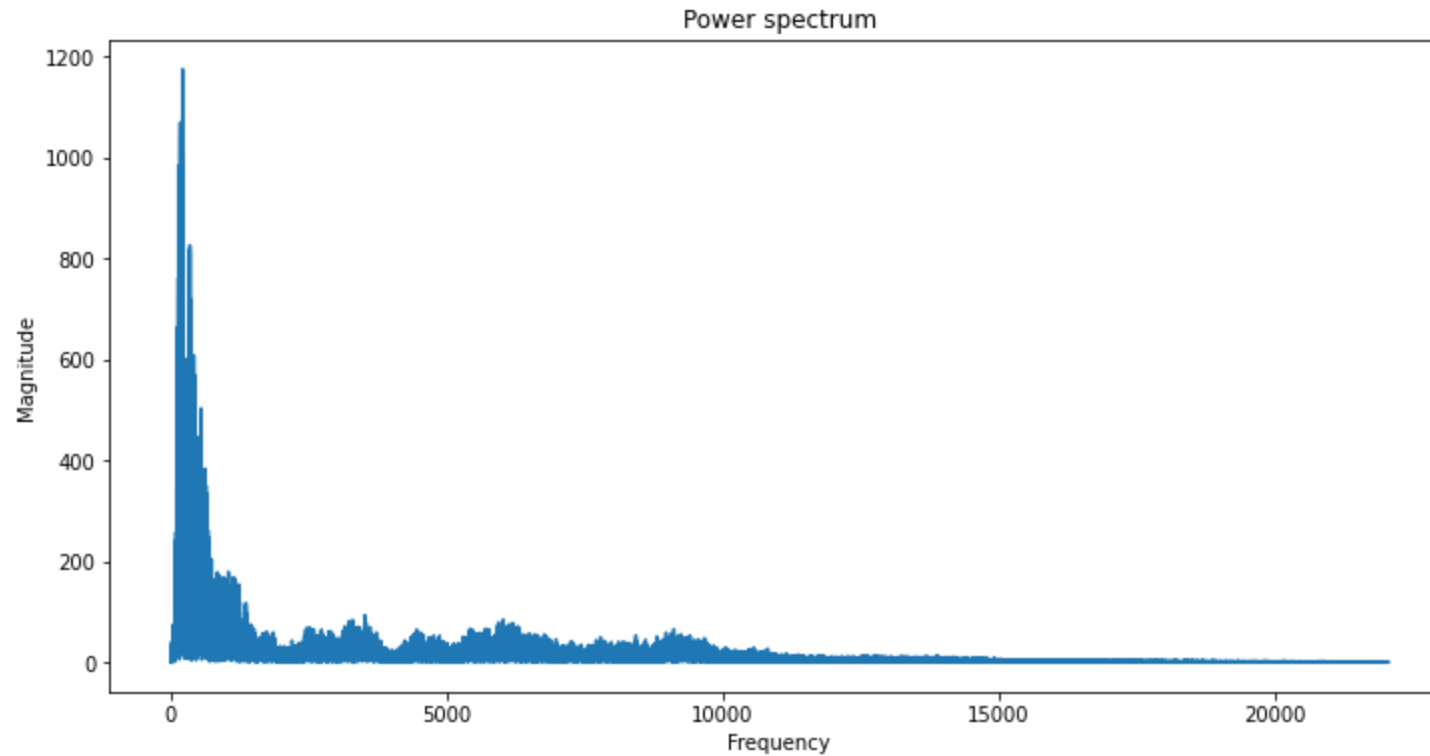




### 3. 데이터 전처리



## 음성 전처리 - Spectrum



Waveform

푸리에 변환

주기함수의 합으로 분해

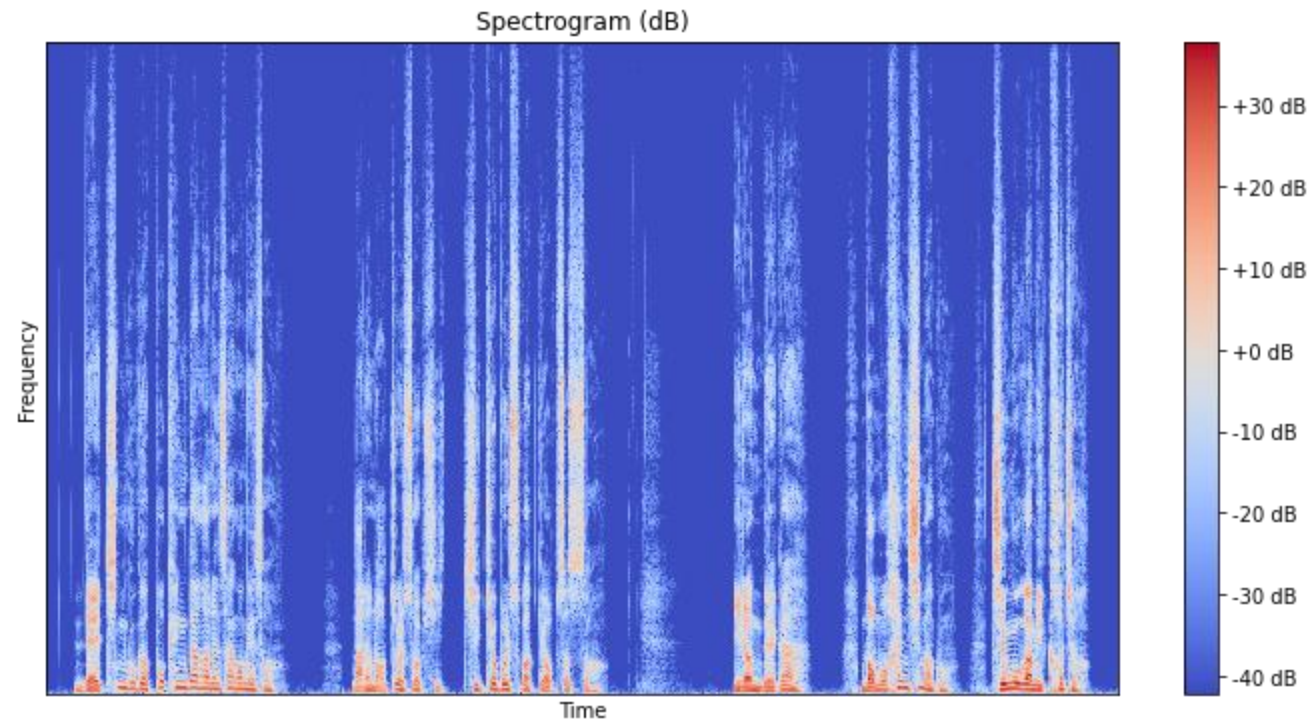




### 3. 데이터 전처리



## 음성 전처리 - Spectrogram



Waveform

STFT  
(Short-Time Fourier Transformation)

일정 간격(Frame)마다 푸리에 변환 적용

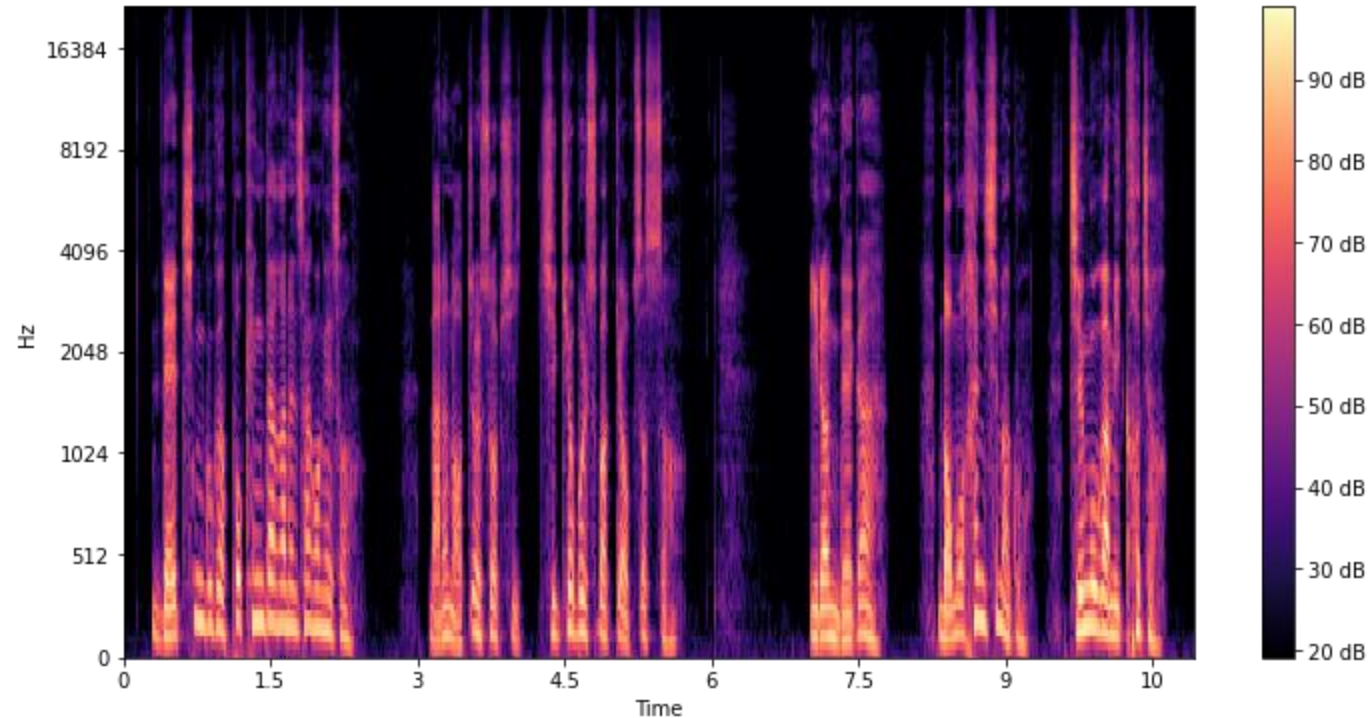




### 3. 데이터 전처리



## 음성 전처리 - Mel Spectrogram



Spectrogram

Mel Scale 적용  
(Log Scale)

사람 귀의 민감도를 반영하여 저주파수 강조

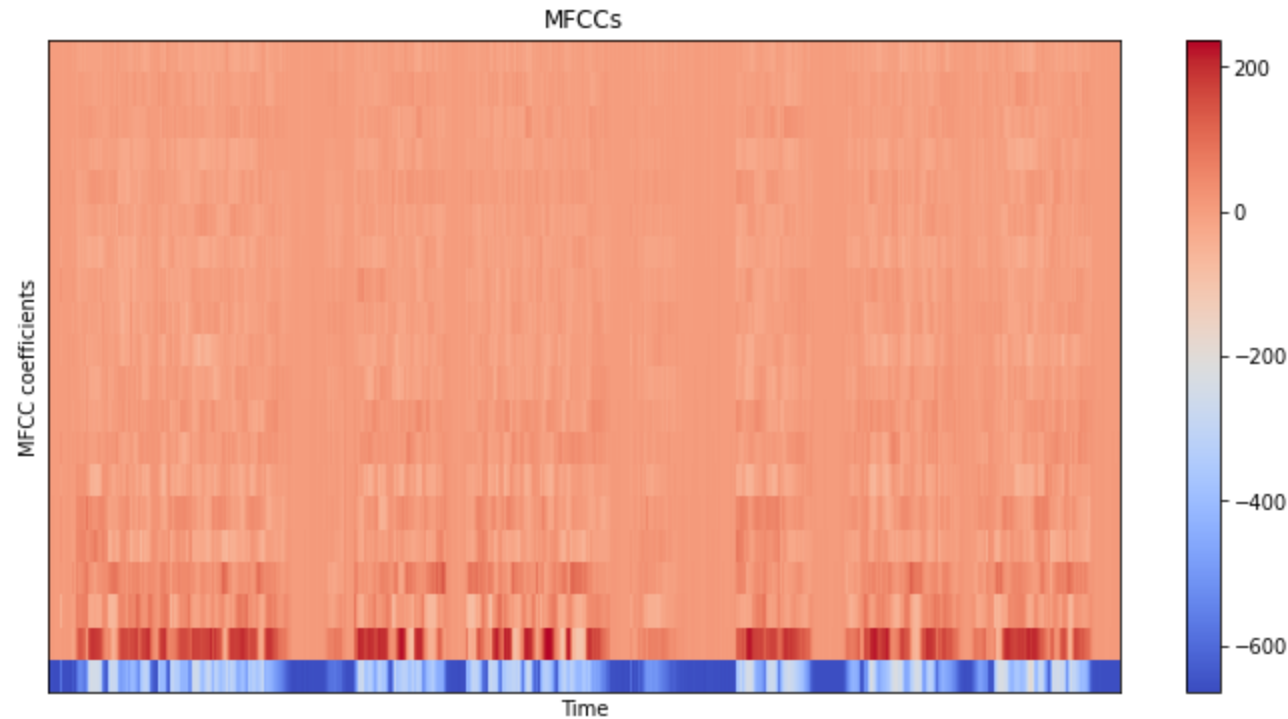




### 3. 데이터 전처리



## 음성 전처리 - MFCC (Mel Frequency Cepstral Coefficient)



Spectrum

Mel Scale 적용  
(Log Scale)

Cepstral Analysis

MFCC



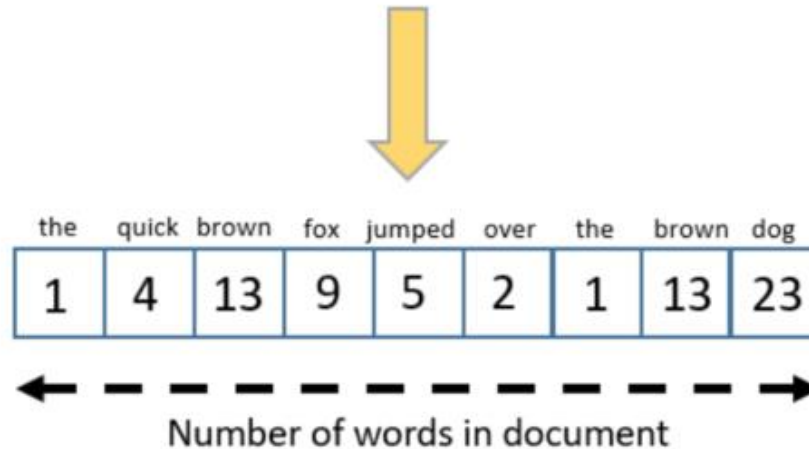


### 3. 데이터 전처리



# 텍스트 전처리 - Index Based Encoding

The quick brown fox jumped over the brown dog



전체 데이터셋의 단어들에 번호를 부여하여 각 단어를 정수로 인코딩



### 3. 데이터 전처리



## 텍스트 전처리 - One Hot Encoding

Rome = [1, 0, 0, 0, 0, 0, ..., 0]

Paris = [0, 1, 0, 0, 0, 0, ..., 0]

Italy = [0, 0, 1, 0, 0, 0, ..., 0]

France = [0, 0, 0, 1, 0, 0, ..., 0]

word V

전체 데이터셋의 단어 가짓수를 차원으로 One-Hot Encoding

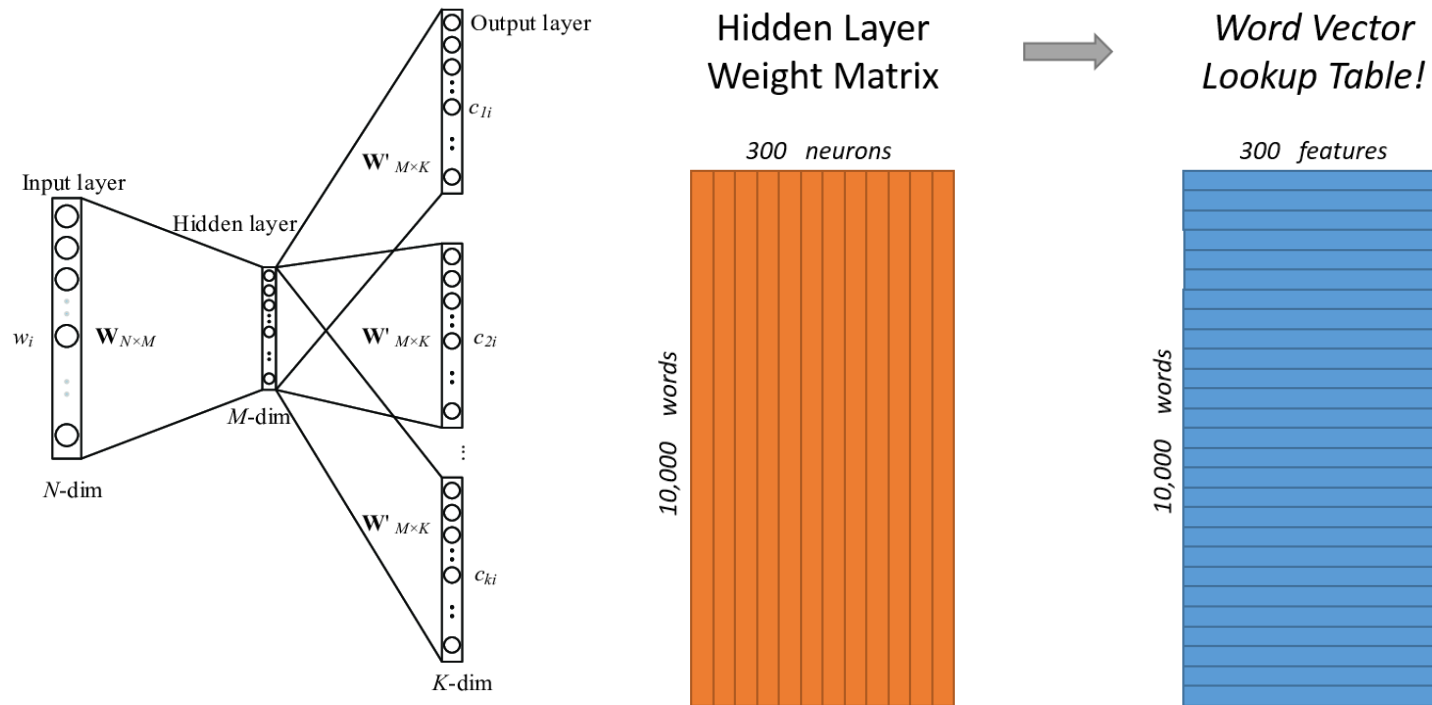




### 3. 데이터 전처리



## 텍스트 전처리 - Word2Vec



One-Hot Encoding 된 단어 벡터들을 저차원 벡터로 변환







## 4-1. 음성 감정 분류



## 4-1. 음성 감정 분류



### 음성 데이터 - csv 구조

NO.	성우	연령	성별	상황키워드	감정_대분류	감정_소분류	사람문장
M_000001	LSW	청소년	남성	학업 및 진로	기쁨	감사하는	오늘 반장 선거에서 내가 반장이 됐어! 친구들이 날 믿어줘서 너무 고마워
F_000021	JSY	청소년	여성	가족관계	당황	고립된	내 부모님은 내가 친구들과 어울려 다니는 걸 마음에 안 들어 하셔
F_000023	KJY	청년	여성	대인관계	불안	스트레스 받는	힘들게 가고 싶던 대학교에 왔지만 자퇴하고 싶어졌어
M_000040	PSM	청년	남성	연애, 결혼, 출산	상처	질투하는	그 애가 다른 사람과 연애를 시작했어. 날 좋아한다고 했던 그 애가.
:							
F_005000	JYC	중년	여성	건강	상처	가난한, 불우한	큰 병에 걸렸는데 병원비가 감당이 안돼
M_005000	KSG	노년	남성	재정	기쁨	감사하는	이번 달도 월세가 잘 입금됐네

약 1만 문장으로 구성






## 4-1. 음성 감정 분류



### 음성 데이터 - csv 구조

NO.	성우	연령	성별	상황키워드	감정_대분류	감정_소분류	사람문장
M_000001	LSW	청소년	남성	학업 및 진로	기쁨	감사하는	오늘 반장 선거에서 내가 반장이 됐어! 친구들이 날 믿어줘서 너무 고마워
F_000021	JSY	청소년	여성	가족관계	당황	고립된	내 부모님은 내가 친구들과 어울려 다니는 걸 마음에 안 들어 하셔
F_000022	JSY	청소년	여성	학업 및 진로	기쁨	감사하는	오늘 반장 선거에서 내가 반장이 됐어! 친구들이 날 믿어줘서 너무 고마워
M_000040	PSM	청년	남성	연애, 결혼, 출산	상처	질투하는	그 애가 다른 사람과 연애를 시작했어. 날 좋아한다고 했던 그 애가.
F_005000	JYC	중년	여성	건강	상처	가난한, 불우한	큰 병에 걸렸는데 병원비가 감당이 안돼
M_005000	KSG	노년	남성	재정	기쁨	감사하는	이번 달도 월세가 잘 입금됐네


 감성대화말뭉치AI데이터\_Wave\_남자성우\_5000 : [M\_000001.wav ~ M\_005000.wav]


 감성대화말뭉치AI데이터\_Wave\_여자성우\_5000 : [F\_000001.wav ~ F\_005000.wav]

약 1만 문장으로 구성





## 4-1. 음성 감정 분류



### Data input

```
def extract_feature(file_name, mfcc, chroma, mel):
    with soundfile.SoundFile(file_name) as sound_file:
        X = sound_file.read(dtype="float32")
        sample_rate=sound_file.samplerate
        if chroma:
            stft=np.abs(librosa.stft(X))
            result=np.array([])
            if mfcc:
                mfccs=np.mean(librosa.feature.mfcc(y=X, sr=sample_rate, n_mfcc=40).T, axis=0)
                result=np.hstack((result, mfccs))
            if chroma:
                chroma=np.mean(librosa.feature.chroma_stft(S=stft, sr=sample_rate).T,axis=0)
                result=np.hstack((result, chroma))
            if mel:
                mel=np.mean(librosa.feature.melspectrogram(X, sr=sample_rate).T,axis=0)
                result=np.hstack((result, mel))
        return result

emotions={
    '01':'neutral',
    '02':'calm',
    '03':'happy',
    '04':'sad',
    '05':'angry',
    '06':'fearful',
    '07':'disgust',
    '08':'surprised'
}

#DataFlair - Emotions to observe
observed_emotions=['happy','surprised','angry','fearful','disgust','sad']

data = read_csv('D:/users/채소연/바탕 화면/Text_total.csv',encoding='cp949')
m = data['감정_대분류']
```

```
def load_data(test_size=0.2):
    x,y=[],[]
    i=0
    for file in glob.glob(
        "D:/users/채소연/바탕 화면/감성대화말뭉치(원천데이터)_음성데이터/감성대화말뭉치AI데이터_Wave_*/*.wav"
    ):
        file_name=os.path.basename(file)
        emotion=m[i]
        if emotion not in observed_emotions:
            continue
        feature=extract_feature(file, mfcc=True, chroma=True, mel=False)
        x.append(feature)
        y.append(emotion)
        print(file_name)
        i=i+1

    return train_test_split(np.array(x), y, test_size=test_size, random_state=9)
```

음성 데이터와 감정 분류 결과를 비교할 CSV 데이터 INPUT  
MFCC와 Mel Spectrogram을 활용해서 전처리





## 4-1. 음성 감정 분류



### Data input

```
#DataFlair - Split the dataset
x_train,x_test,y_train,y_test=load_data(test_size=0.25)

#DataFlair - Get the shape of the training and testing datasets
print((x_train.shape, x_test.shape))

#DataFlair - Get the number of features extracted
print(f'Features extracted: {x_train.shape[1]}')

#DataFlair - Initialize the Multi Layer Perceptron Classifier
model=MLPClassifier(alpha=0.01, batch_size=256, epsilon=1e-08, hidden_layer_sizes=(300,), learning_rate='adaptive', max_iter=1500,early_stopping=False, warm_start=True)

#DataFlair - Train the model
model.fit(x_train,y_train)

import matplotlib.pyplot as plt

from sklearn.neural_network import MLPClassifier
from sklearn.preprocessing import MinMaxScaler
from sklearn import datasets
from sklearn.exceptions import ConvergenceWarning

#DataFlair - Predict for the test set
y_pred=model.predict(x_test)

#DataFlair - Calculate the accuracy of our model
accuracy=accuracy_score(y_true=y_test, y_pred=y_pred)

#DataFlair - Print the accuracy
print("Accuracy: {:.2f}%".format(accuracy*100))
```

MLP Classifier를 활용해서 모델 학습



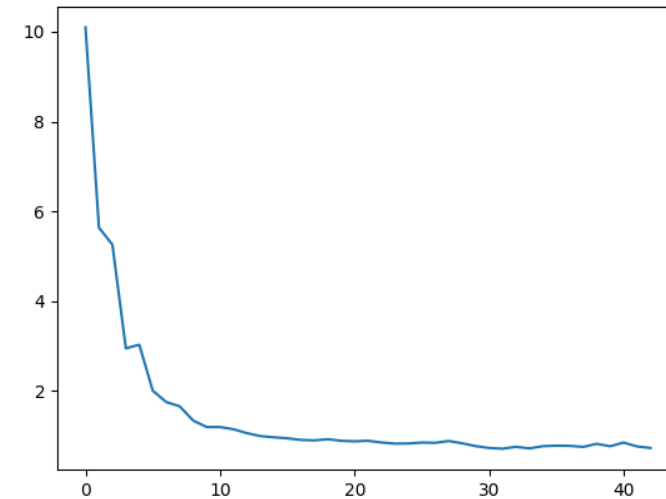
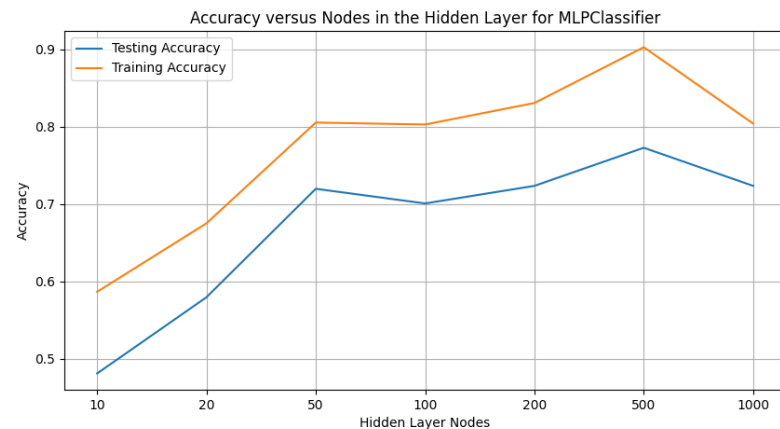


## 4-1. 음성 감정 분류



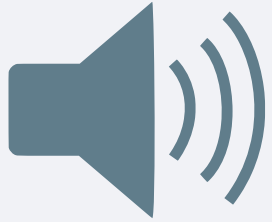
### Accuracy/Loss plot

Accuracy: 69.47%  
((7279, 180), (2427, 180),  
Features extracted: 180  
Accuracy: 69.47%



최종적으로 수렴하는 것을 확인할 수 있음





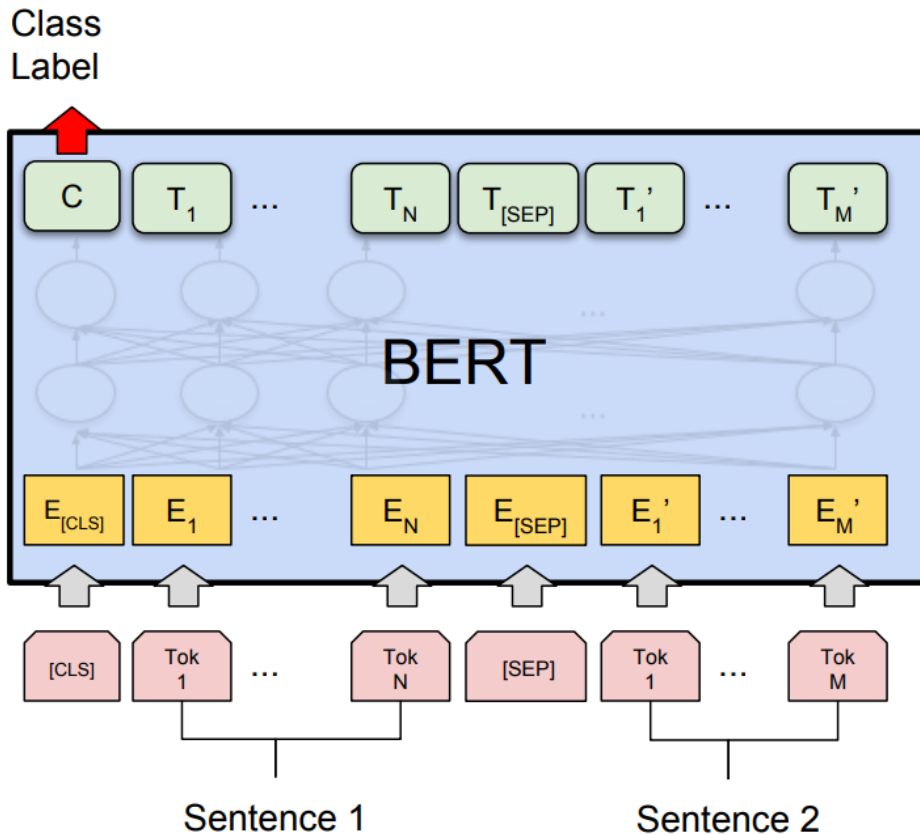
## 4-2. 텍스트 감정분류



## 4-2. 텍스트 감정분류



### BERT란?



- 사전 훈련 언어 모델
- 사전 훈련 임베딩을 통해 성능 높임
- 그 외 임베딩 방식
  - Word2Vec
  - GloVe
  - Fasttext







## 4-2. 텍스트 감정분류



### 일반 모델링 과정





## 4-2. 텍스트 감정분류



### BERT를 사용한 모델링 과정

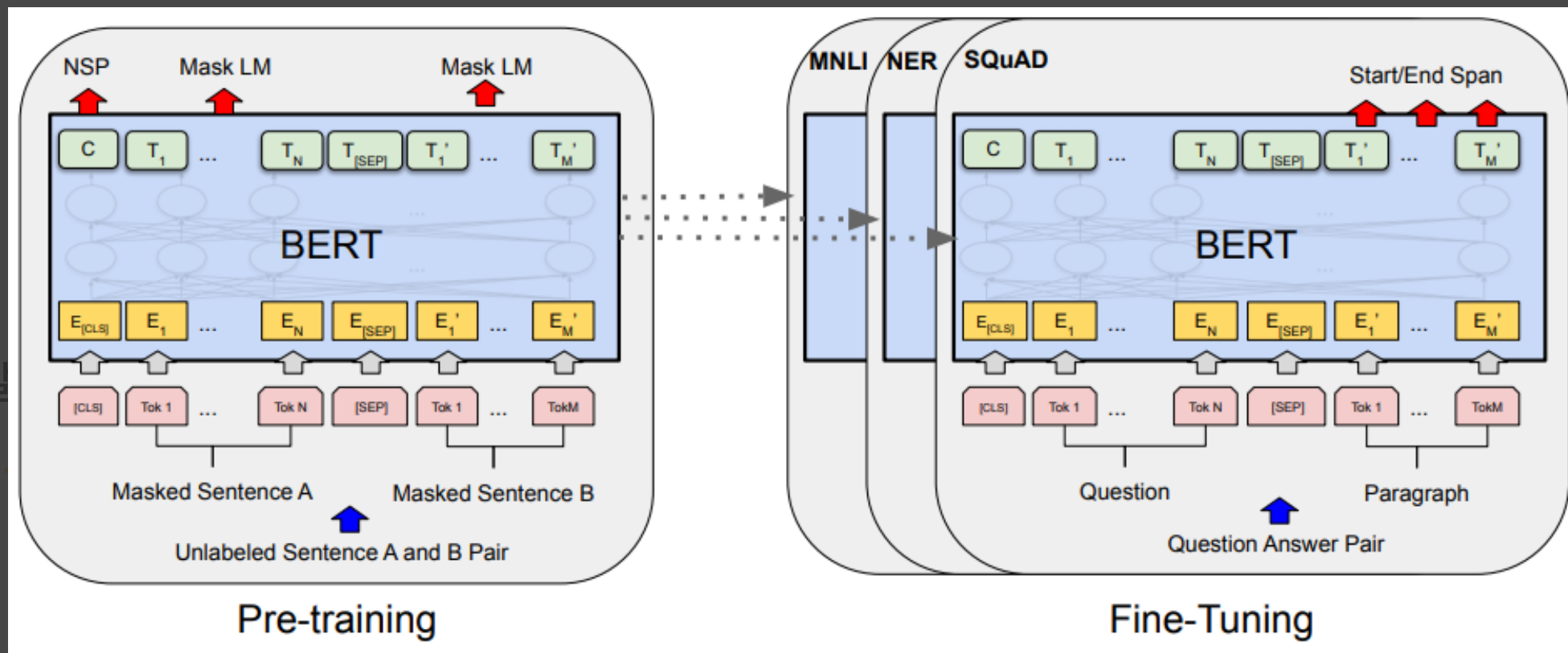




## 4-2. 텍스트 감정분류



### BERT를 사용한 모델링 과정

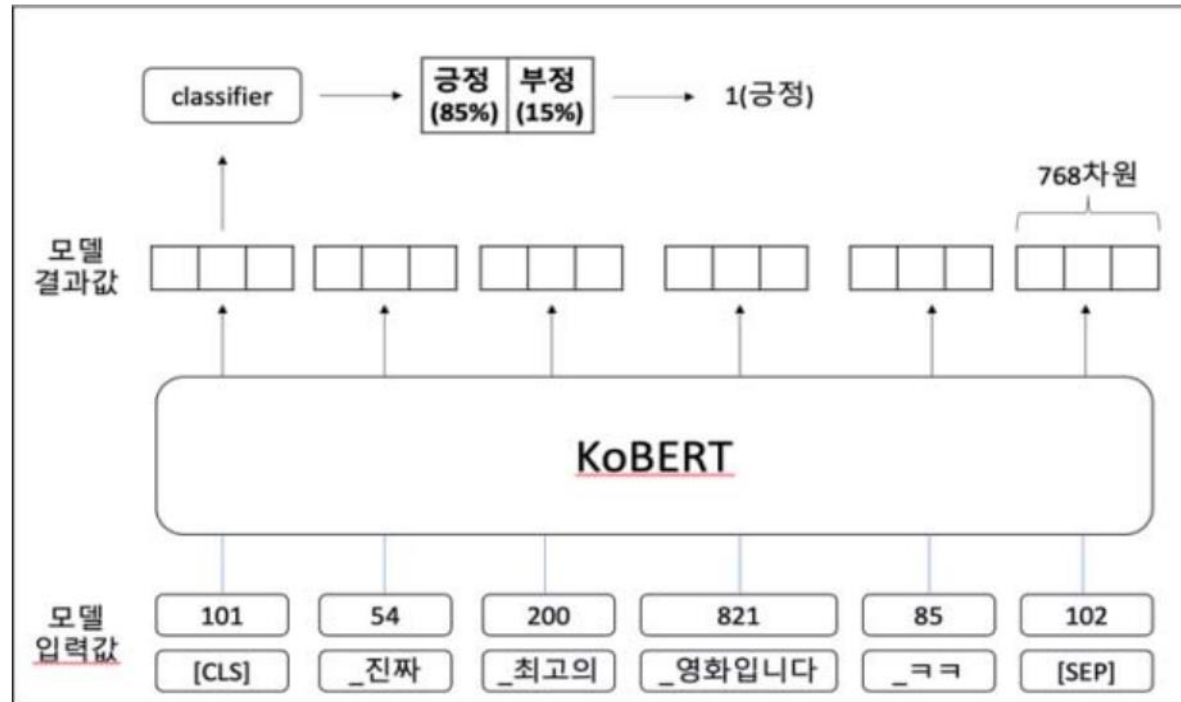




## 4-2. 텍스트 감정분류



### KoBERT란?



- SKTBrain에서 개발
- BERT 구조 활용
- 한국어 위키 5백만 문장 학습
- 한국어 뉴스 2천만 문장 학습
- BERT의 다국어 모델보다 성능 높음





## 4-2. 텍스트 감정분류



### 텍스트 데이터 - CSV 구조

NO.	성우	연령	성별	상황키워드	감정_대분류	감정_소분류	사람문장
M_000001	LSW	청소년	남성	학업 및 진로	기쁨	감사하는	오늘 반장 선거에서 내가 반장이 됐어! 친구들이 날 믿어줘서 너무 고마워
F_000021	JSY	청소년	여성	가족관계	당황	고립된	내 부모님은 내가 친구들과 어울려 다니는 걸 마음에 안 들어 해서
F_000023	KJY	청년	여성	대인관계	불안	스트레스 받는	힘들게 가고 싶던 대학교에 왔지만 자퇴하고 싶어졌어
M_000040	PSM	청년	남성	연애, 결혼, 출산	상처	질투하는	그 애가 다른 사람과 연애를 시작했어. 날 좋아한다고 했던 그 애가.
:							
F_005000	JYC	중년	여성	건강	상처	가난한, 불우한	큰 병에 걸렸는데 병원비가 감당이 안돼
M_005000	KSG	노년	남성	재정	기쁨	감사하는	이번 달도 월세가 잘 입금됐네

약 1만 문장





## 4-2. 텍스트 감정분류



### Input 데이터

감정_대분류	사람문장
기쁨	오늘 반장 선거에서 내가 반장이 됐어! 친구들이 날 믿어줘서 너무 고마워
당황	내 부모님은 내가 친구들과 어울려 다니는 걸 마음에 안 들어 하셔서
불안	힘들게 가고 싶던 대학교에 왔지만 자퇴하고 싶어졌어
상처	그 애가 다른 사람과 연애를 시작했어. 날 좋아한다고 했던 그 애가.
...	
상처	큰 병에 걸렸는데 병원비가 감당이 안돼
기쁨	이번 달도 월세가 잘 입금됐네

감정\_대분류가 라벨링 된 List 데이터





## 4-2. 텍스트 감정분류



### 토큰화 진행 (BERT)

[CLS]	_이	_영화	_꽤	_재밌어	_ㅋㅋ	[SEP]
-------	----	-----	----	------	-----	-------



Embedding 진행

101	54	200	821	85	94	102
-----	----	-----	-----	----	----	-----

감정\_대분류가 라벨링 된 List 데이터





## 4-2. 텍스트 감정분류



### 모델링 진행

101	54	200	821	85	94	102
-----	----	-----	-----	----	----	-----



KoBERT Classifier



'기쁨'으로 분류

기쁨(86%)	불안(3%)	당황(3%)	분노(3%)	슬픔(3%)	상처(2%)
---------	--------	--------	--------	--------	--------





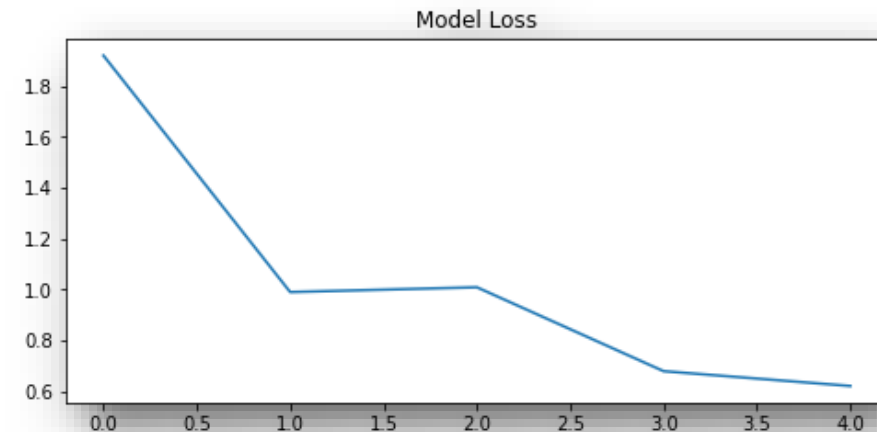
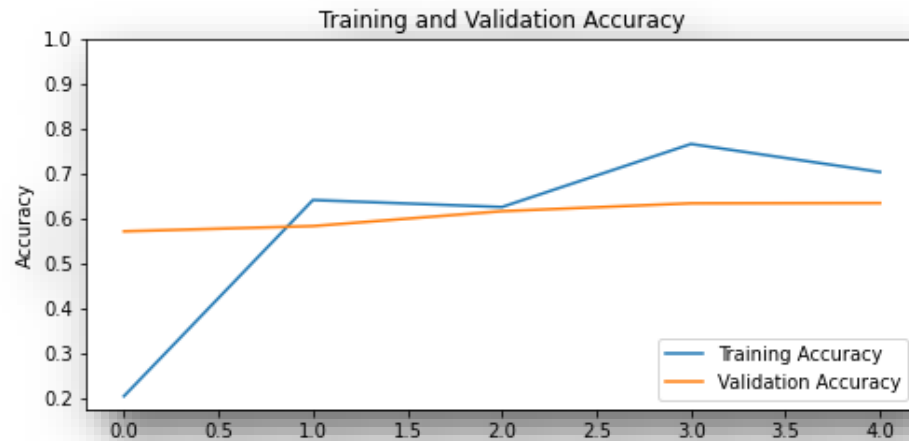


## 4-2. 텍스트 감정분류



### Accuracy plot/Loss plot

train accuracy: 0.809, validation accuracy: 0.635



Loss는 최종적으로 수렴하나, 정확도는 epoch이 증가할 수록 과적합 현상을 보임





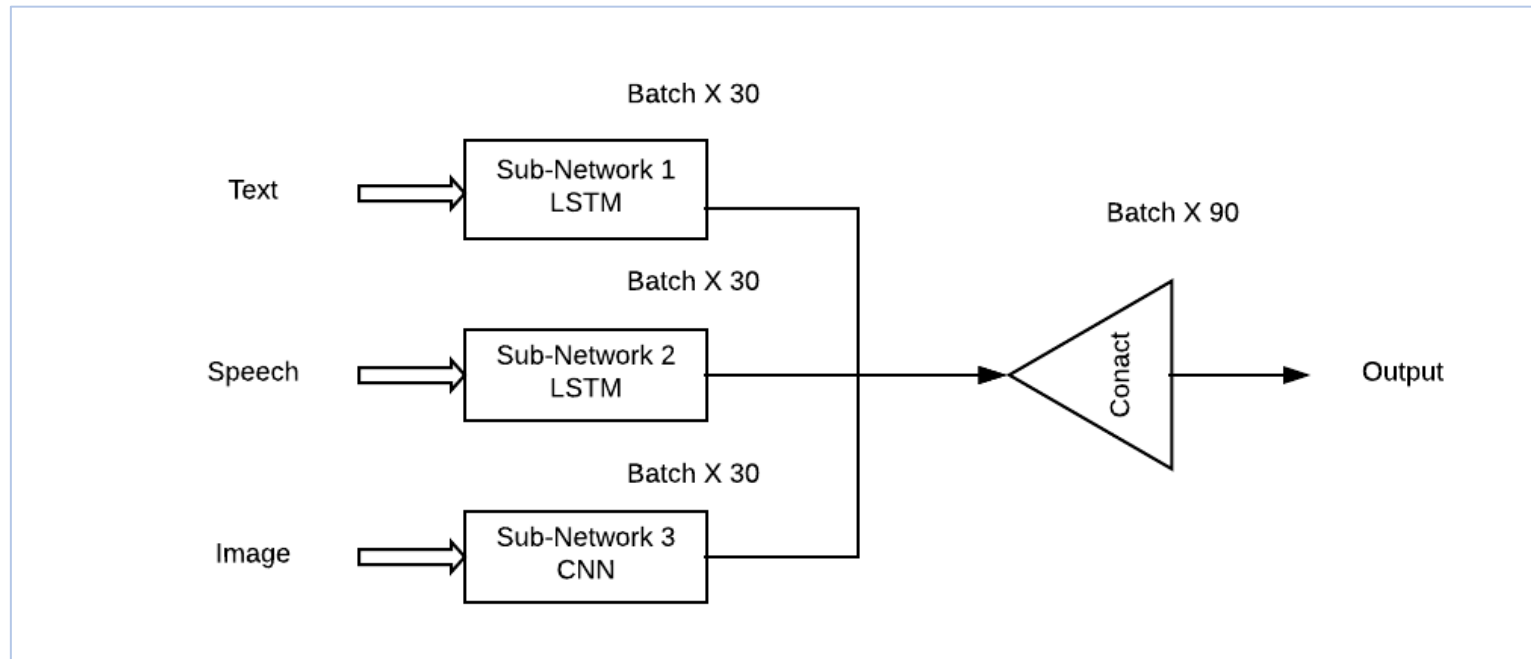
## 5. Multi-Modal 감정 분류 모델



## 5. Multi-Modal 감정 분류 모델



### Multi-Modal Deep Learning



2가지 이상의 데이터 유형을 Input으로 사용



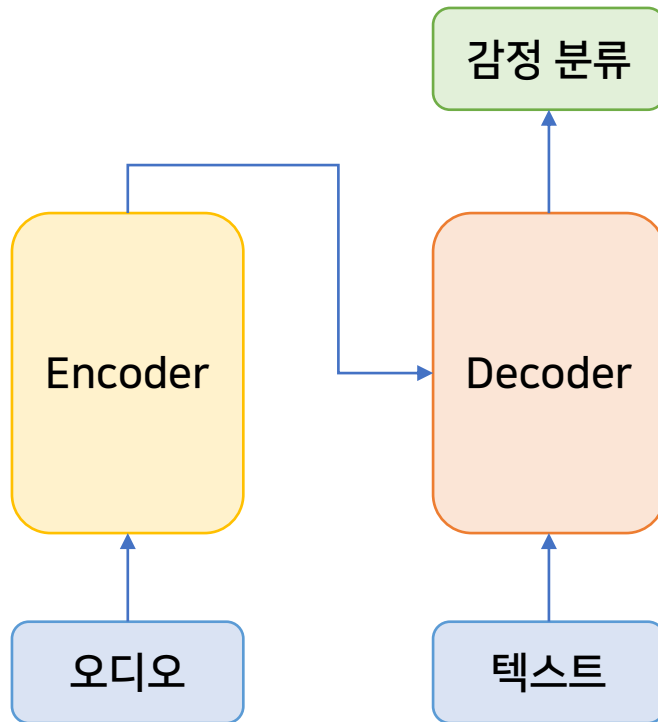


## 5. Multi-Modal 감정 분류 모델

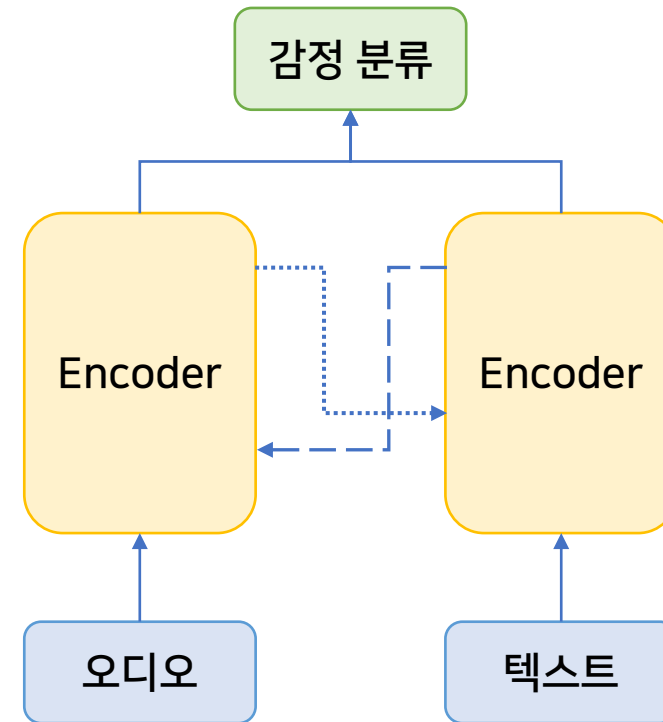


### 모델 고안

#### 1. Multi-modal Transformer



#### 2. Cross Attention Transformer



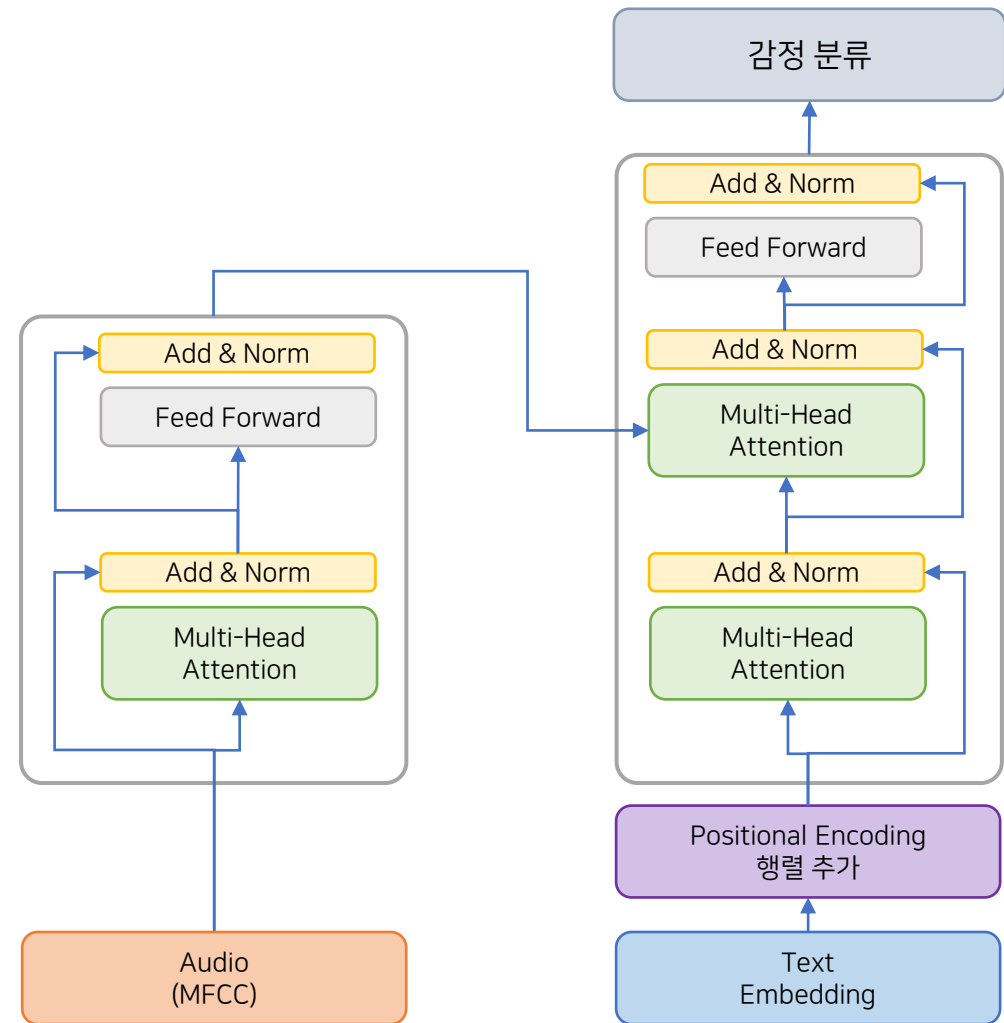


## 5. Multi-Modal 감정 분류 모델



### Multi-modal Transformer

속성	내용
파라미터 수	약 350만 개
감정분류 Cell	FC Layer 활용 예정
특이사항	<ul style="list-style-type: none"><li>음성의 특성을 텍스트에 반영</li><li>최종 분류시 텍스트 Feature만 사용</li><li>텍스트 임베딩: 모델 내에서 진행</li></ul>

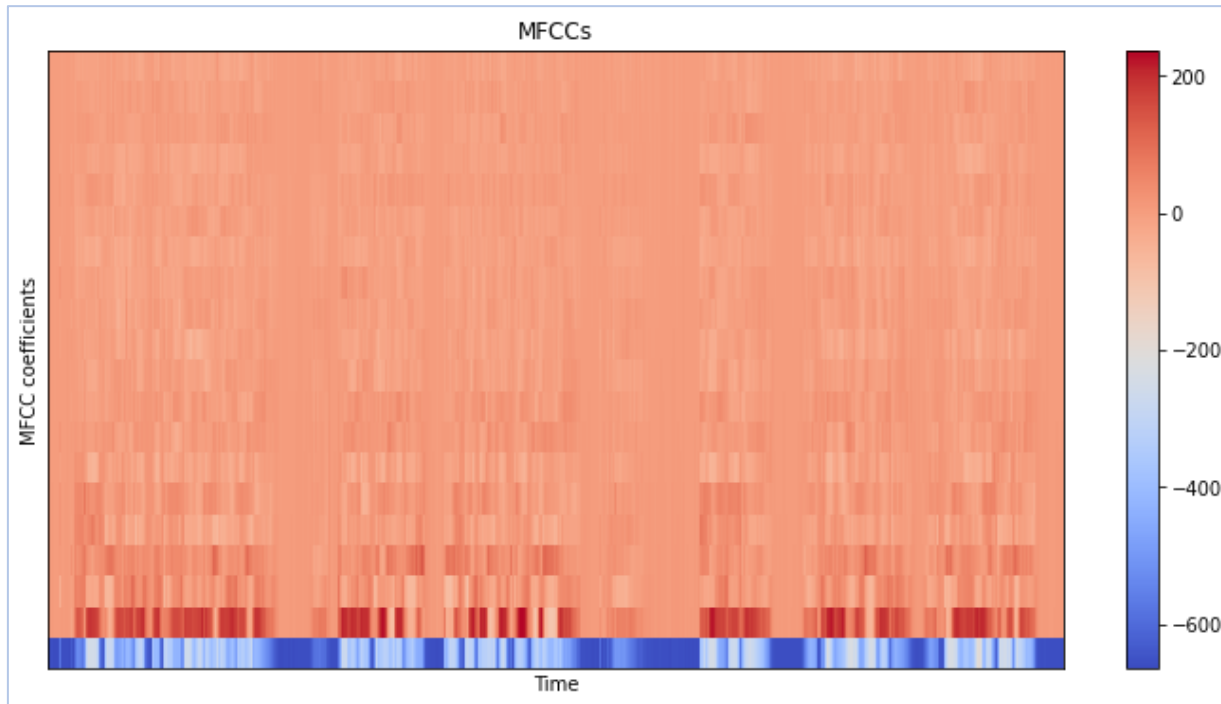




## 5. Multi-Modal 감정 분류 모델



### Multi-modal Transformer – Audio Input



- MFCC ( $n_{mfcc} = 20$ )
- 길이 1000으로 통일
- 1차, 2차 차분값 연결



[1, 1000, 60]





## 5. Multi-Modal 감정 분류 모델



### Multi-modal Transformer – Text Input

```
tensor([[ 272,   10, 5619,   23, 9598,  130,   2,   1,   1,   1,   1,   1,
         1,   1,   1,   1,   1,   1,   1,   1],
        [   5,  105,  729,   16, 1639,   63,  90,   8,  47,  427,   13, 1613,
         0,   8,   18,  264,   29,  42,  229,  229],
        [ 114,   45,   30,  763,   1,   1,   1,   1,   1,   1,   1,   1,
         1,   1,   1,   1,   1,   1,   1,   1],
        [  70,   40,   97,  101,   22,  287,  105,  97,   1,   1,   1,   1,
         1,   1,   1,   1,   1,   1,   1,   1],
        [ 168,   1,   1,   1,   1,   1,   1,   1,   1,   1,   1,   1,
         1,   1,   1,   1,   1,   1,   1,   1]])
```

torchtext의 Field 사용



[60, ] 고정 Input 생성



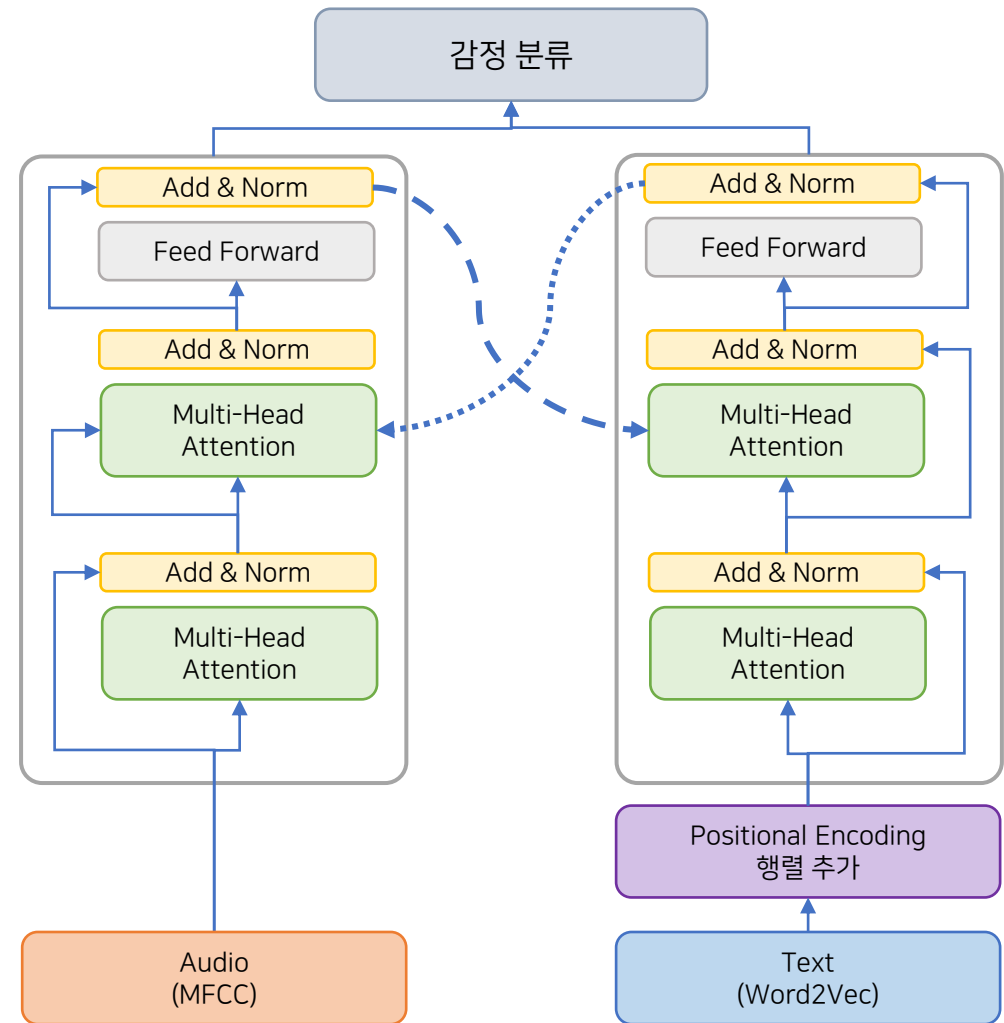


## 5. Multi-Modal 감정 분류 모델



### Cross Attention Transformer

속성	내용
파라미터 수	약 400만 개
감정분류 Cell	Convolutional Layer 사용 예정
특이사항	<ul style="list-style-type: none"><li>서로의 특성을 상호 반영</li><li>최종 분류시 양쪽 Feature Stack</li><li>텍스트 임베딩: Word2Vec</li></ul>



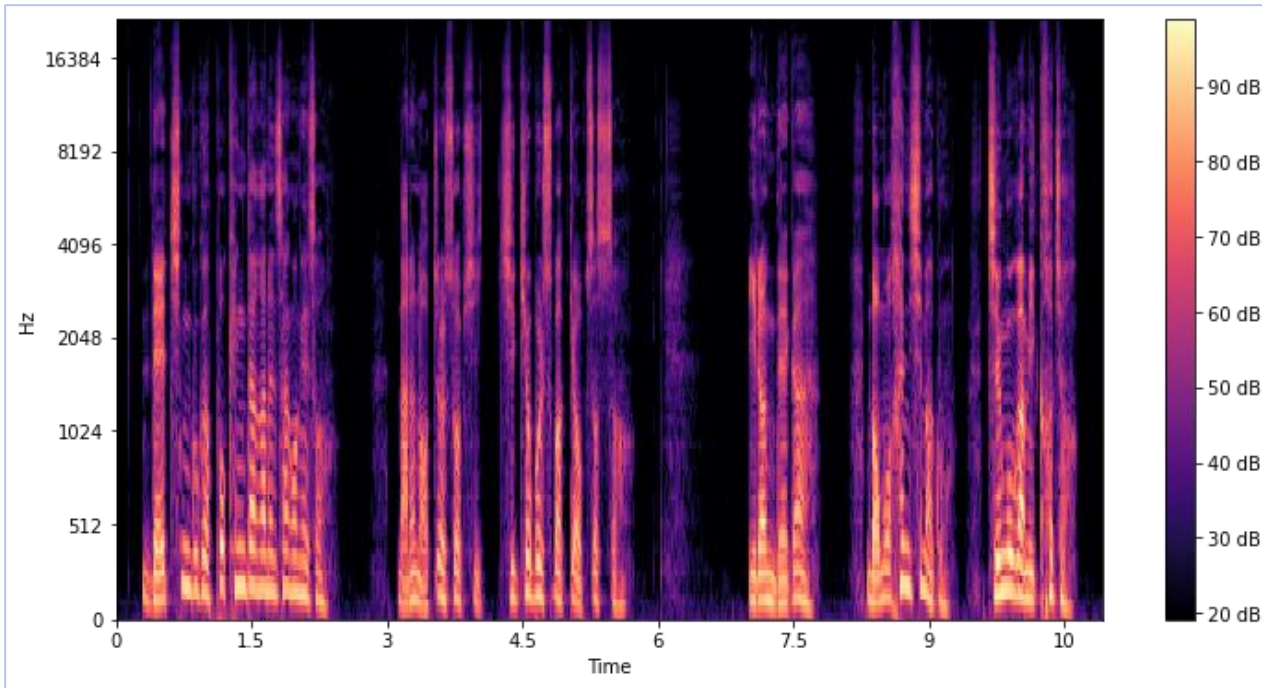




## 5. Multi-Modal 감정 분류 모델



### Cross Attention Transformer – Audio Input



- Mel\_Spectrogram(n\_mels=128)
- 길이 1000으로 통일
- 1차 차분값 연결



[1, 1000, 256]

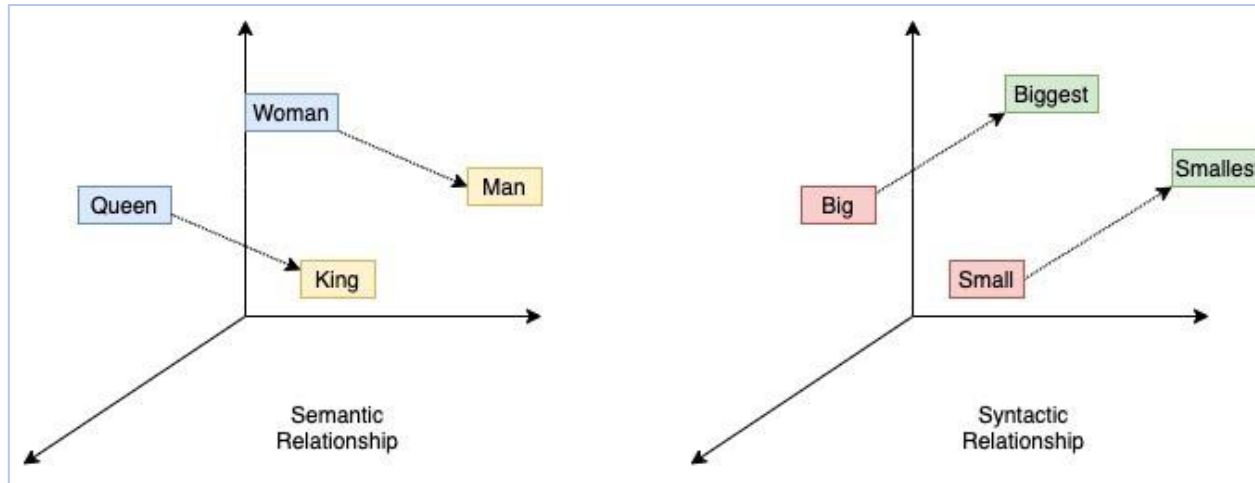




## 5. Multi-Modal 감정 분류 모델



### Cross Attention Transformer – Text Input



- Tokenizer: Okt(konlpy)
- 데이터셋 단어 수: 19612개
- 임베딩 차원: 256
- 문장 길이 제한: 40



[1, 40, 256]



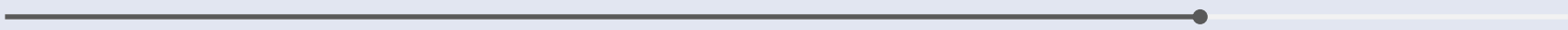


## 3주차 계획



### 트랜스포머 모델 학습

- 고안된 트랜스포머 모델을 학습시켜 음성 분석 모델과 텍스트 분석 모델과의 정확도 비교
- 우울증 데이터를 받아서 분석 - 우울증과 감정의 분포를 확인
- 음성인식 모델과 연동 / 음성을 텍스트화하여 트랜스포머 모델로 분석
- 플랫폼을 만들어서 직접 테스트할 수 있도록 만들 예정!!





감사합니다