

# 딥러닝 기반 소셜 미디어 텍스트 우울 경향 분석

평일 저녁 4조  
김지현, 김예지, 박민수, 한혜승

Contents

**문제 정의 및 목표 설정**

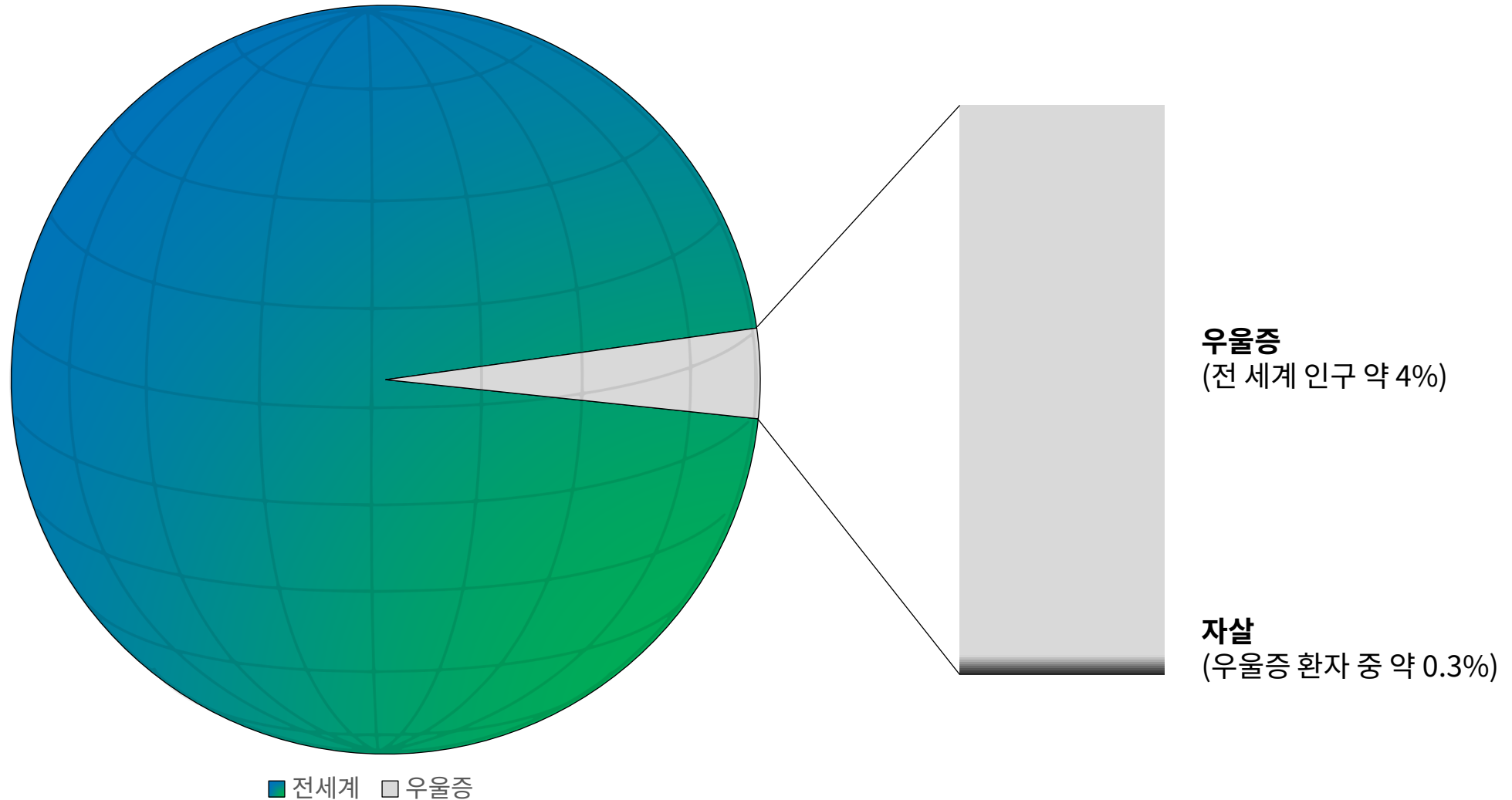
데이터 수집 및 전처리

EDA

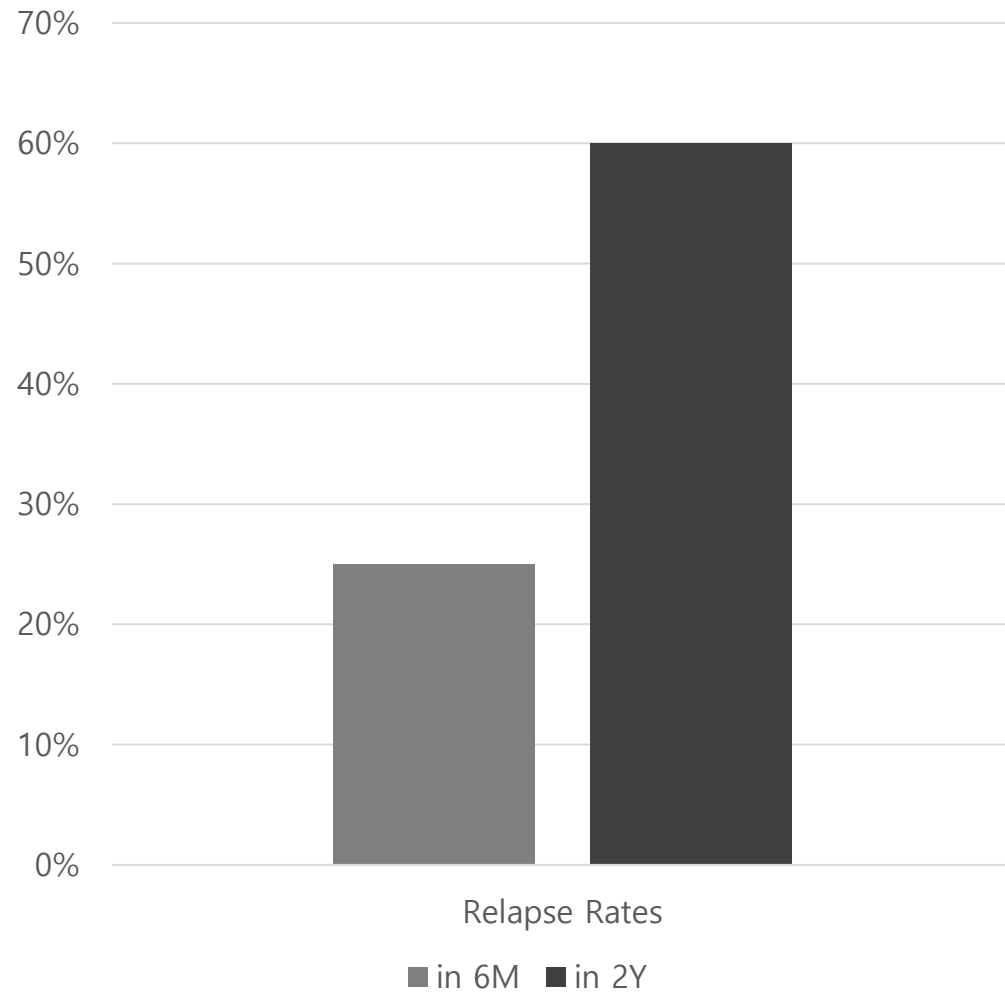
모델링

결과

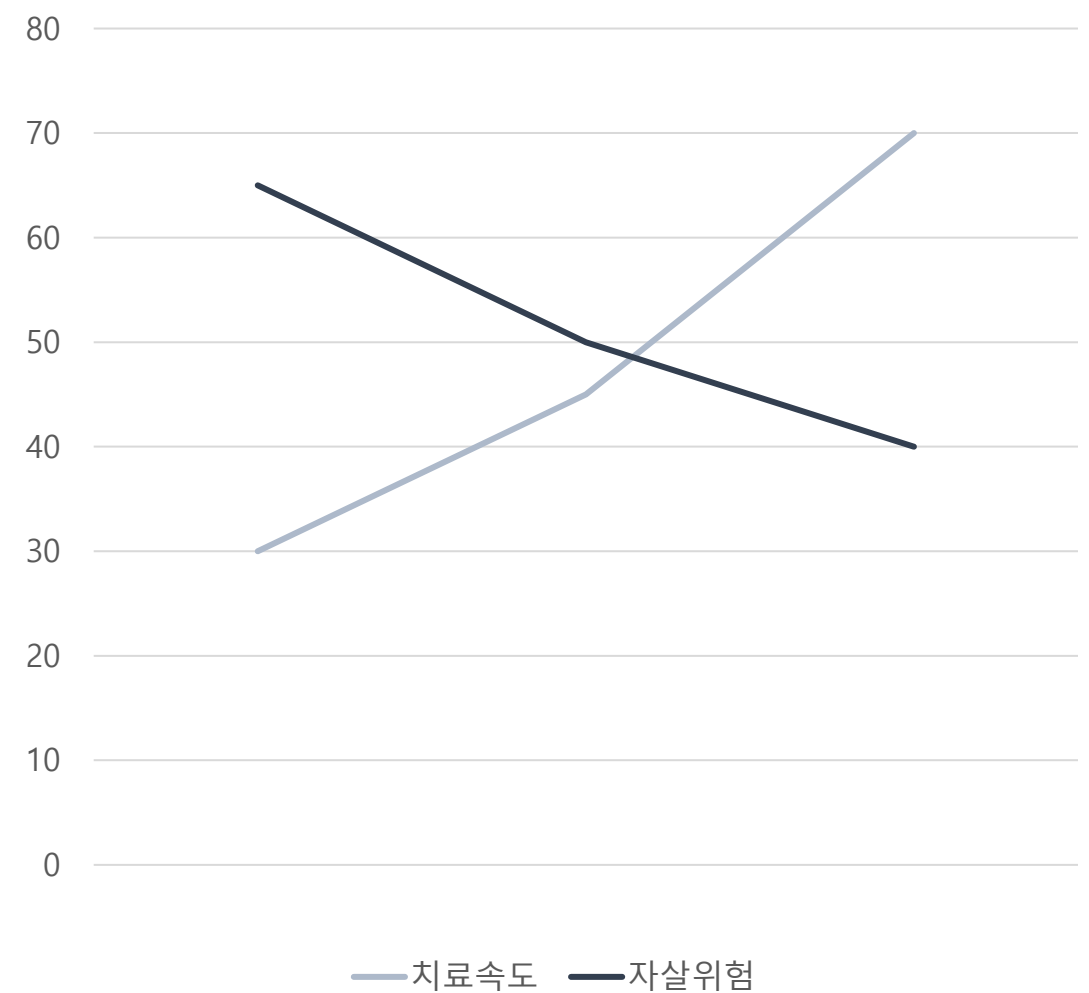
# 1. 문제 정의 및 목표 설정

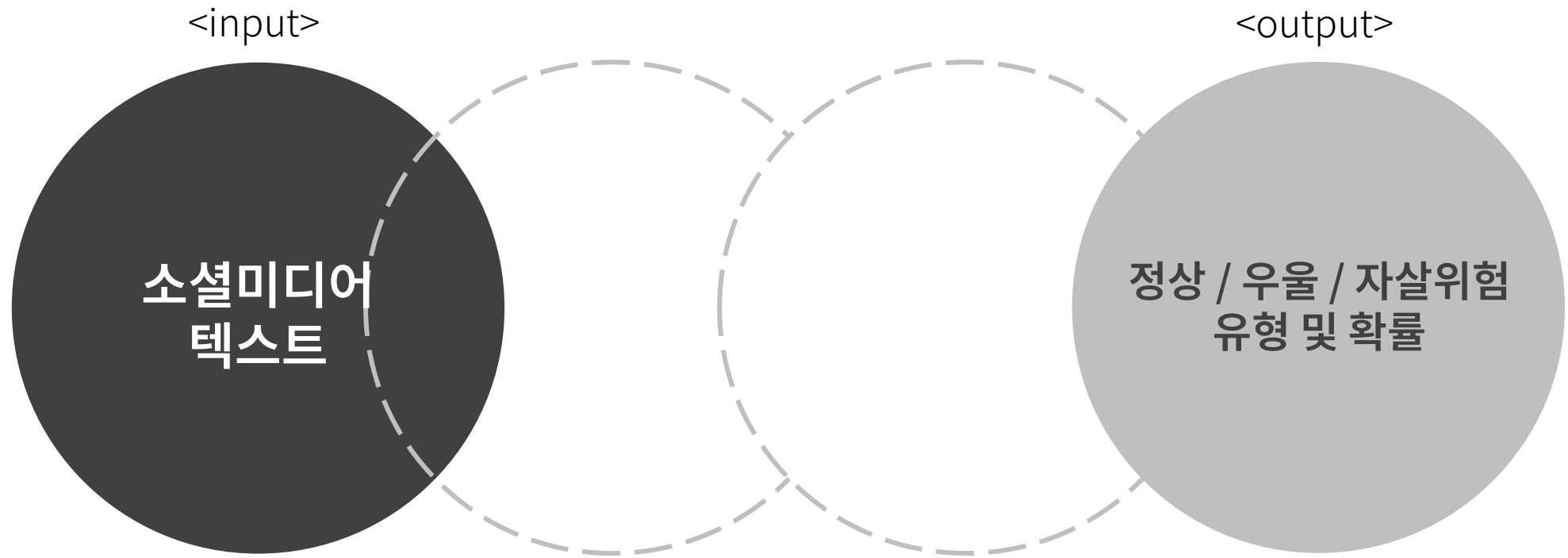


우울증을 치료하지 않을 경우 재발률



우울증 조기 치료의 영향





★

이 프로젝트를 통해 우울증 경향을 조기 예측하여 적절한 대응 조치를 취하고  
우울증이라는 사회 문제 해결에 기여하기를 기대함

Contents

문제 정의 및 목표 설정  
**데이터 수집 및 전처리**  
EDA  
모델링  
결과

## 2. 데이터 수집 및 전처리

<Reddit, kaggle>

- Depression dataset
- Suicide and Depression Detection dataset

0(정상)	약 12만건
1(우울)	약 3000건
2(자살)	약 11만건

(전처리)  
소문자 변환  
특수문자/이모티콘 제거  
불용어 제거  
숫자 제거  
빈 문자열 제거

Contents

문제 정의 및 목표 설정  
데이터 수집 및 전처리

**EDA**

모델링

결과

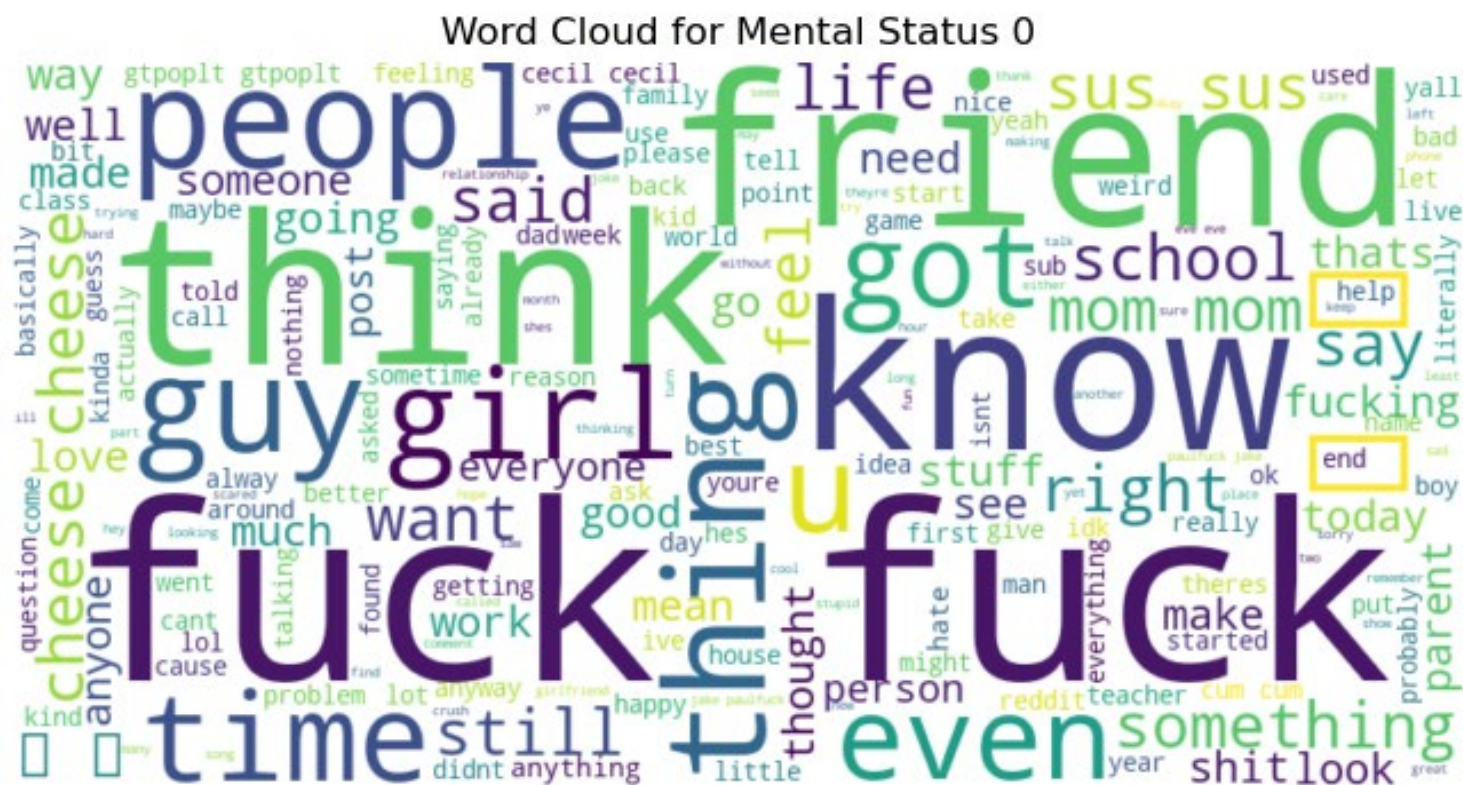


### 3. EDA

Total number of words (including duplicates): 12,710,934

Total of unique words: 210,972

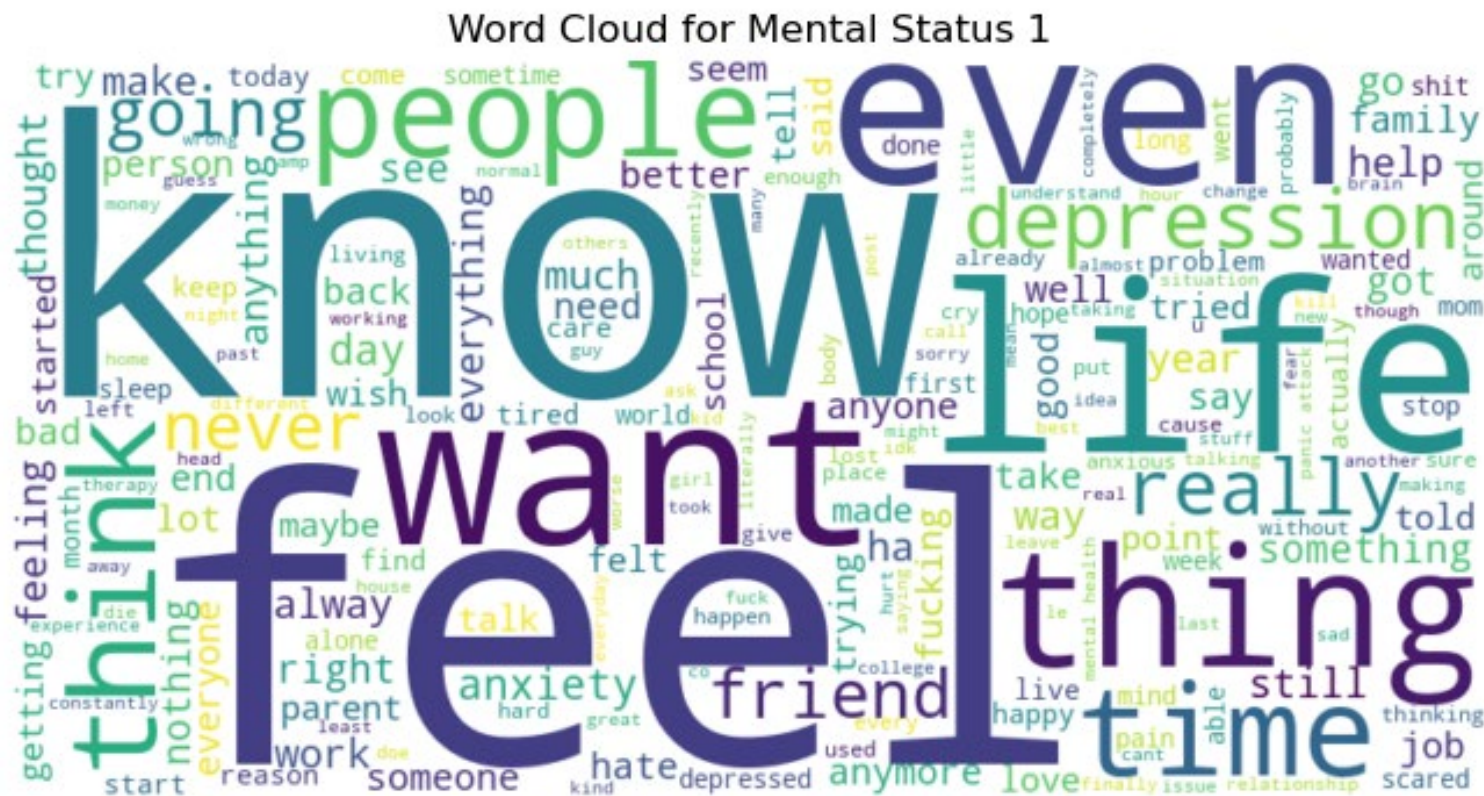
### 3. EDA - normal



## TF-IDF

Normal
Know
Want
People
Day
Really
Got
Time
Think
Good
feel

### 3. EDA - depressed

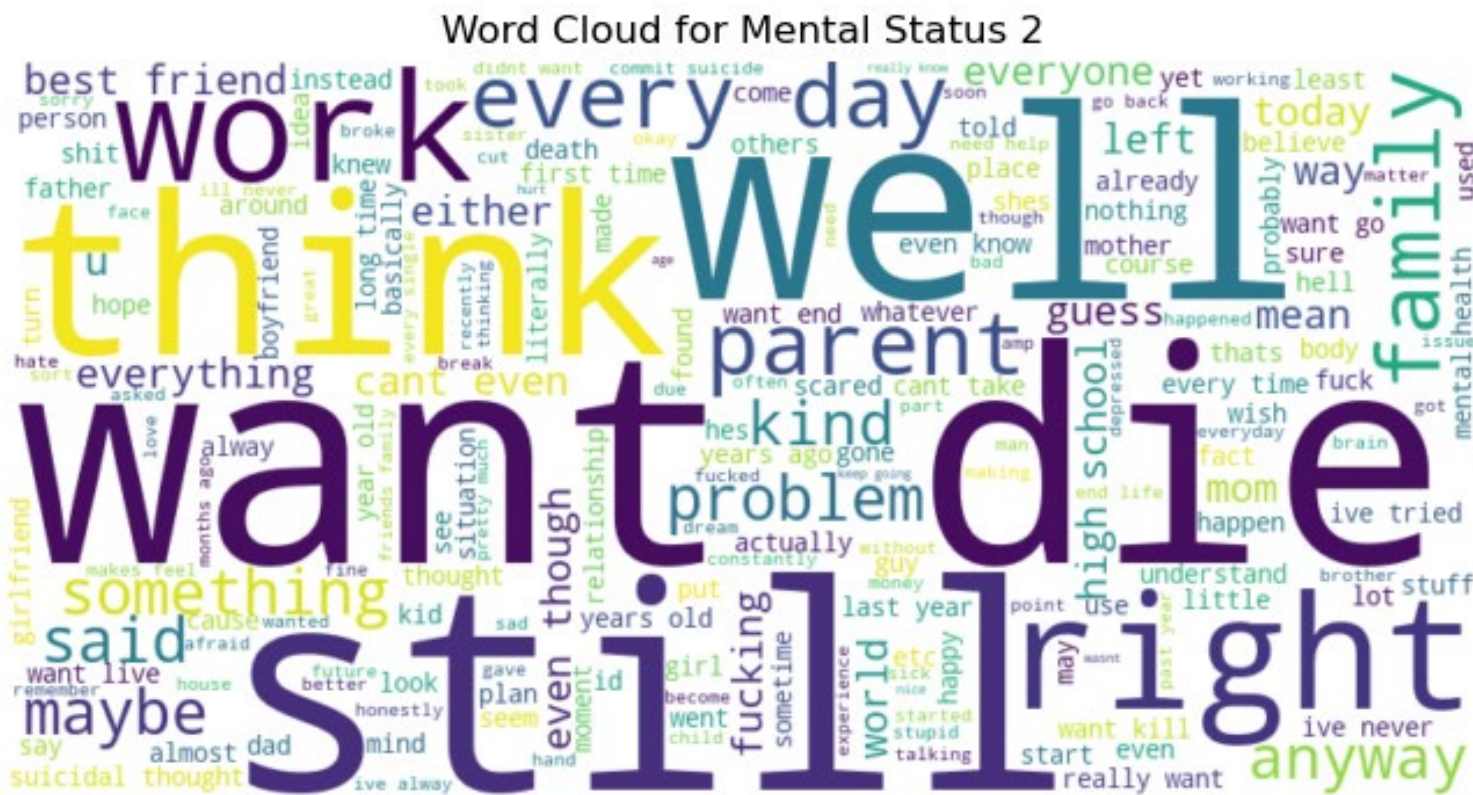


## TF-IDF

Depressed
Depression
Feel
Want
Anxiety
Know
Life
People
Time
Day
really



### 3. EDA - suicidal



## TF-IDF

Suicidal
Want
Feel
Life
Know
People
Really
Time
Going
Think
help

Contents

문제 정의 및 목표 설정  
데이터 수집 및 전처리  
EDA  
**모델링**  
결과

## 4-1. 모델 선정

### / NaiveBayes

조건부 확률을 기반으로 각 특징이 독립적이라고 가정하여 빠르고 효율적으로 분류를 수행하는 머신러닝 알고리즘

### / BERT

양방향 트랜스포머 모델로, 사전 훈련된 대규모 언어 모델을 통해 문맥에 대한 깊은 이해를 제공

### / Bi-LSTM

양방향 LSTM 네트워크를 사용해 텍스트의 앞뒤 문맥을 동시에 고려하여 시퀀스 정보를 효과적으로 학습

### / GRU

게이트 순환 유닛(GRU)으로, LSTM보다 간단하고 연산 효율적인 순환 신경망으로 시퀀스 데이터를 처리

# # 데이터 전처리



## 4-2. 모델평가

<b>NaiveBayes/50000</b>	precision	recall	f1-score	support
Normal	0.98	0.77	0.86	24076
Depression	0.43	0.93	0.59	773
Suicide	0.83	0.98	0.89	23103
Accuracy			0.87	47952
macro avg	0.74	0.89	0.78	47952
weighted avg	0.90	0.87	0.87	47952

<b>Bi-LSTM/baseline</b>	precision	recall	f1-score	support
Normal	0.93	0.93	0.93	9592
Depression	0.63	0.80	0.71	307
Suicide	0.93	0.93	0.93	9282
Accuracy			0.93	19181
macro avg	0.83	0.89	0.86	19181
weighted avg	0.93	0.93	0.93	19181

<b>BERT/baseline</b>	precision	recall	f1-score	support
Normal	0.97	0.98	0.98	23955
Depression	0.95	0.89	0.92	756
Suicide	0.98	0.98	0.98	23241
Accuracy			0.98	47952
macro avg	0.97	0.95	0.96	47952
weighted avg	0.98	0.98	0.98	47952



## 4-3. BERT 모델 선정

<i>BERT/baseline</i>	precision	recall	f1-score	support
Normal	0.97	0.98	0.98	23955
Depression	0.95	0.89	0.92	756
Suicide	0.98	0.98	0.98	23241
Accuracy			<b>0.98</b>	47952
macro avg	<b>0.97</b>	<b>0.95</b>	<b>0.96</b>	47952
weighted avg	0.98	0.98	0.98	47952

Text: I worked hard today! It's rewarding. – Normal (0.9994)

Text: I **don't wanna live** anymore - **Normal** (0.8831)

Text: I wanna **suicide**. – **Normal** (0.6292)

## 4-3. BERT 모델 선정

<i>BERT/30000</i>	precision	recall	f1-score	support
Normal	0.98	0.98	0.98	23981
Depression	0.88	0.92	0.90	766
Suicide	0.98	0.98	0.98	23205
Accuracy			<b>0.97 (-0.01)</b>	47952
macro avg	0.94	0.96	0.95	47952
weighted avg	0.97	0.97	0.97	47952

Text: I don't wanna live anymore - Suicide (0.7856)

Text: I don't see the point in anything anymore. - Suicide (0.6727)

Text: I wanna **suicide**. - **Normal** (0.7332)

## 4-4. 모델 보완

### 목표

분명한 상태를 드러내는 몇십가지의 텍스트들을 넣었을 때  
명확히 분류해낼 수 있는 모델 구축

### 개선사항

#### 1. 특정 단어에 가중치 추가

- EDA, 동의어를 찾아 감정을 나타내는 직접적인 단어는 가중치를 줘 강한 판단력을 기대함.
- 자살(1.8), 우울(1.5) 가중치를 추가해 우울보단 자살 키워드를 우선 판단하게 학습시킴.
- 자살이 더 높은 위험성을 가진 주제라고 판단

#### 2. Early stopping 설정

- 학습의 효율성 높임. 일정 수준 이상 성능 향상이 되지 않으면 학습 종료.

## # 토픽 분석

ALL

Topic 1:

fuck didnt said mom got friend told know friends time

Topic 2:

want feel know ive life people really think help time

Topic 3:

school got need time post know really class play good

Topic 4:

fucking people day shit hate fuck want youre know say

Topic 5:

life ive want years job time work school going family

## depressed

Topic 1:  
depression people feel life want friend know person make http

Topic 2:  
feel want know life time thing really day think year

Topic 3:  
depression year le pain amp la got pression http told

## suicidal

Topic 1:  
job school ive life years time got work year family

Topic 2:  
time ive got didnt started years day went felt life

Topic 3:  
life people want feel fucking ive hate know world live

Contents

문제 정의 및 목표 설정  
데이터 수집 및 전처리  
EDA  
모델링  
**결과**

## 6. 결과 - 창작 텍스트 평가

### 방법

토픽분석 EDA 기반

정상 4 / 우울 14 / 자살 14개의 문장 창작

### 라벨별성능

0. Normal (정상): 100%

1. Depressed (우울): 66.7%

- 절반 이상 맞추지만, 오답이 존재

2. Suicidal (자살): 31.25%

- 자살 텍스트에서 오답 빈번하게 발생

## 카테고리, 텍스트, 한글 해석, 예측 라벨, 정답 라벨

토픽	Depression ...	우울증은 사...	Depressed	Depressed
토픽	Depression ...	우울증은 내...	Depressed	Depressed
토픽	Some days, ...	어떤 날은 시...	Normal	Depressed
토픽	I often fee...	나는 종종 압...	Depressed	Depressed
토픽	The depress...	오랫동안 짐...	Depressed	Depressed
토픽	After a yea...	우울증과 끝...	Depressed	Depressed
토픽	Balancing l...	우울증, 학교...	Depressed	Depressed
토픽	Balancing l...	지난 1년 동...	Normal	Depressed
토픽	Anxiety att...	직장에서 불...	Suicidal	Suicidal
토픽	Every day a...	매일 직장에...	Normal	Suicidal
토픽	I've spent ...	나는 수년간 ...	Suicidal	Suicidal
토픽	I've strugg...	나는 수년간 ...	Normal	Suicidal
토픽	I've been t...	나는 수년이 ...	Suicidal	Suicidal
토픽	I've got no...	시간이 흘러...	Suicidal	Suicidal
토픽	I hate how ...	세상이 나에...	Depressed	Suicidal
토픽	I hate how ...	세상이 나를 ...	Depressed	Suicidal
토픽	I want to e...	나는 오랫동안...	Depressed	Suicidal
토픽	I want to f...	내 삶에 의미...	Depressed	Suicidal
토픽	I've tried ...	나는 친구들...	Suicidal	Suicidal
토픽	I really kn...	사람들과 친...	Normal	Suicidal



① I've tried to talk to friends, but no matter how much help I want, it feels like no one really understands.

나는 친구들과 이야기를 나누려 했지만, 아무리 도움을 받고 싶어도 아무도 나를 진정으로 이해하지 못하는 것 같다.

② I've spent years juggling jobs, school, and life, but it all feels like too much, even for my family to understand.

나는 수년간 직장, 학교, 그리고 삶의 균형을 맞추기 위해 애써왔지만, 이 모든 것이 너무 무겁게 느껴지며, 가족조차 이해하지 못하는 것 같다.

③ Anxiety attacks me at work, and no matter how much time passes, I don't know if I can handle it without help.

직장에서 불안이 나를 덮치고, 시간이 지나도 도움 없이는 이를 감당할 수 있을지 모르겠다.

④ Balancing life with depression, school, family, and anxiety feels like a job I never signed up for.

우울증, 학교, 가족, 불안을 함께 균형 맞추는 것은 내가 선택하지 않은 일처럼 느껴진다.

	정답라벨	30000(가중치)	30000	baseline
1	Suicidal	Suicidal	Normal	Normal
2	Suicidal	Suicidal	Normal	Normal
3	Suicidal	Suicidal	Normal	Normal
4	Depressed	Depressed	Normal	Normal

# 결론

**이번 프로젝트에서 딥러닝을 사용해 우울증과 자살과 같은 사회적 문제를  
조기에 발견하고, 예방할 수 있는 가능성을 확인할 수 있었습니다.  
추가적인 모델 개선을 통해 더욱 정확한 분석과 대응이 가능할 것으로 기대됩니다.**

**감사합니다**