

# Capstone Project

Detection of Suicidal Posts through Reddit Posts

Johan Ng

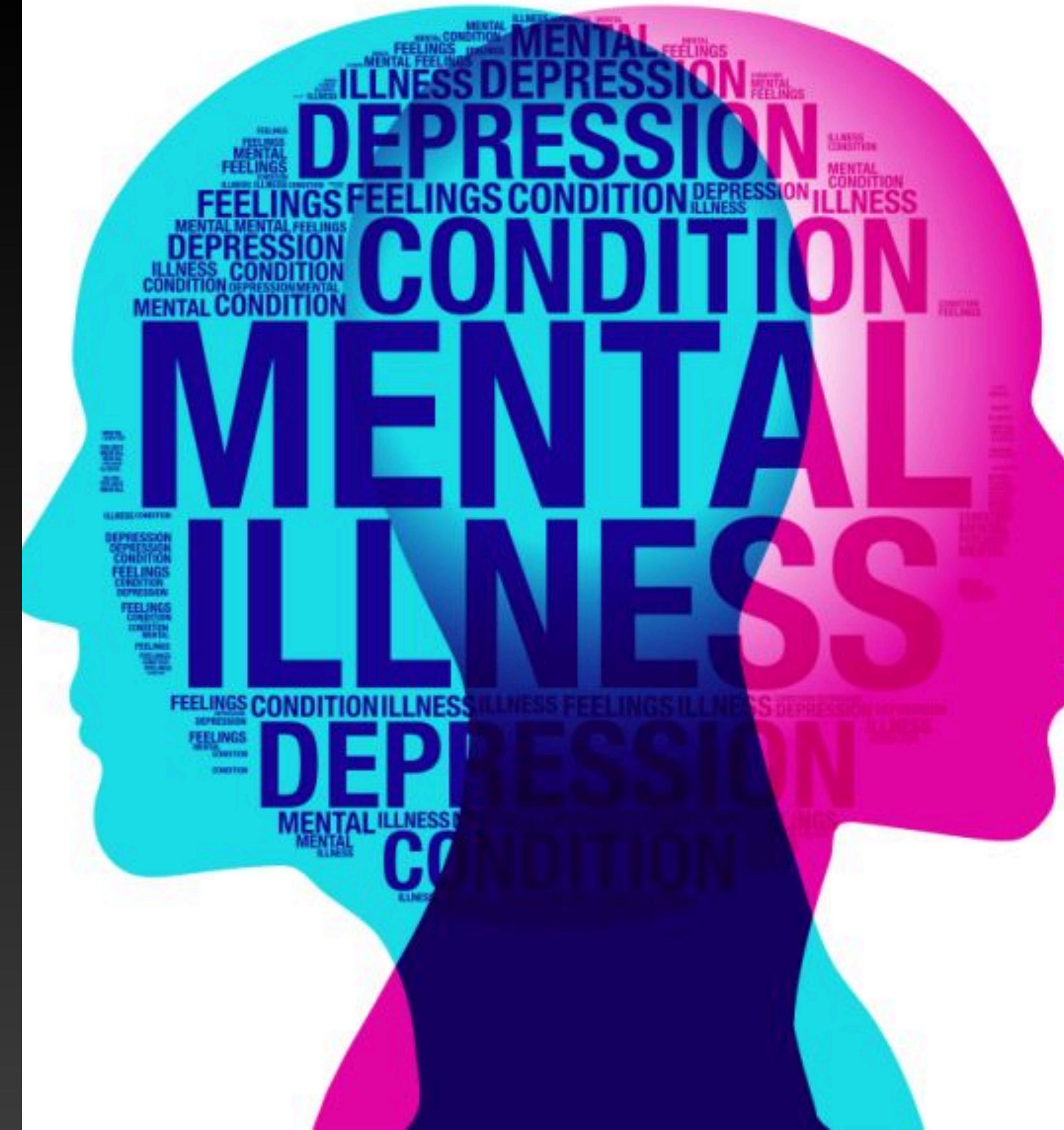
# Content Page

- Background
- Problem Statement
- Dataset and Process
  - Text Preprocessing
  - Topic Modelling
  - Detecting Suicidal Posts
- Value Proposition
- Conclusion

# Background

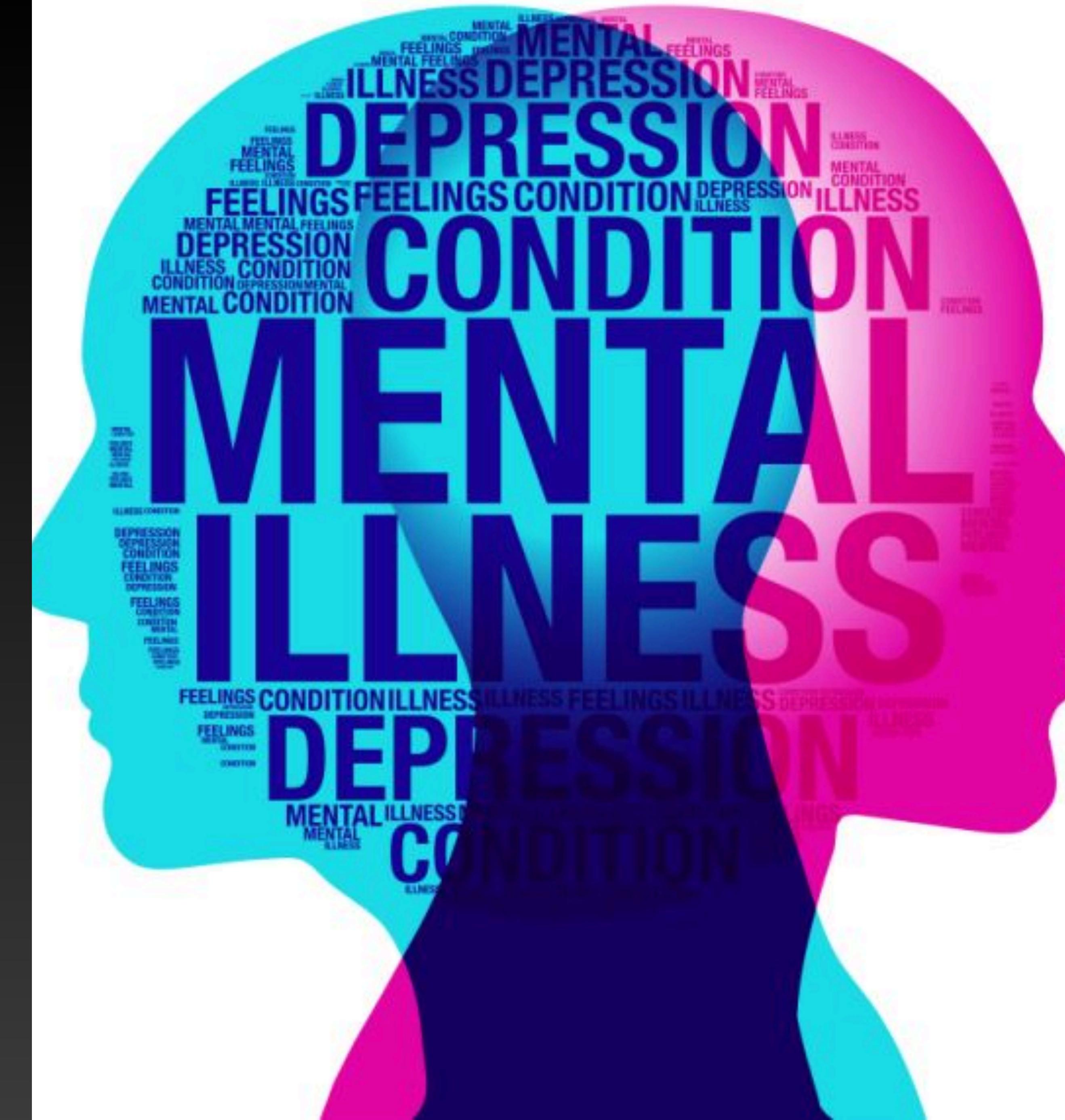
# Suicide: A Global Health Problem

- 800,000 deaths/year globally



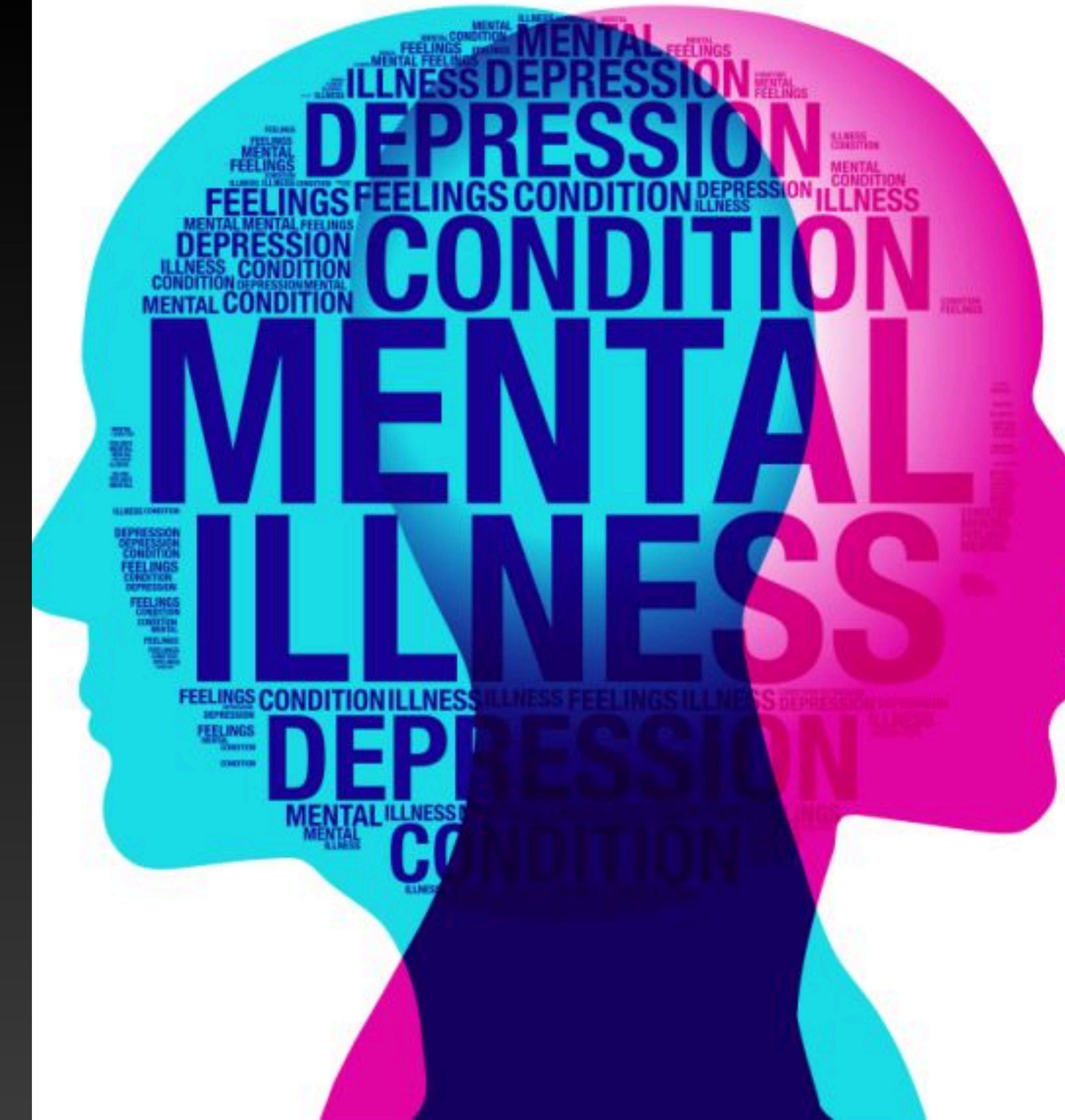
# Suicide: A Global Health Problem

- 800,000 deaths/year globally
- Leading cause of death in 10 - 29 years old (Singapore)



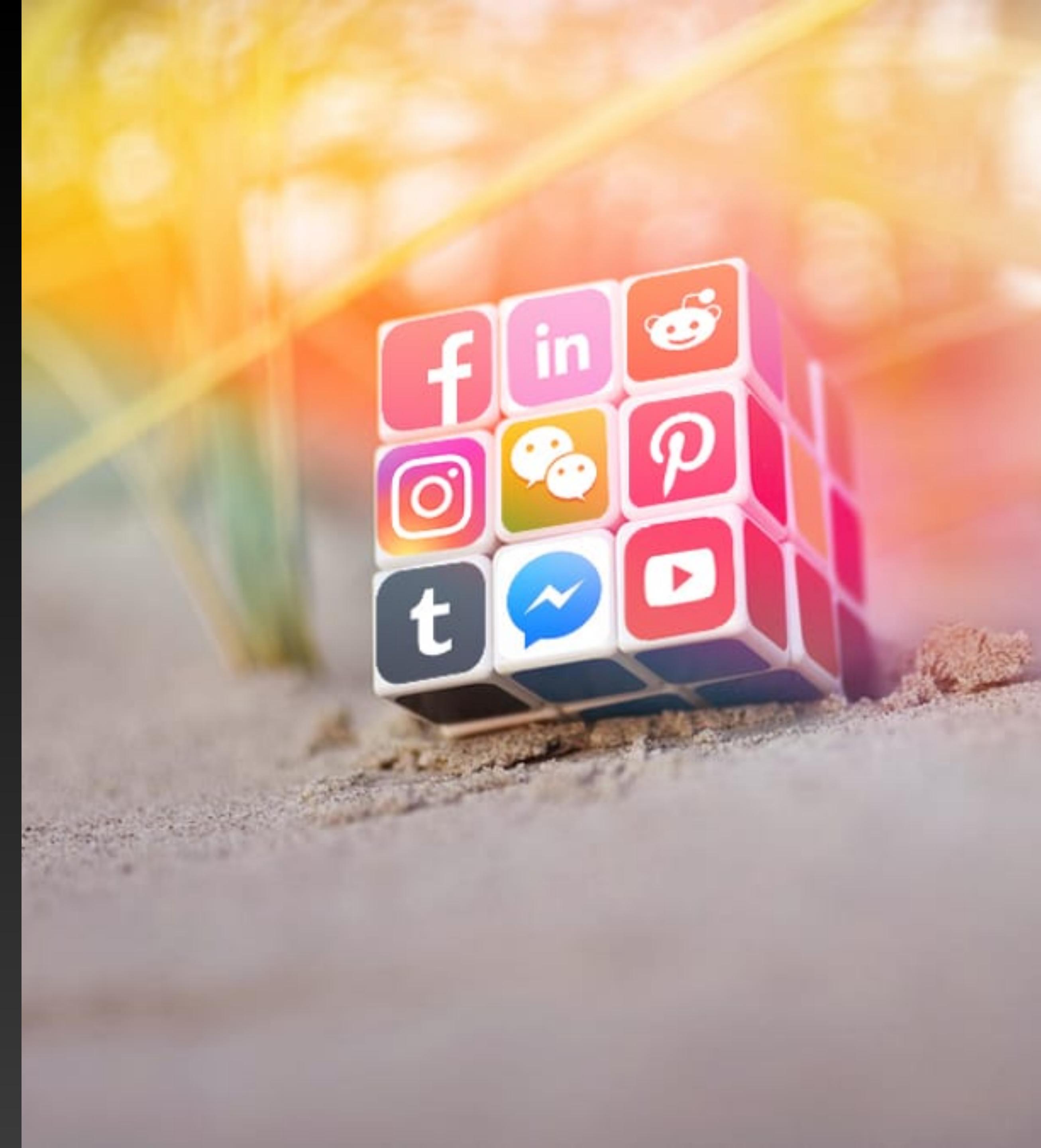
# Suicide: A Global Health Problem

- 800,000 deaths/year
- 3rd Leading cause of death in 15 - 19 years old
- PREVENTABLE



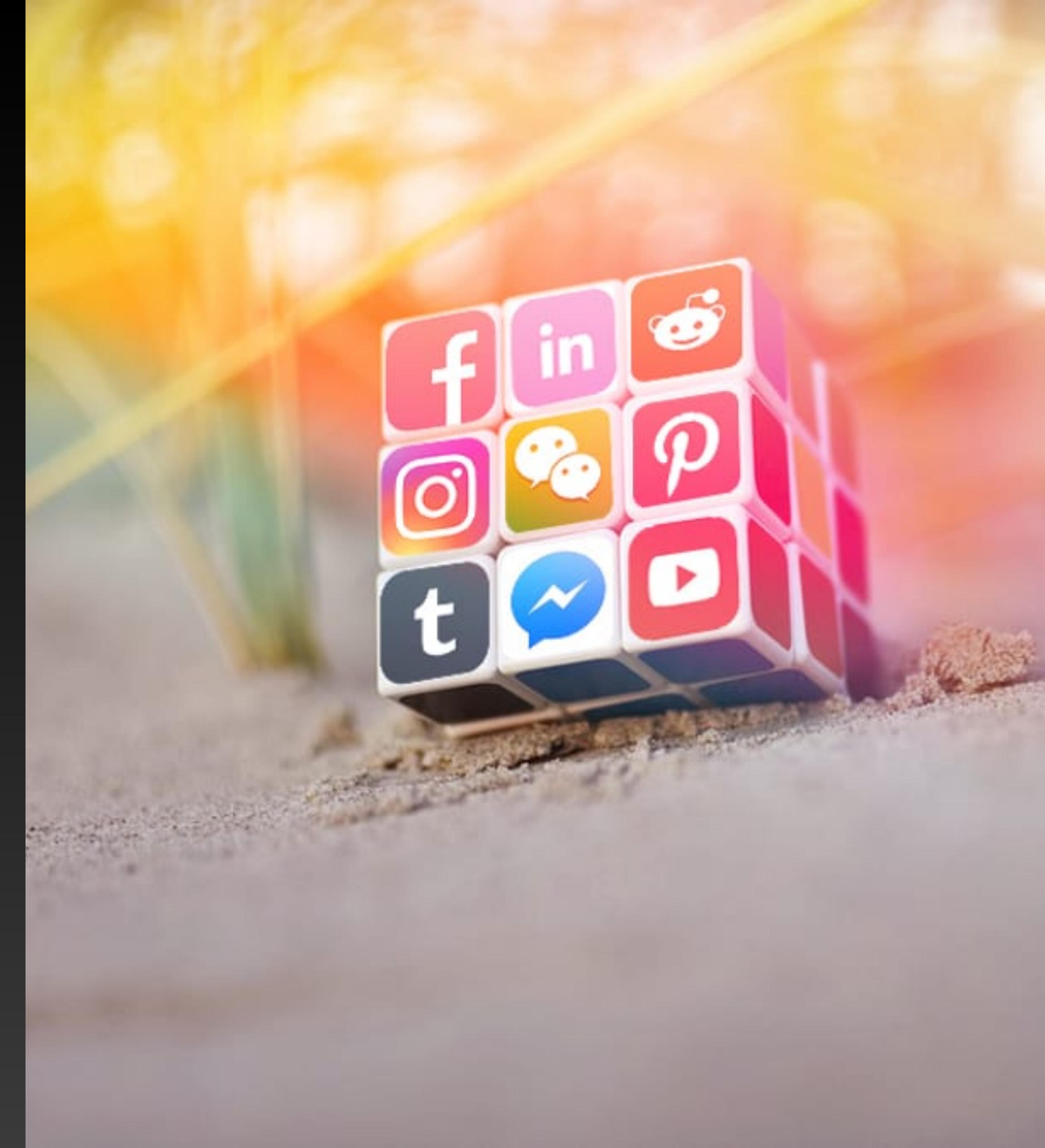
# Social Media and Mental Illness

- Safe space for expressing one's feelings (Anonymity)



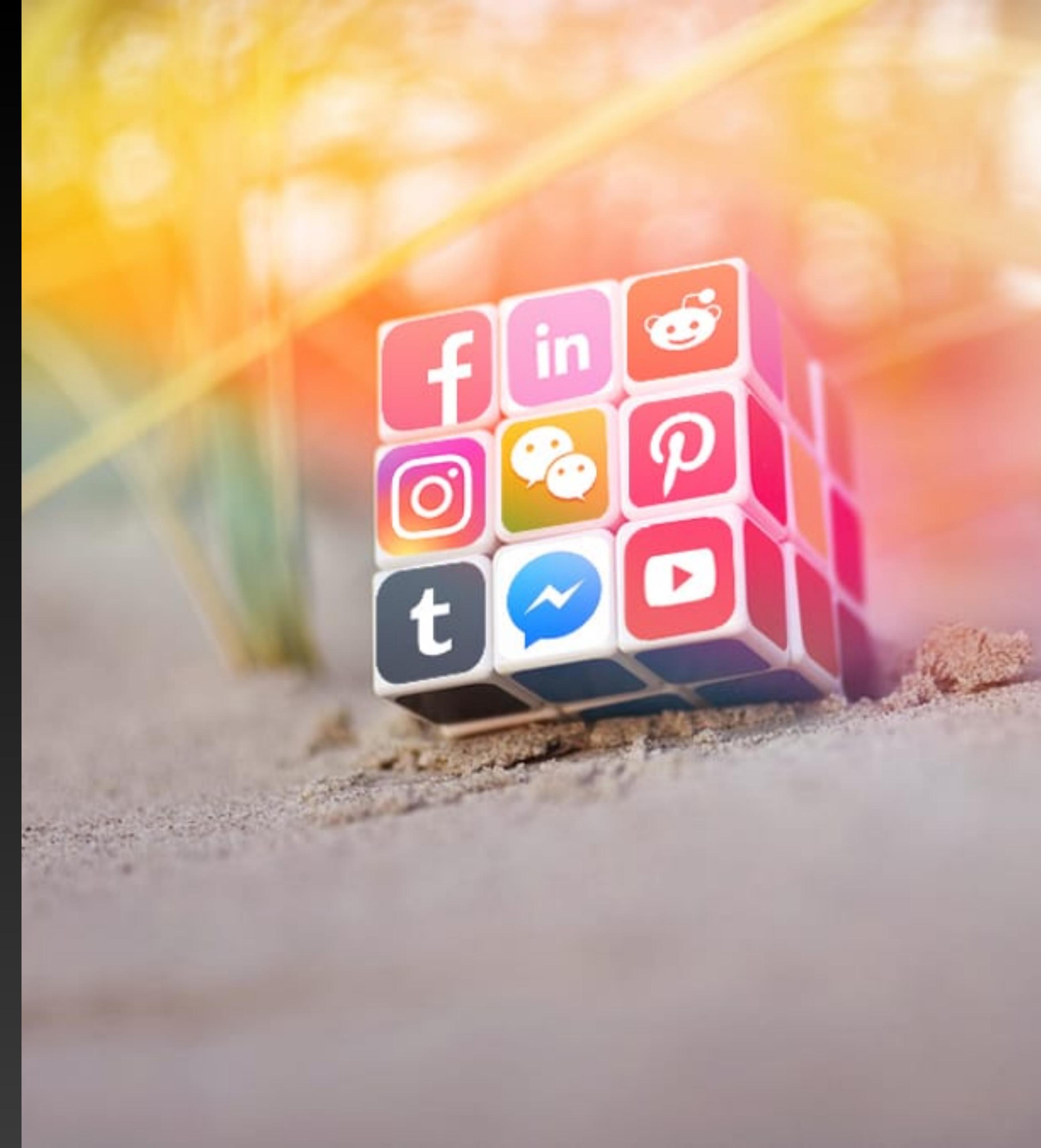
# Social Media and Mental Illness

- Safe space for expressing one's feelings (Anonymity)
- Strong community for expressing empathy (No Stigma)



# Social Media and Mental Illness

- Safe space for expressing one's feelings (Anonymity)
- Strong community for expressing empathy
- Natural medium to observe users' mental illnesses based on their posts



**“Youths are more likely to disclose suicidal thoughts and suicidal risk factors on social media”**

Source: A machine learning approach predicts future risk to suicidal ideation from social media data

# Problem Statement

# Stakeholder

- Software Company

# Stakeholder

- Software Company
  - Specialises in Mental Health Apps

# Stakeholder

- Software Company
  - Specialises in Mental Health App
  - Developing an application to detect suicidal ideation from social media data

# Stakeholder

- Software Company
  - Specialises in Mental Health App
  - Developing an application to predict suicidal ideation from social media data
    - Risk Monitoring tool
    - Prevention tool
    - Support Community

**“Can we use Reddit to detect  
posts with suicidal ideation?**

Problem Statement

# Dataset and Process

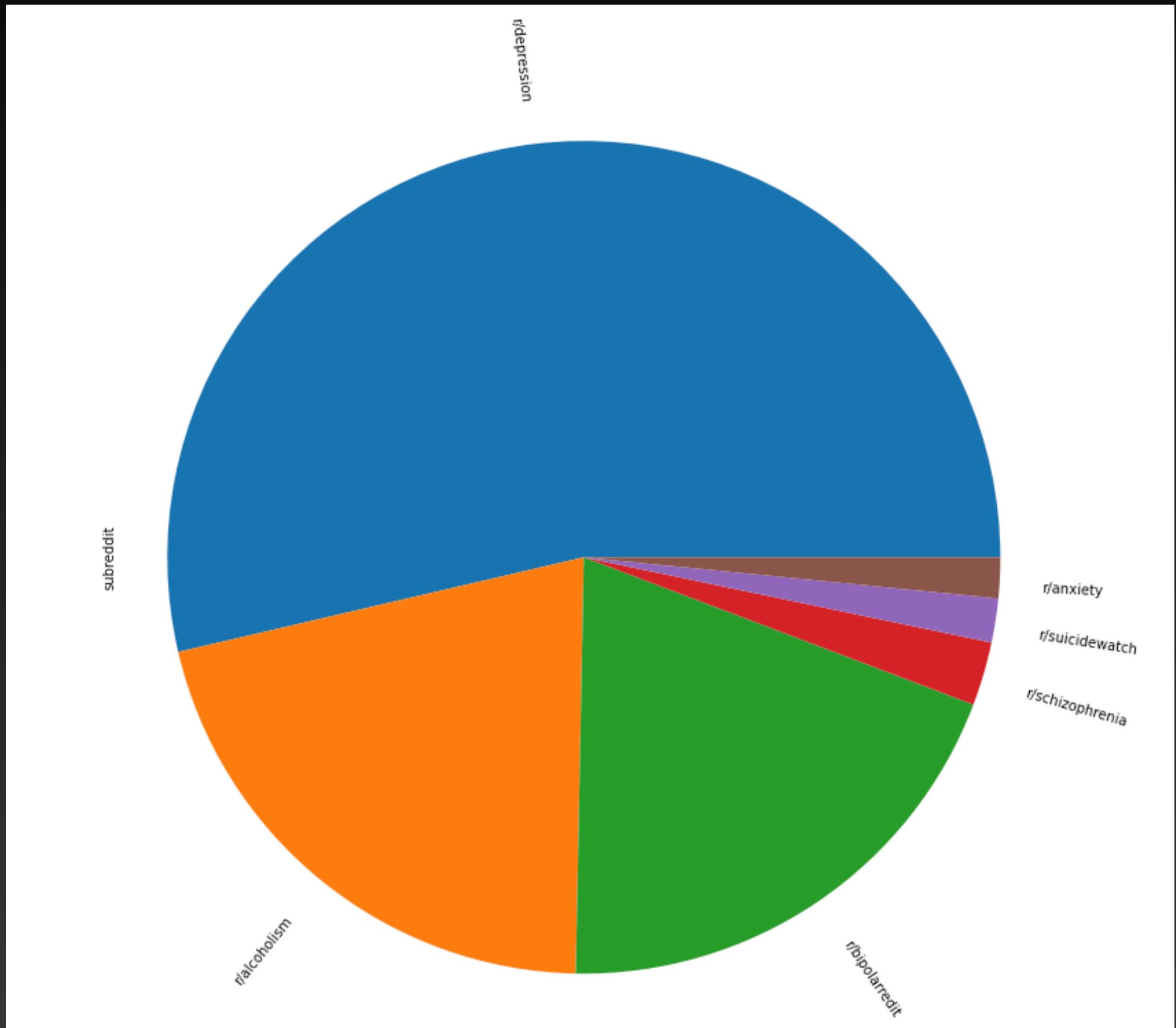
Source of Dataset:

**Natural Language Processing Reveals Vulnerable Mental Health  
Support Groups and Heightened Health Anxiety on Reddit  
During COVID-19: Observational Study**

# Reddit Dataset

- Jan to Apr (2018,2019,2020) - **179,154 unique posts**
- 6 Subreddits:
  - r/depression (53.7%)
  - r/suicidewatch (19.5%)
  - r/anxiety (21.0%)
  - r/schizophrenia (2.5%)
  - r/alcoholism (1.6%)
  - r/bipolarreddit (1.7%)

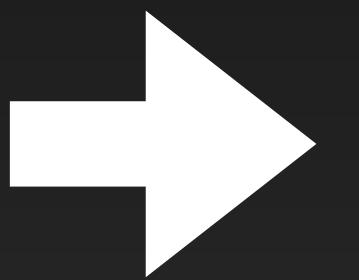
Figure 1: Distribution of subreddits in DataSet



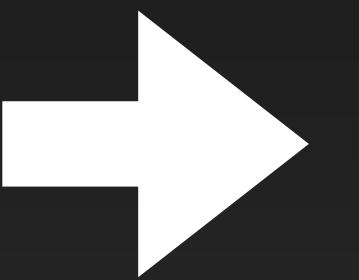
# Process



Reddit Posts



Step 1: Text Preprocessing

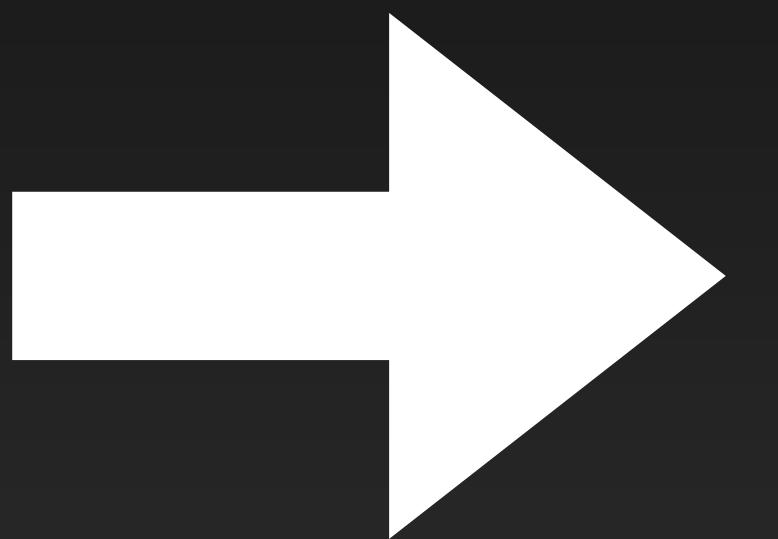


Step 2: Topic Modelling and Interpretation

# Process



**Step 3: Manual Classification of Posts**



**Step 4: Predictive Modelling**

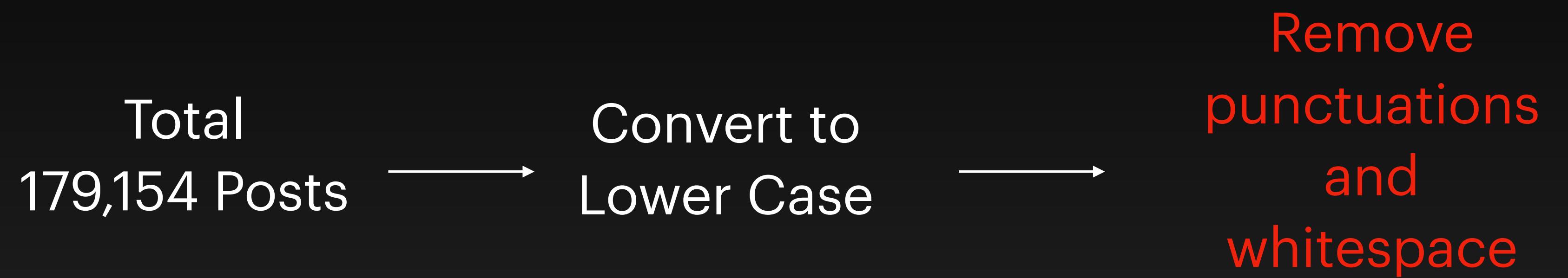
# Step 1: Text Preprocessing



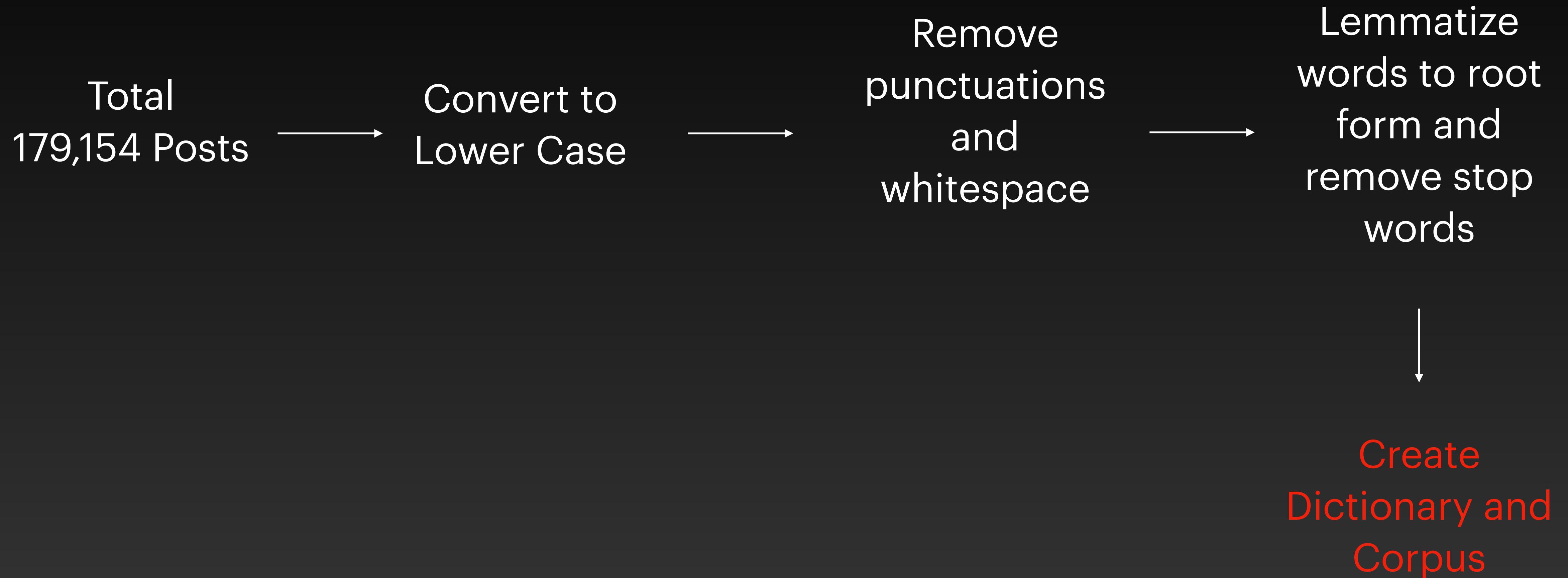
Total  
179,154 Posts



Convert to  
Lower Case





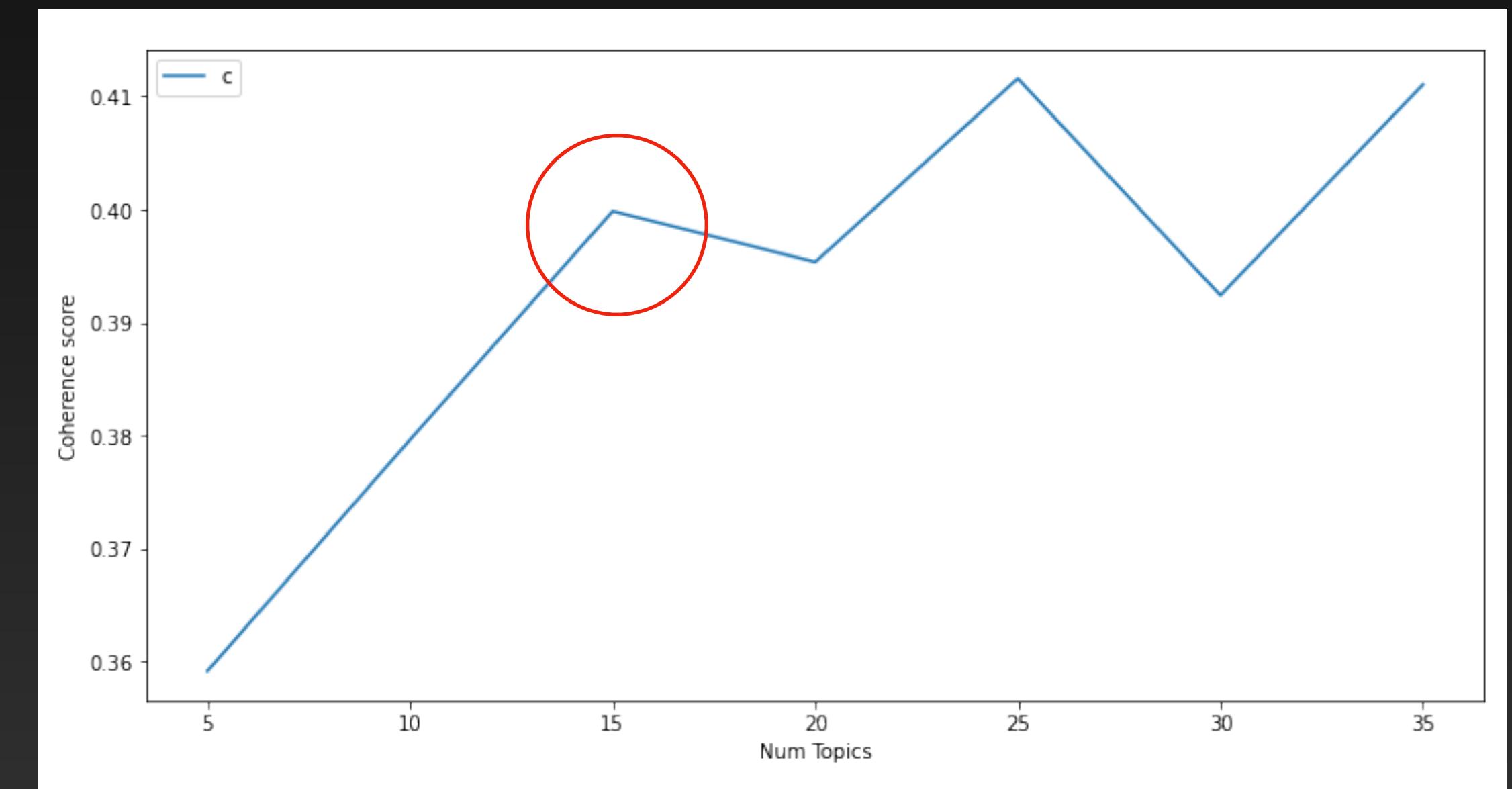


# **Step 2 : Topic Modelling and Interpretation**

# Latent Dirichlet Allocation (LDA)

- Using Coherence Score for optimal no of topics
  - Measure degree of semantic similarity between high scoring words in the topic

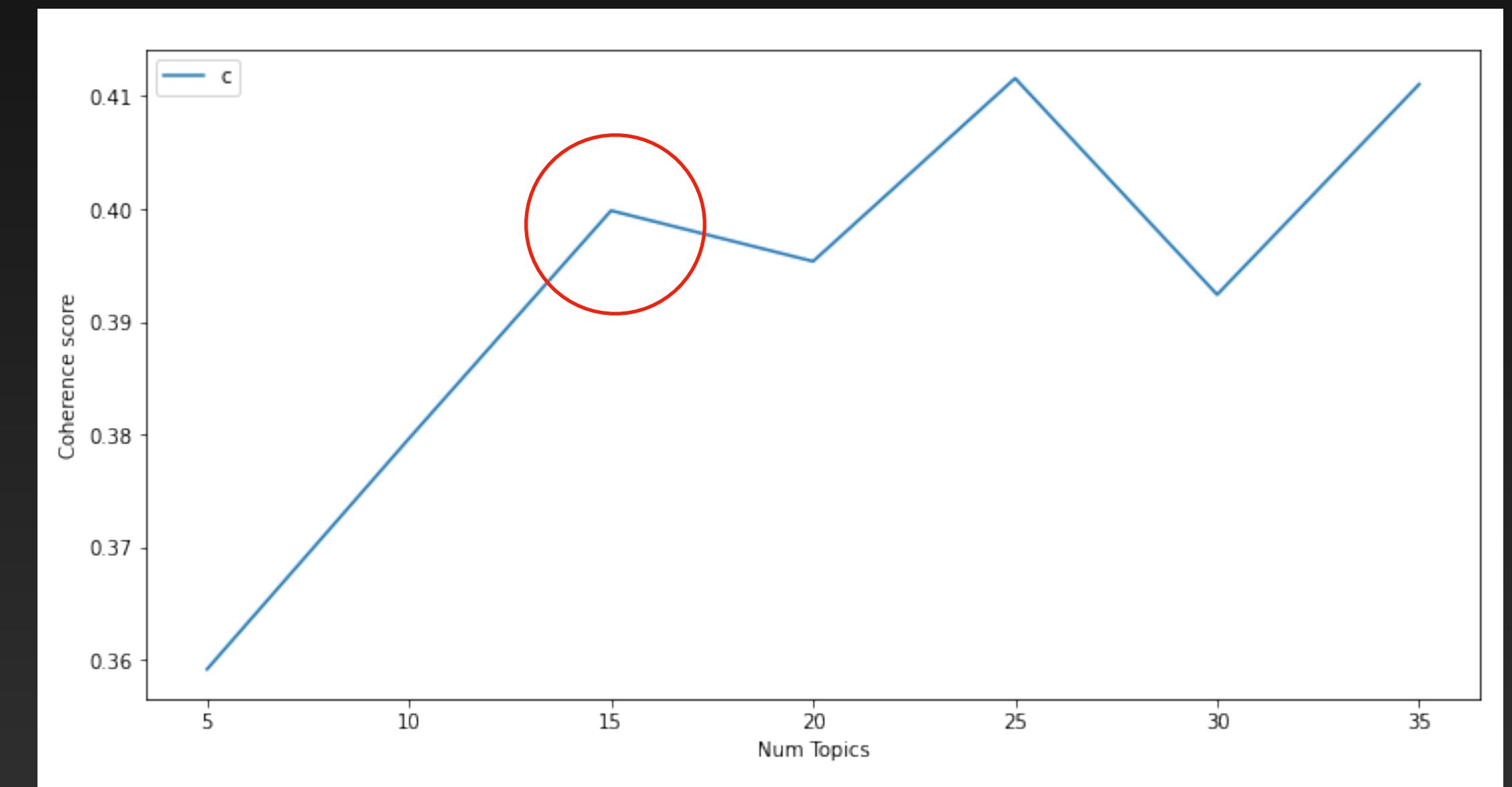
Figure 2: Coherence Score Over Num Topics



# Latent Dirichlet Allocation (LDA)

- Using Coherence Score for optimal no of topics
  - Measure degree of semantic similarity between high scoring words in the topic
- Final Coherence Score: 0.45 with 15 topics

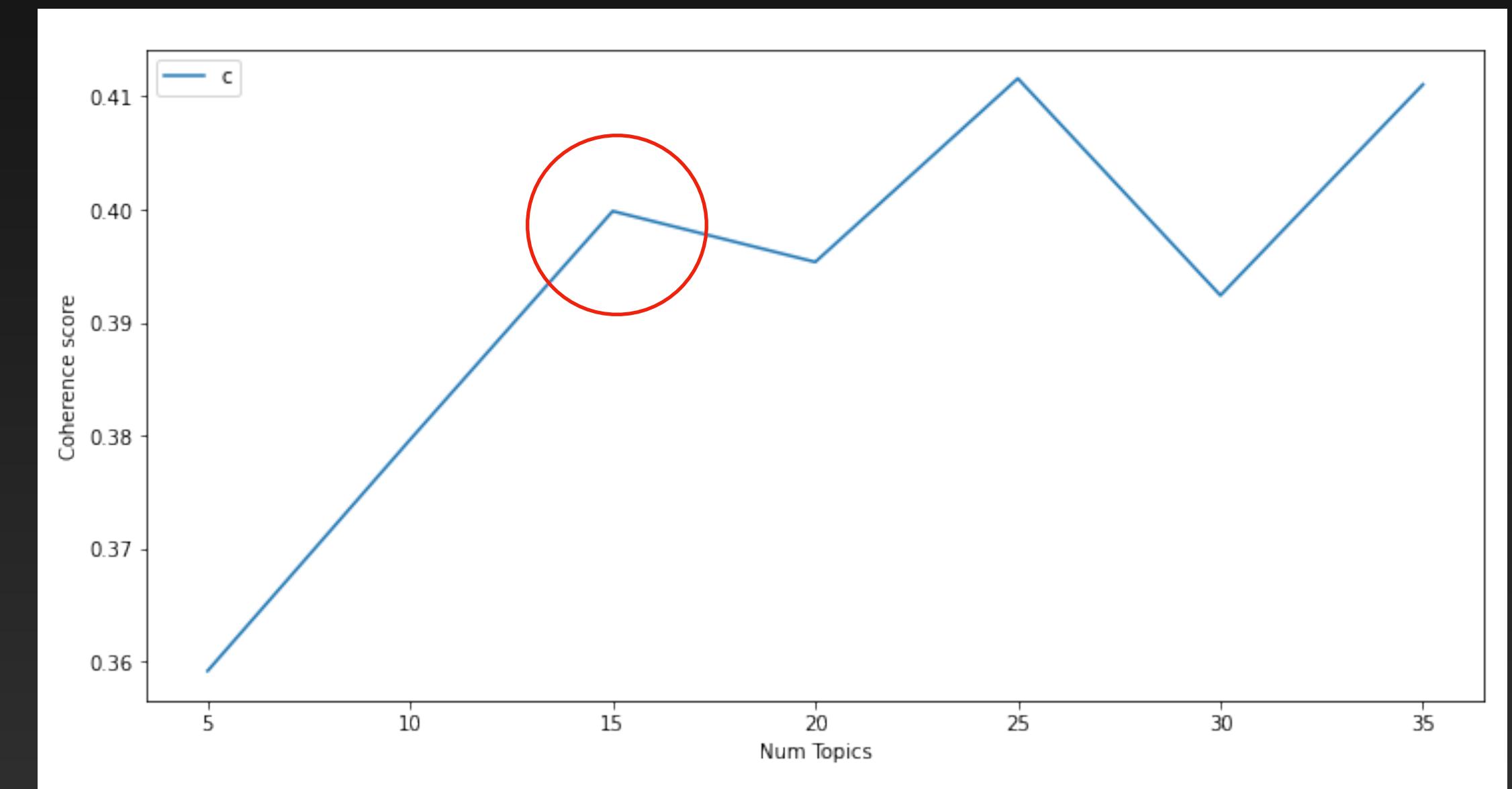
Figure 2: Coherence Score Over Num Topics



# Latent Dirichlet Allocation (LDA)

- Using Coherence Score for optimal no of topics
  - Measure degree of semantic similarity between high scoring words in the topic
- Final Coherence Score: 0.45 with 15 topics
- **Assign a dominant topic to each Reddit post**

Figure 2: Coherence Score Over Num Topics



**Figure 3: Assignment of Topics to Individual Reddit posts**

	subreddit	author	date	post	clean	Document_No	Dominant_Topic	Topic_Perc_Contrib	Keywords	Text
0	depression	hanngapoi	2018/01/01	school makes me suicidal (please help) im a 16...	school make suicidal please help year old girl...	0	9.0	0.6241	help, know, tell, need, want, think, talk, pos...	[school, make, suicidal, help, year, girl, eng...
1	depression	thr0w4w4y989897	2018/01/01	I give up. Failed I just want to wipe my hands...	give fail want wipe hand clean bare minimum re...	1	0.0	0.5127	life, want, live, know, think, kill, anymore, ...	[fail, want, wipe, hand, clean, bare, minimum,...
2	depression	mooseman1800	2018/01/01	depressed over the holidays No calls, messages...	depress holiday call message email holiday lif...	2	10.0	0.6203	friends, talk, friend, want, know, people, lov...	[depress, holiday, call, message, email, holid...
3	depression	throwaway1802360	2018/01/01	It's hard I don't want to go back to school th...	hard dont want back school week hard alone sch...	3	10.0	0.4170	friends, talk, friend, want, know, people, lov...	[hard, want, school, week, hard, school, hard,...
4	depression	atemylife	2018/01/01	Year in Pixels I'm going to be doing one of th...	year pixels one year pixels things track much ...	4	13.0	0.3861	school, college, class, high, year, grade, stu...	[year, pixels, go, year, pixels, things, track...

Topic 0:  
**Suicidal Ideation**

Topic 5:  
**Strong Negative Words**

Topic 10:  
**Relationships**

Topic 1:  
**Anxiety**

Topic 6:  
**Coping Mechanisms**

Topic 11:  
**Sleep**

Topic 2:  
**Common Words**

Topic 7:  
**Employment**

Topic 12:  
**Alcoholism**

Topic 3:  
**Mental Health Conditions**

Topic 8:  
**Family**

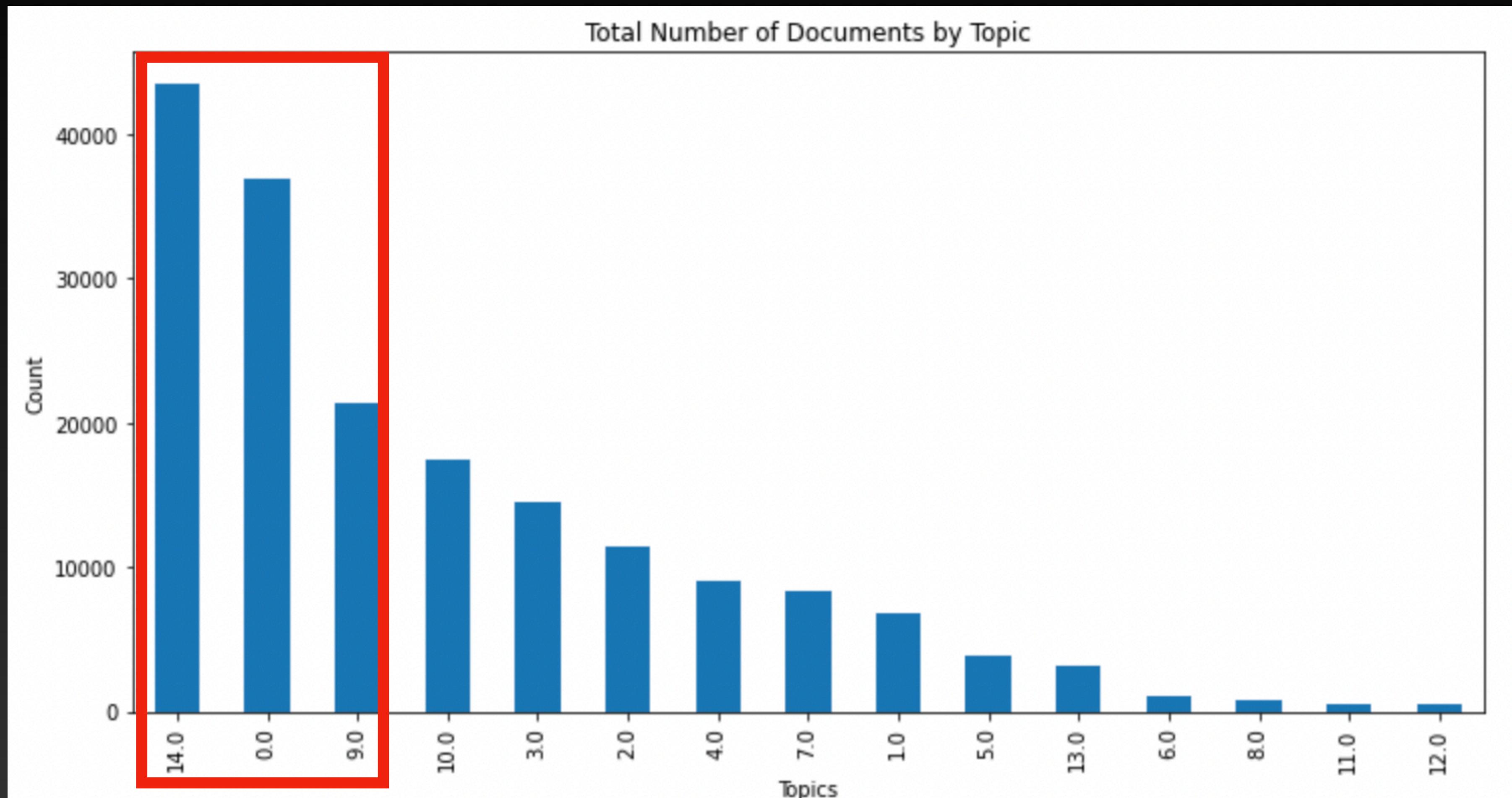
Topic 13:  
**Academics**

Topic 4:  
**Words used by persons with  
Schizophrenia & Bipolar Disorder**

Topic 9:  
**Seeking Advice & Help**

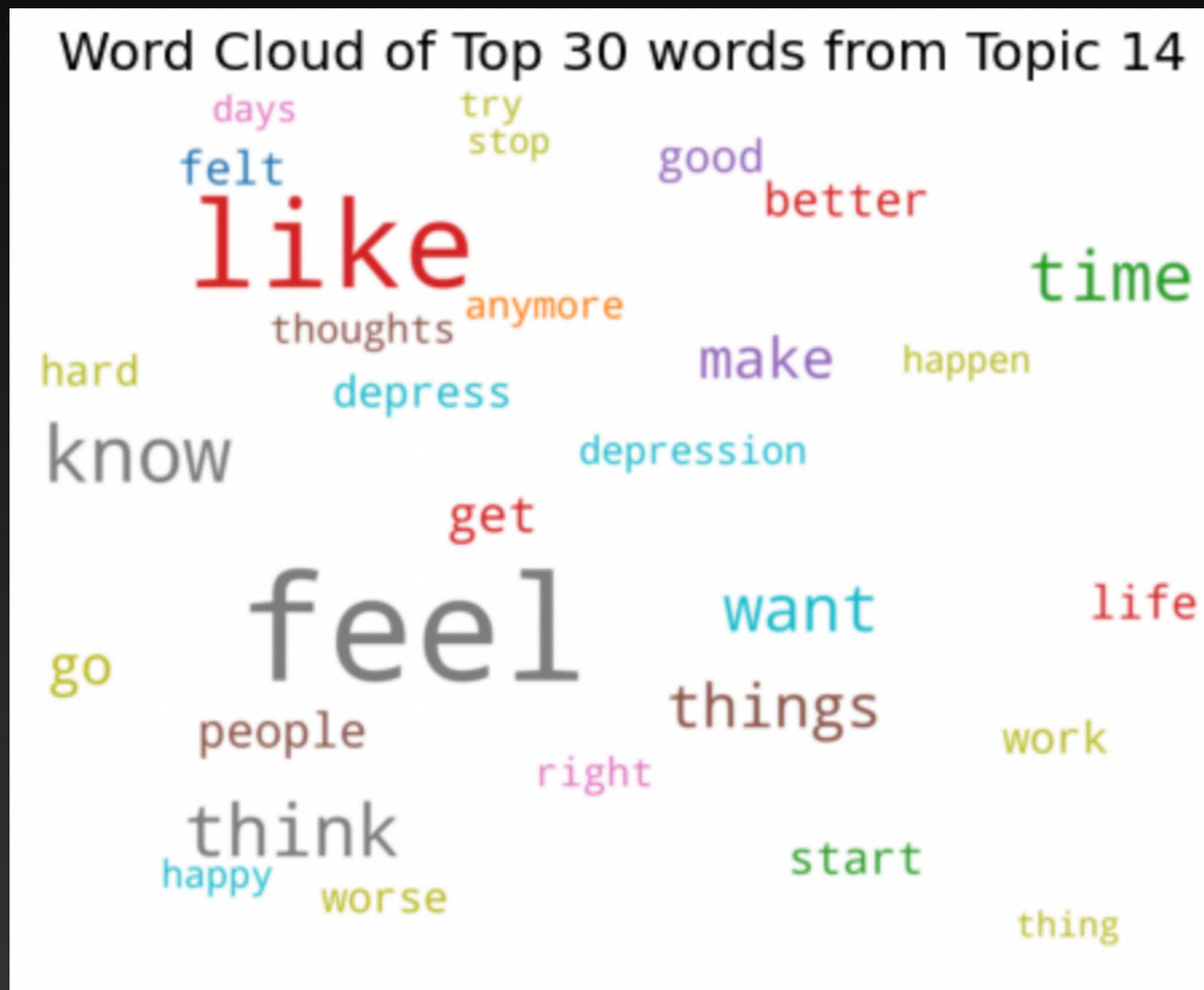
Topic 14:  
**Feelings**

**Figure 4: Top Dominant Topics**



# Topic 14

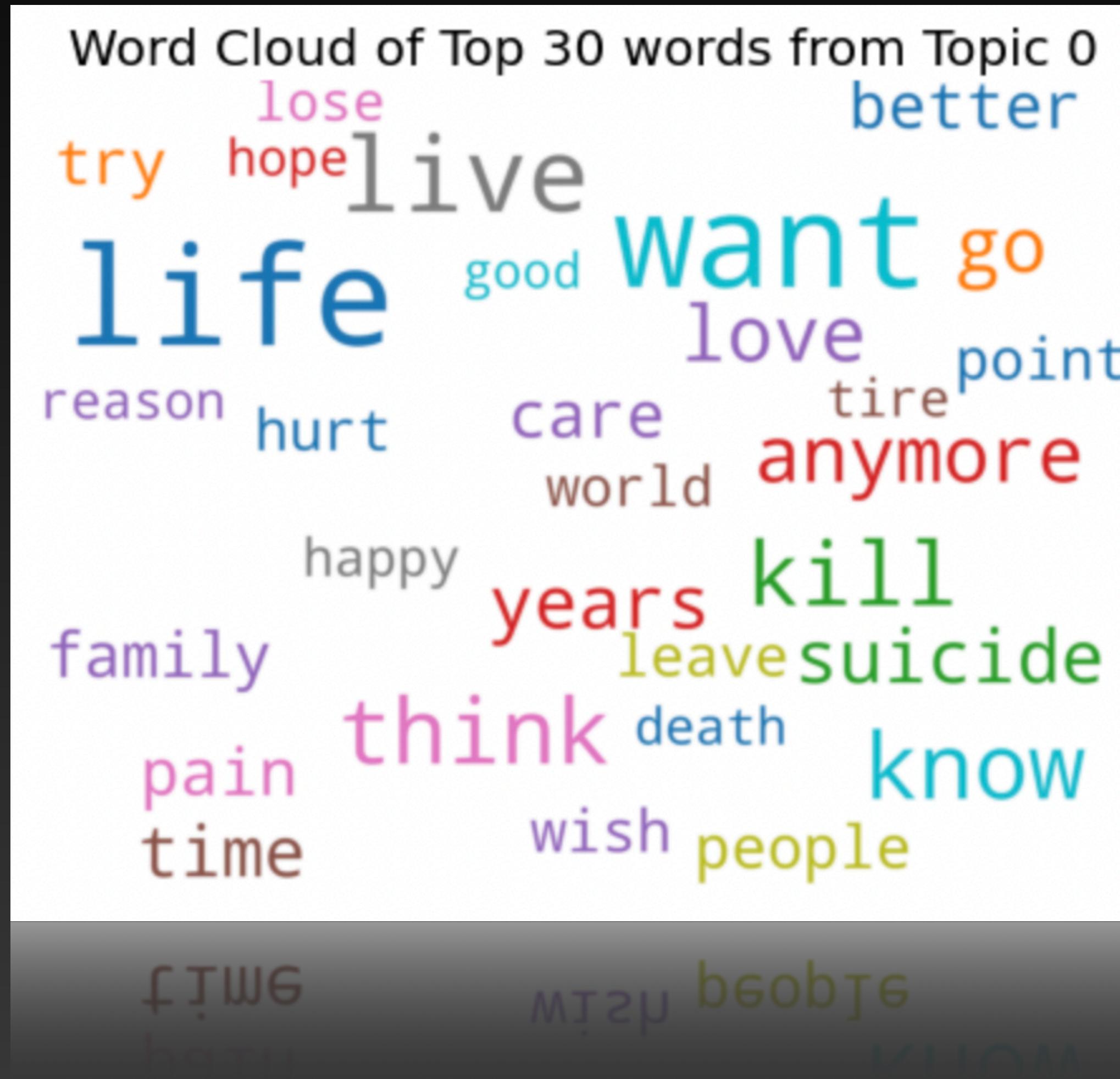
Figure 5: Topic 14 Word Cloud



Theme Assigned:  
**Feelings**

# Topic 0

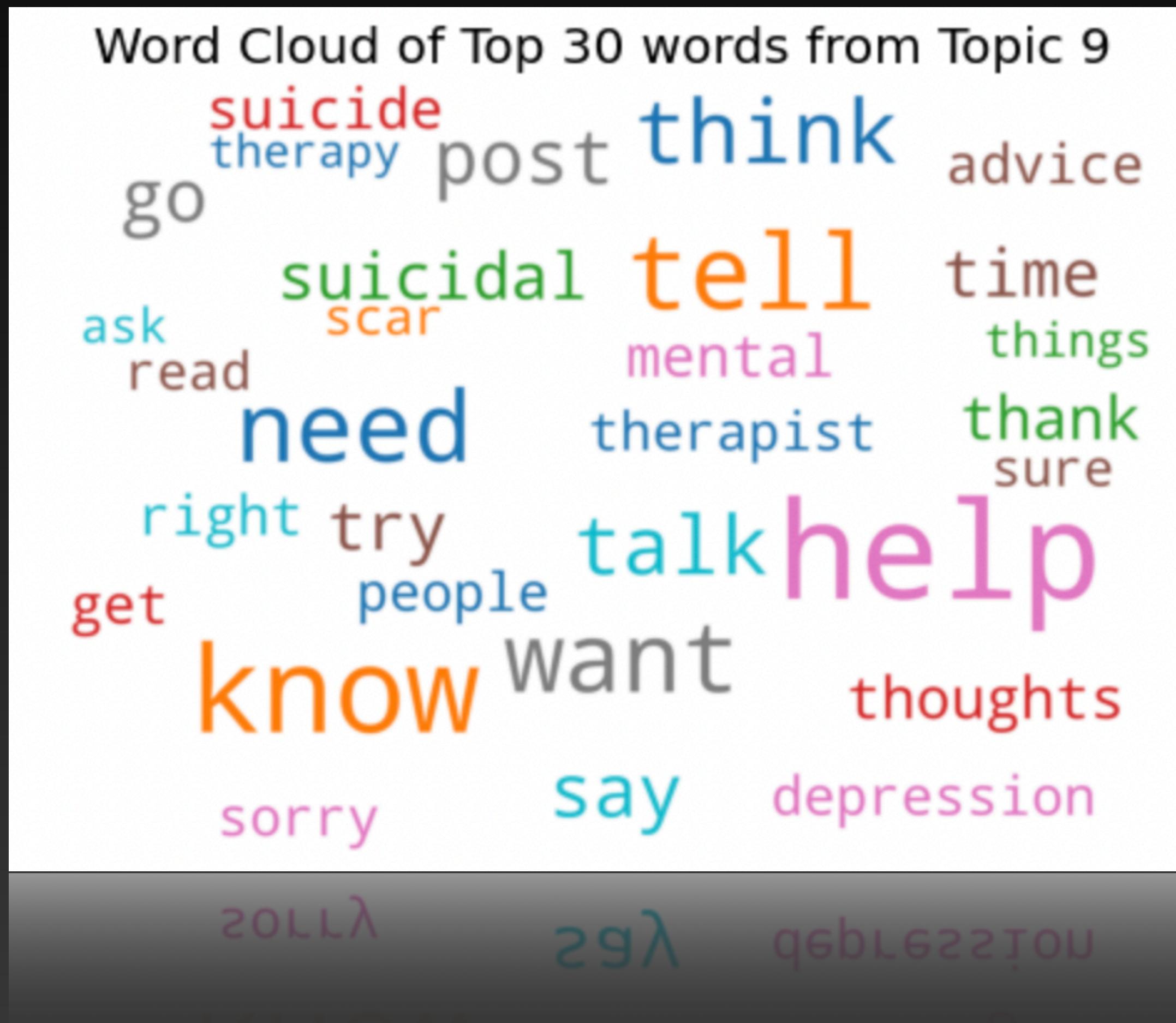
Figure 6: Topic 0 Word Cloud



Theme Assigned:  
**Suicidal  
Ideation**

# Topic 9

# Figure 7: Topic 9 Word Clouds

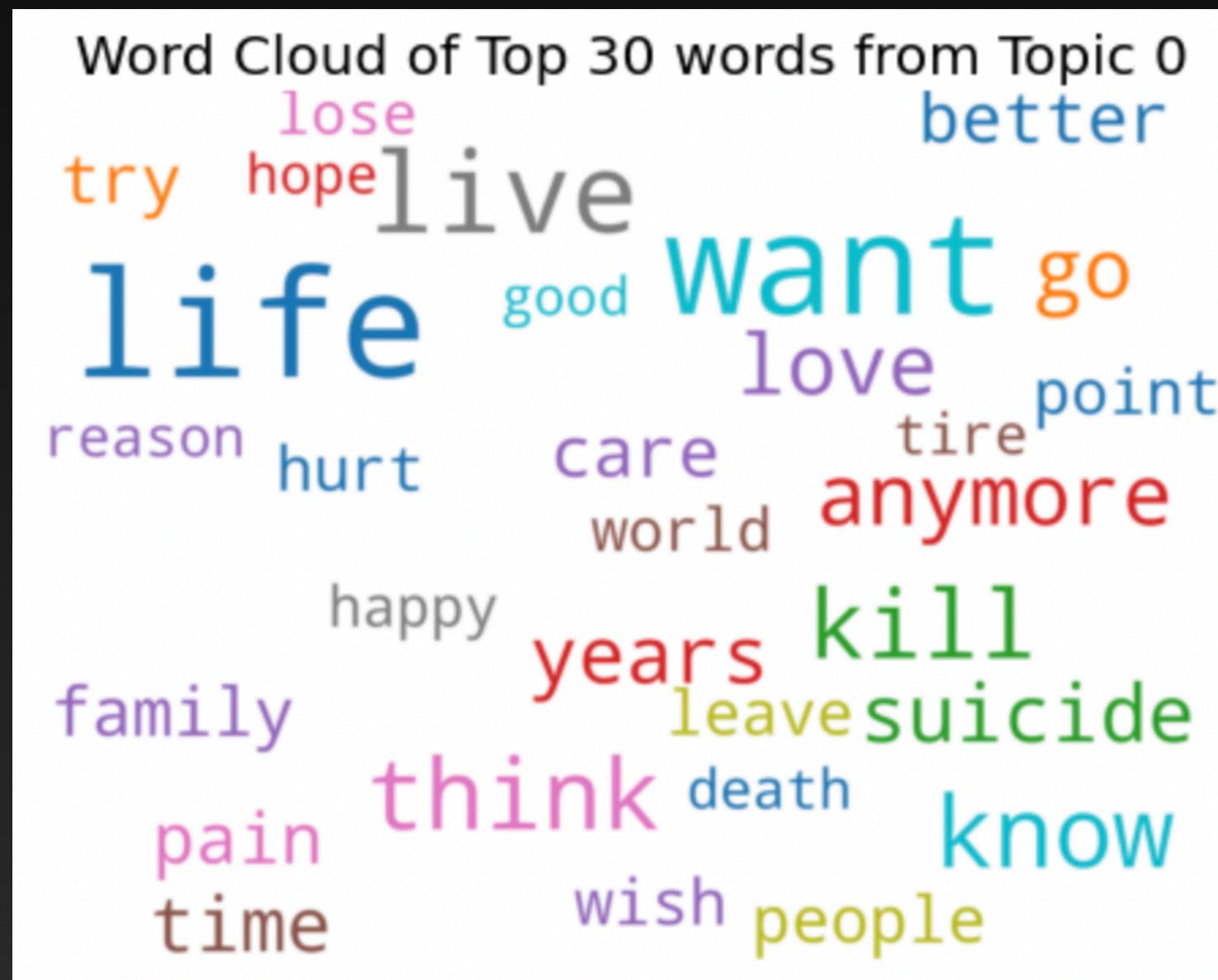


# Theme Assigned: **Seeking Advice**

# **Step 3: Manual Classification**

# Suicidal Posts

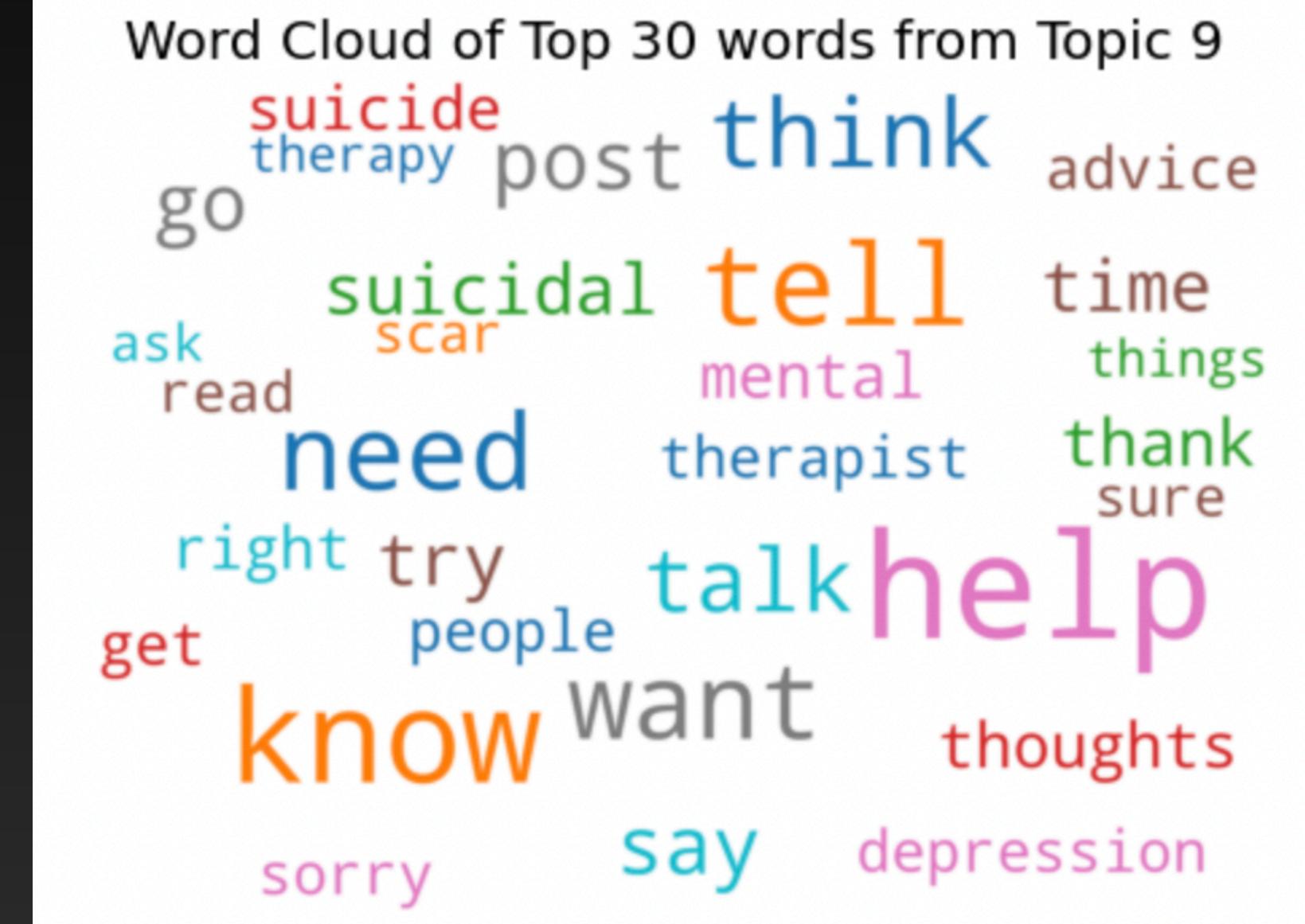
# Suicidal Ideation



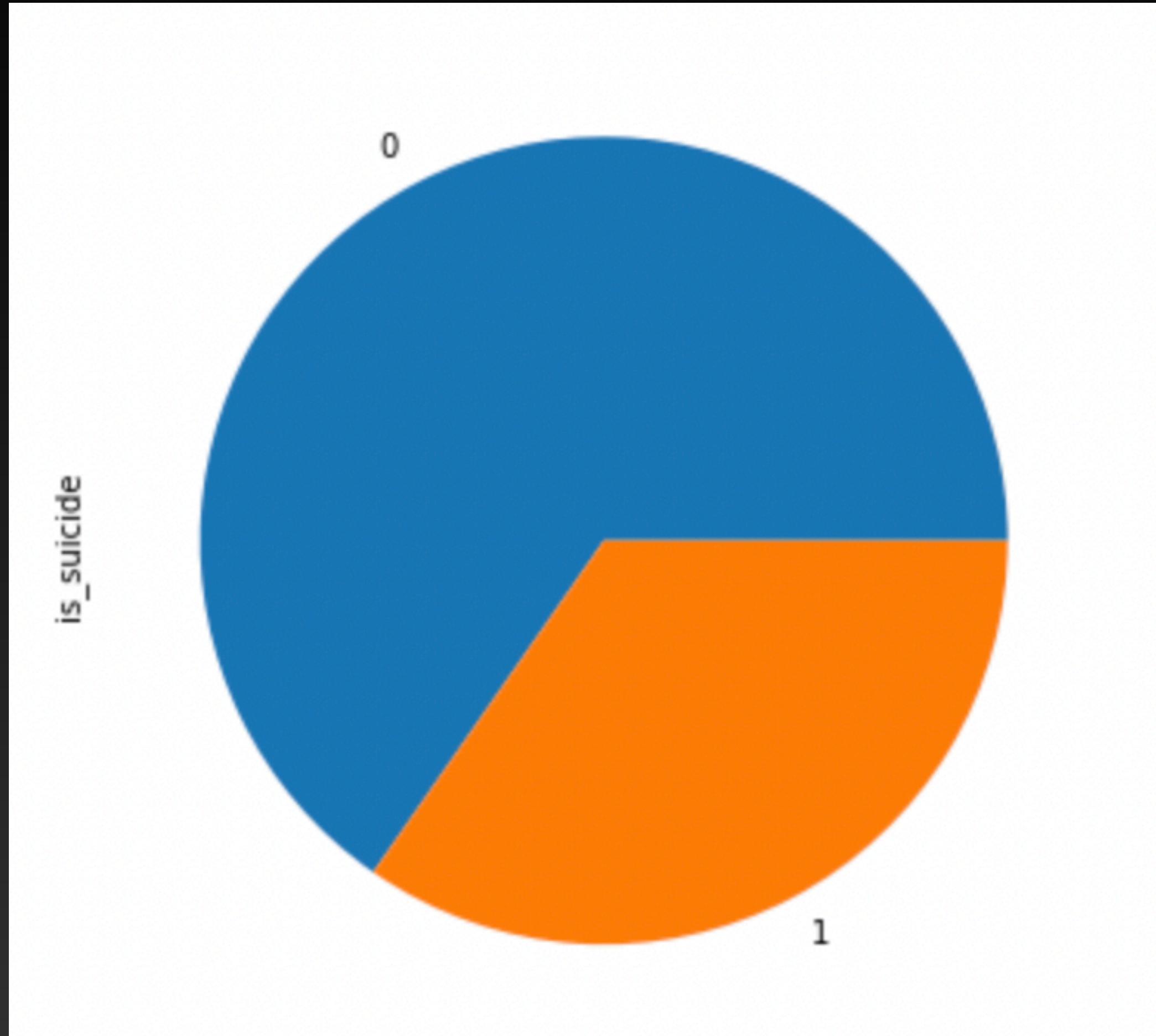
# Strong Negative Words



# Seeking Advice



**Figure 8: Breakdown of Suicidal Posts**



**Suicidal Posts (1)**

34.7%

**Non-Suicidal Posts (0)**

65.3%

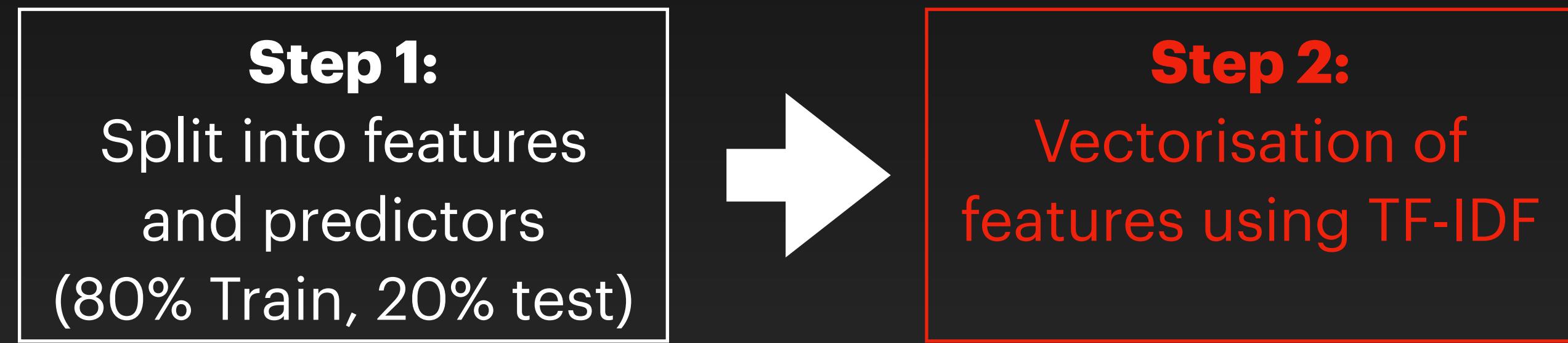
# Step 4: Predictive Modelling

# Modelling Process

## **Step 1:**

Split into features  
and predictors  
(80% Train, 20% test)

# Modelling Process



# Modelling Process



# Summary of Models

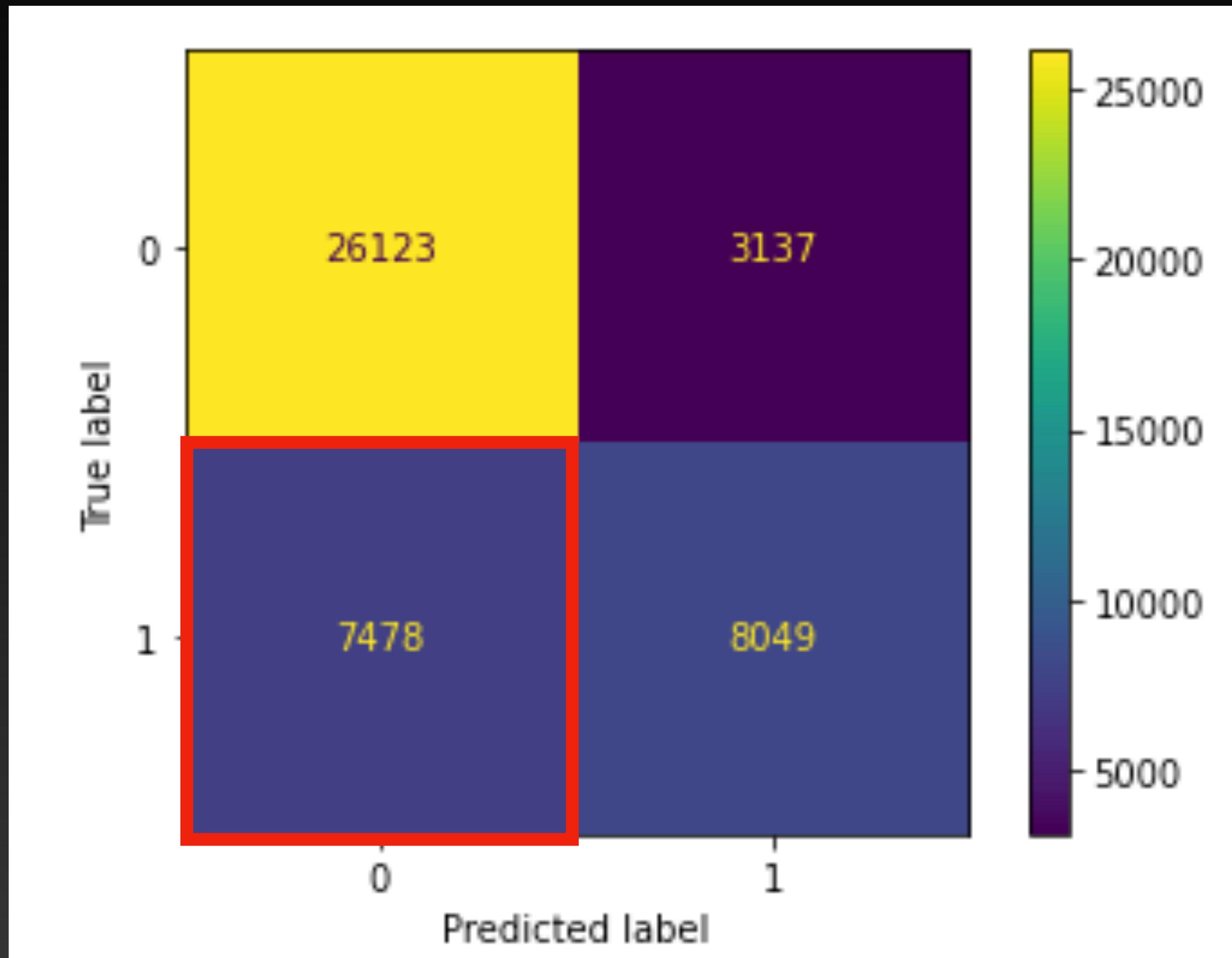
Machine Learning Algorithm	Train Accuracy	Test Accuracy	Precision	Recall
Multinomial Naive Bayes	70%	69%	72%	70%
Logistic Regression	73%	73%	72%	73%
KNeighbors	79%	70%	69%	70%
Random Forest	96%	76%	75%	76%

# Summary of Models

Machine Learning Algorithm	Train Accuracy	Test Accuracy	Precision	Recall
Multinomial Naive Bayes	70%	69%	72%	70%
Logistic Regression	73%	73%	72%	73%
KNeighbors	79%	70%	69%	70%
Random Forest	96%	76%	75%	76%

The model is **76% accurate** in correctly classifying between suicidal posts and non-suicidal posts

# Evaluation of Final Model



**1 = Suicidal Posts**

**0 = Non Suicidal Post**

**Key Concern:**  
Too much False  
Negatives

# Value proposition

**\$1.7 million/person**

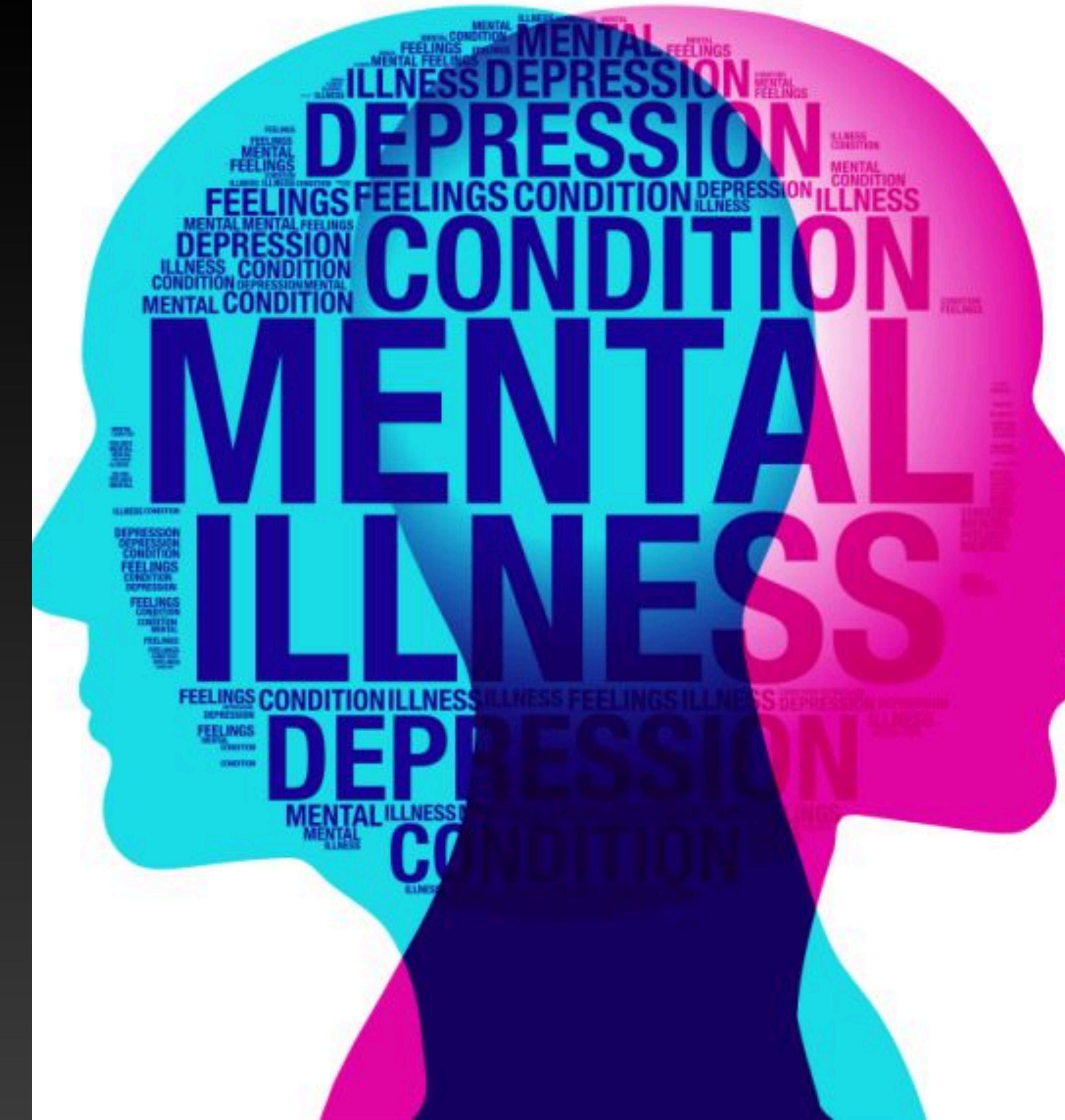
**Estimated Average Cost of Youth Suicide in Singapore (Lost Earnings)**

**“There is an economical impact in suicide prevention and that we need to leverage on social media data as part of our collective efforts to address suicide in Singapore”**

# Conclusion

# Summary

- Model is able to predict suicidal posts at **76% accuracy**
- Need to Address the False Negatives
- Consider other predictors (User's multiple posts)
- Deep Learning techniques to improve prediction



# Reference

Doran CM, Kinchin I (2020) Economic and epidemiological impact of youth suicide in countries with the highest human development index. PLoS ONE 15(5): e0232940. <https://doi.org/10.1371/journal.pone.0232940>

Low, D. M., Rumker, L., Torous, J., Cecchi, G., Ghosh, S. S., & Talkar, T. (2020). Natural Language Processing Reveals Vulnerable Mental Health Support Groups and Heightened Health Anxiety on Reddit During COVID-19: Observational Study. *Journal of medical Internet research*, 22(10), e22635.

Shepard DS, Gurewich D, Lwin AK, Reed GA Jr, Silverman MM. Suicide and Suicidal Attempts in the United States: Costs and Policy Implications. *Suicide Life Threat Behav.* 2016 Jun;46(3):352-62. doi: 10.1111/sltb.12225. Epub 2015 Oct 29. PMID: 26511788; PMCID: PMC5061092.

Roy, A., Nikolitch, K., McGinn, R. et al. A machine learning approach predicts future risk to suicidal ideation from social media data. *npj Digit. Med.* **3**, 78 (2020). <https://doi.org/10.1038/s41746-020-0287-6>