Forecast Depression Level and Risk of Suicide

Team: QWQ

Group members: Amber Wang, Caroline Xu, Zimo Wu, Wanqing Li

Mentors: Farnoosh & Shilpa

Content

- Introduction
- Method
- Result
- Conclusion & Discussion

Content

- **Introduction**
- Method
- Result
- Conclusion & Discussion

Introduction - Why this topic?

Depression is a common illness worldwide which can lead to suicide

By WHO 2021, over **700 000 die** due to suicide, suicide is the **4th leading cause** of death in **15-29-year-olds**

People experiencing depression tend to express their feelings on social networks

Depression

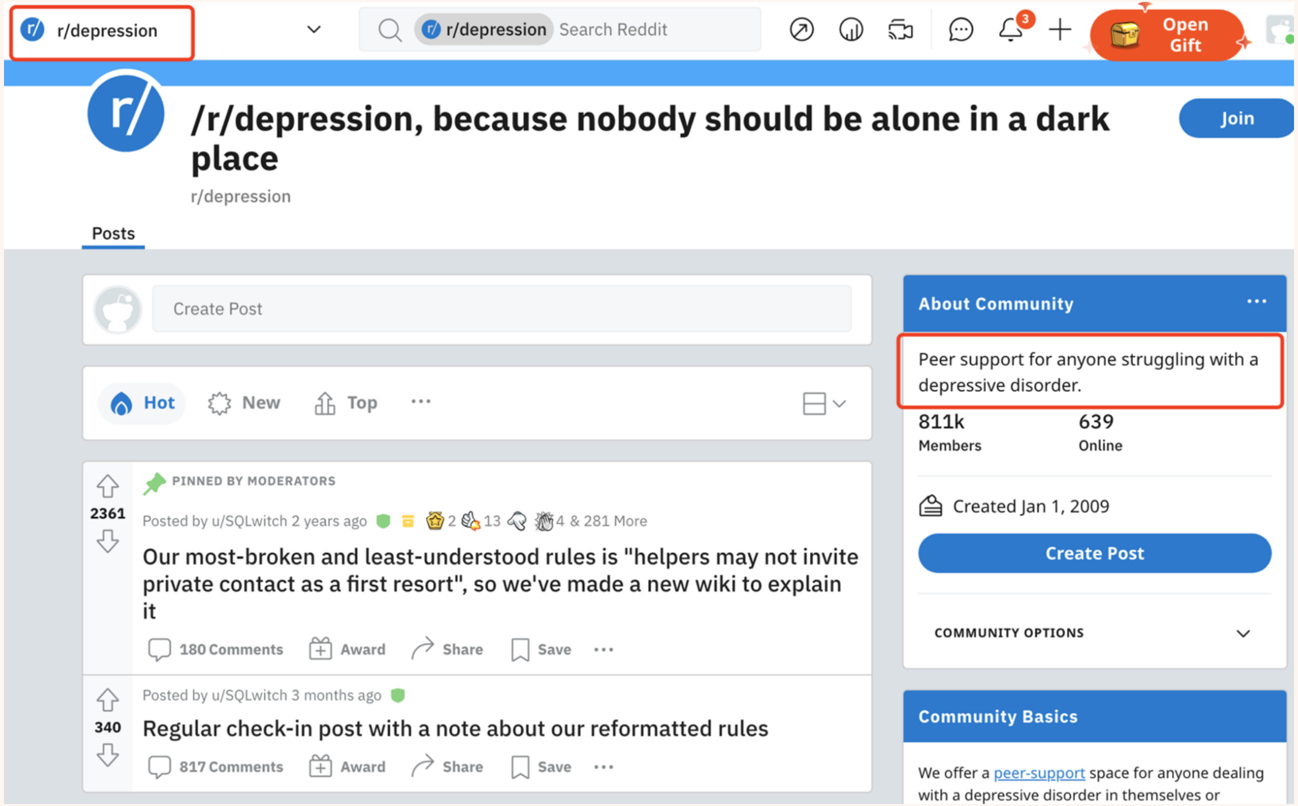
Suicide

Social Media

forecasting the **depression level** and **risk of suicide** through **analysing social media posts**

Introduction - Dataset

500 Redditors' Posts with 5-label Depression Classification, postes are from reddit r/depression



Introduction - Dataset

A **list of posts** sent by a particular user

	User	Post	Label
0	user-0	['Its not a viable option, and youll be leavin...	Supportive
1	user-1	['It can be hard to appreciate the notion that...	Ideation
2	user-2	['Hi, so last night i was sitting on the ledge...	Behavior
3	user-3	['I tried to kill my self once and failed badl...	Attempt
4	user-4	['Hi NEM3030. What sorts of things do you enjo...	Ideation
...	...	...	...
495	user-495	['Its not the end, it just feels that way. Or ...	Supportive
496	user-496	['It was a skype call, but she ended it and Ve...	Indicator
497	user-497	['That sounds really weird.Maybe you were Dist...	Supportive
498	user-498	['Dont know there as dumb as it sounds I feel ...	Attempt
499	user-499	['>It gets better, trust me.Ive spent long ...	Behavior

500 rows × 3 columns

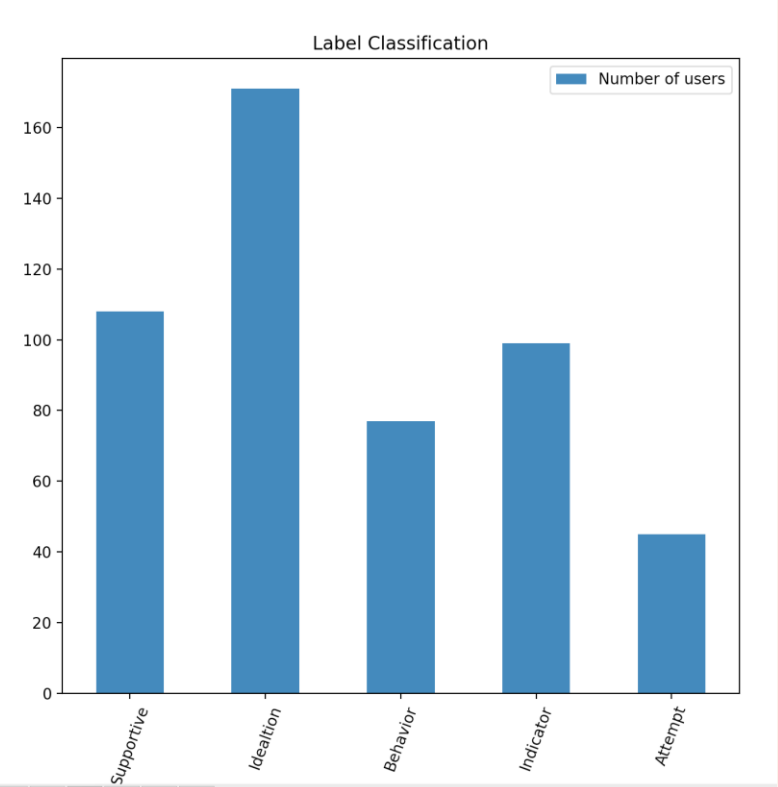
User-Index, from 0 to 499, refer to users have discussed suicide

Labels **developed manually by exports** following the guidelines outlined in Columbia Suicide Severity Rating Scale (C-SSRS)

Introduction - Labels

C-SSRS-based 5-label Classification		
	Supportive	participating in discussion but not showing any sign of being at risk in the past or present
	Indicator	supportive but use at-risk language while sharing personal experience
	Ideation	has thoughts of suicide
	Behavior	having historical self-harm or planning to commit suicide
	Attempt	having deliberate action that may result in intentional death, like writing a public “good bye” note

Introduction - Description of Dataset



Introduction - Description of Dataset

Supportive

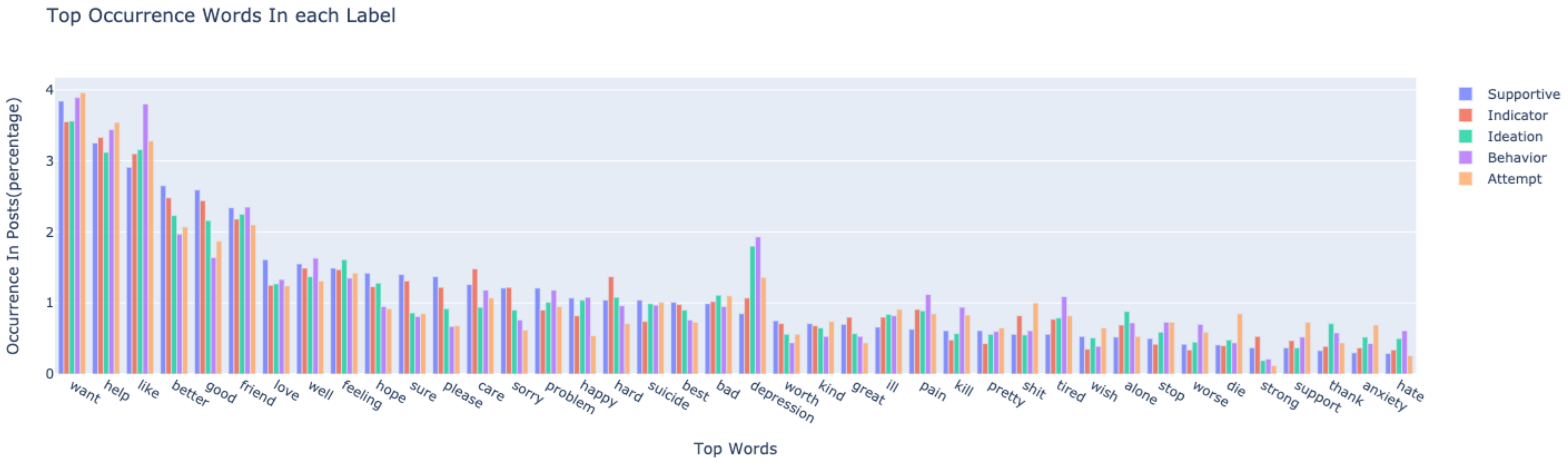
Indicator

Ideation

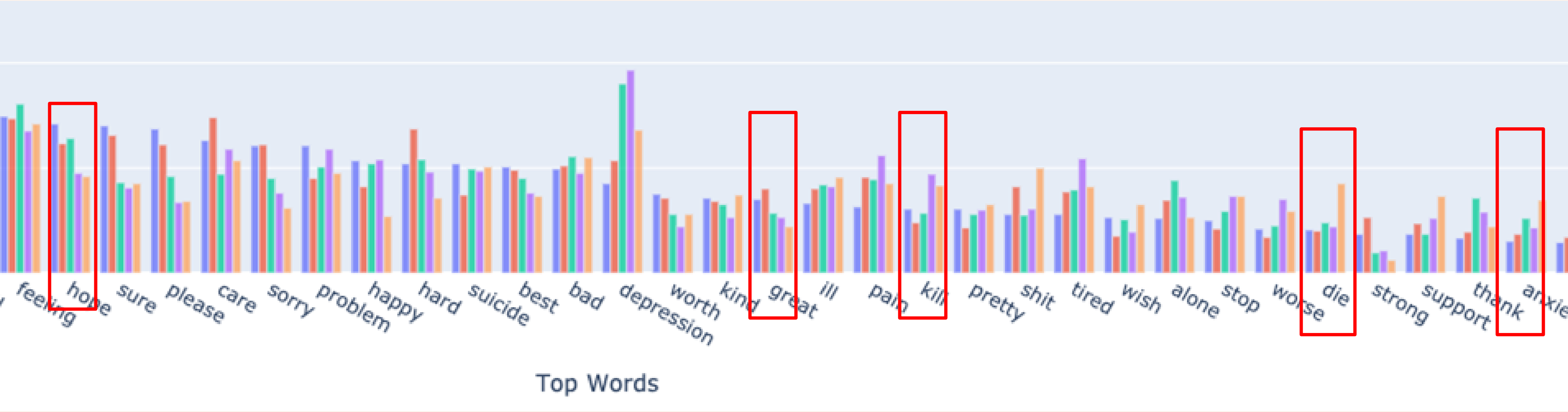
Behavior

Attempt

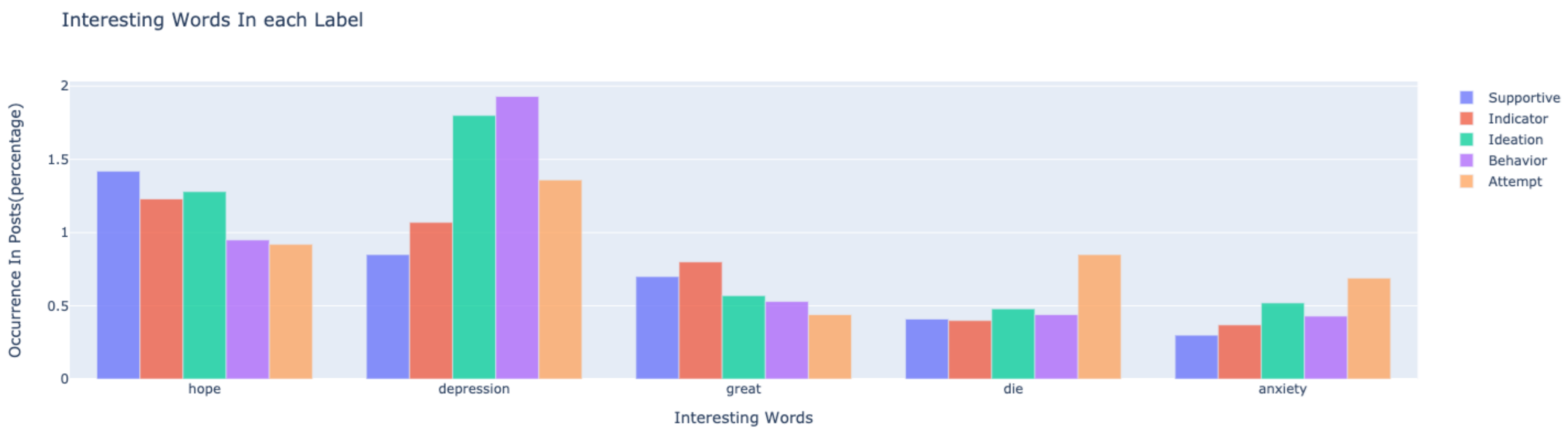
Introduction - Description of Dataset



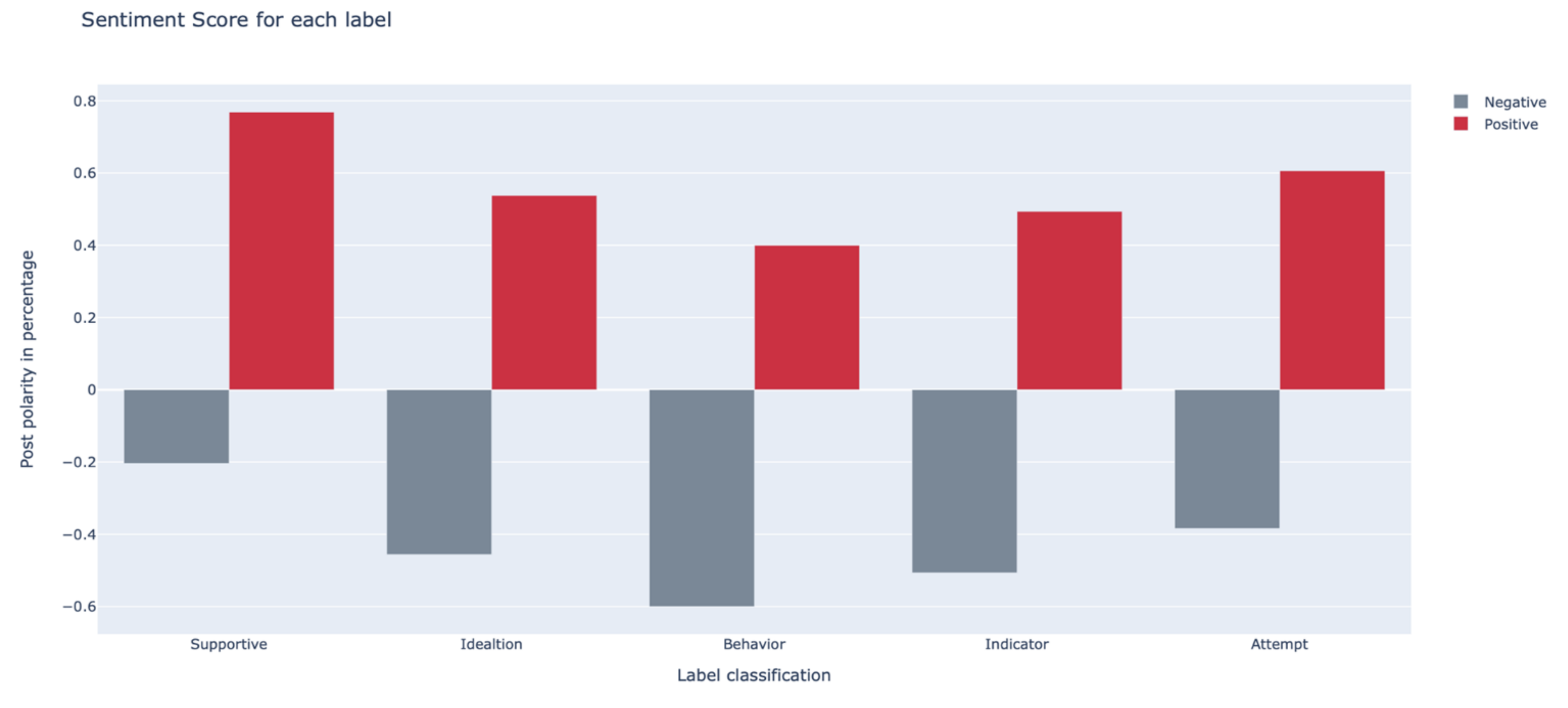
Introduction - Description of Dataset



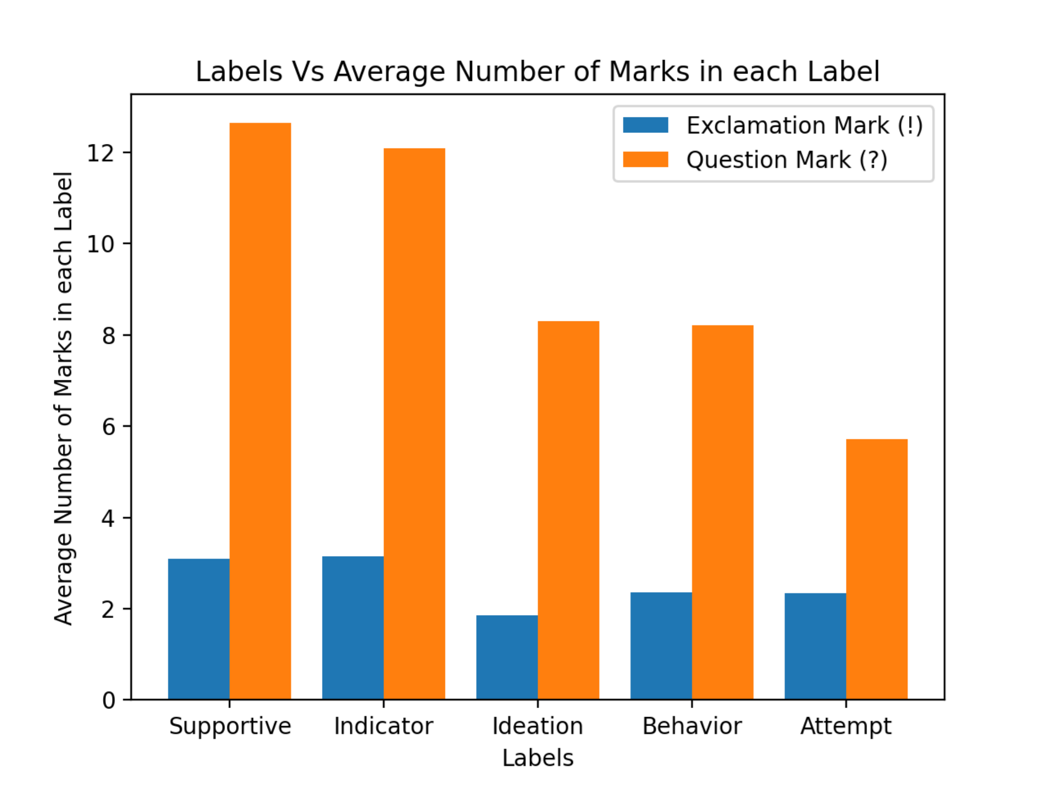
Introduction - Description of Dataset



Introduction - Description of Dataset



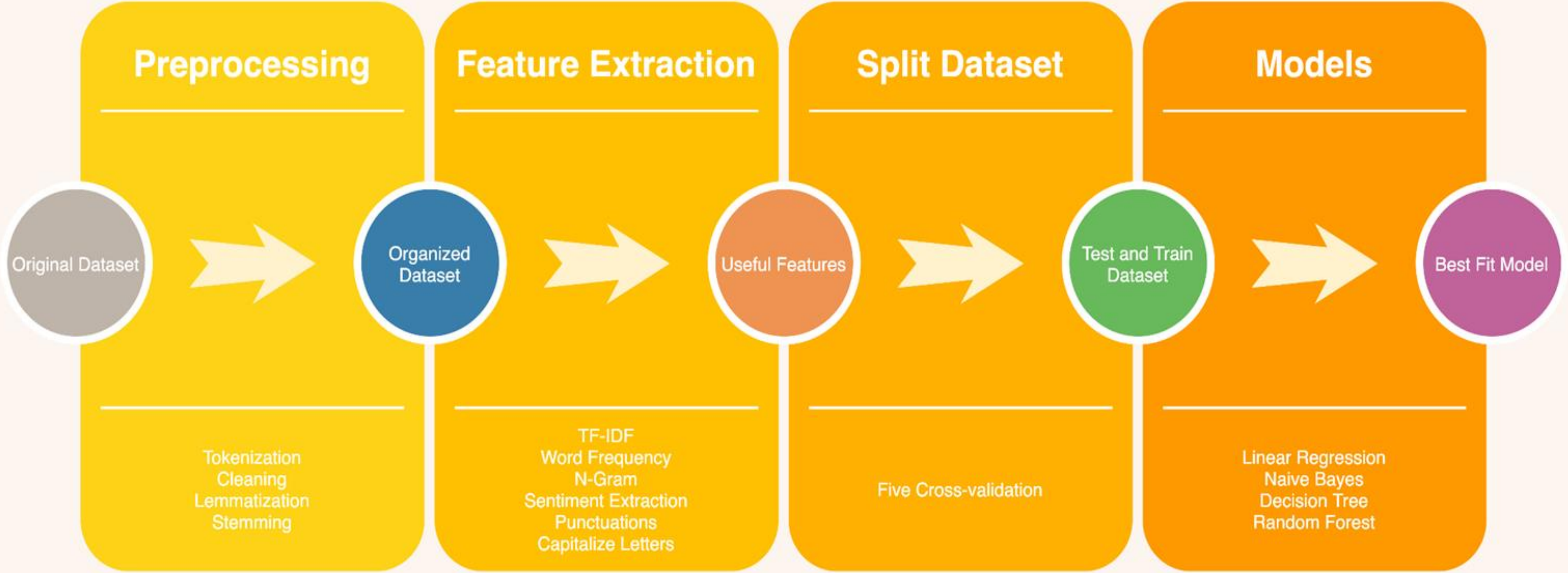
Introduction - Description of Dataset



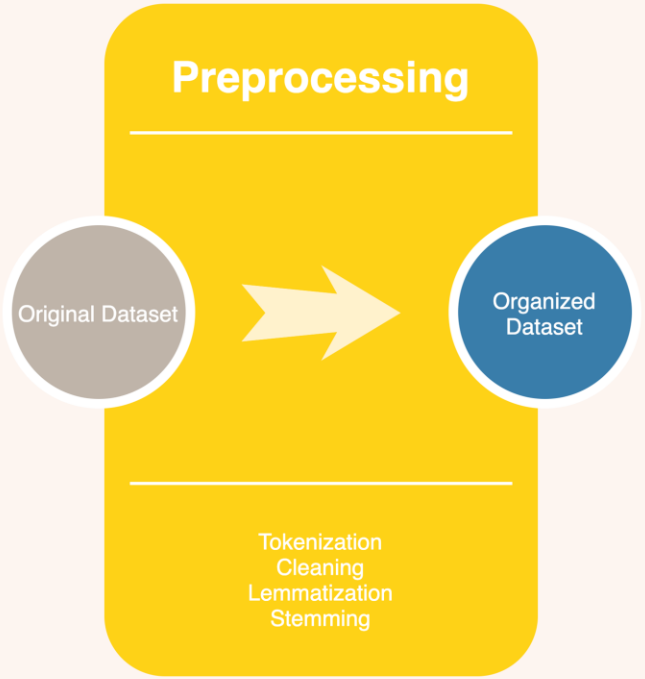
Content

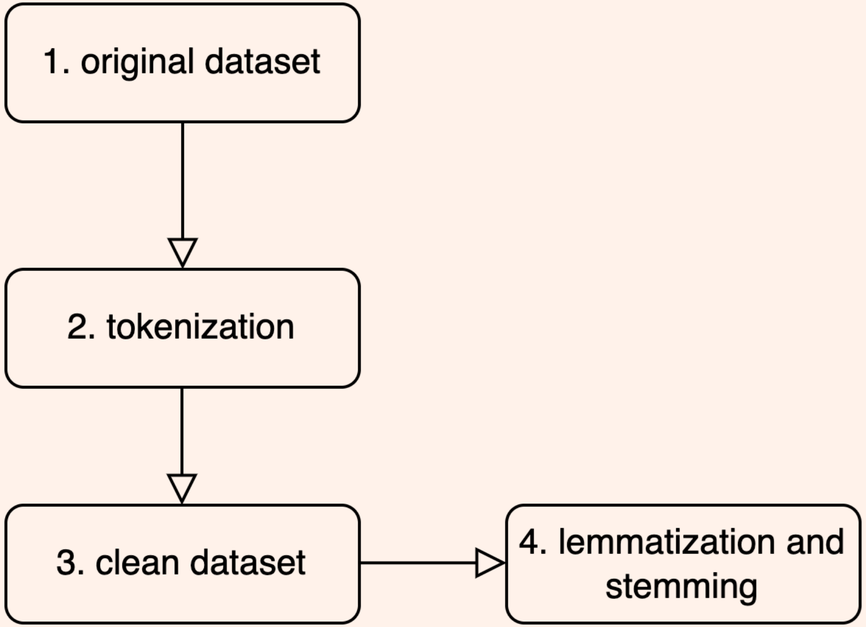
- Introduction
- **Method**
- Result
- Conclusion & Discussion

Method - Machine Learning

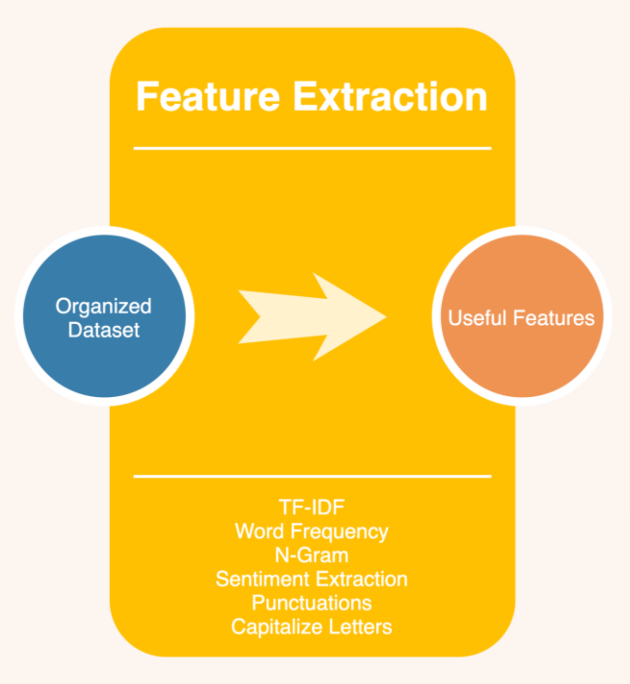


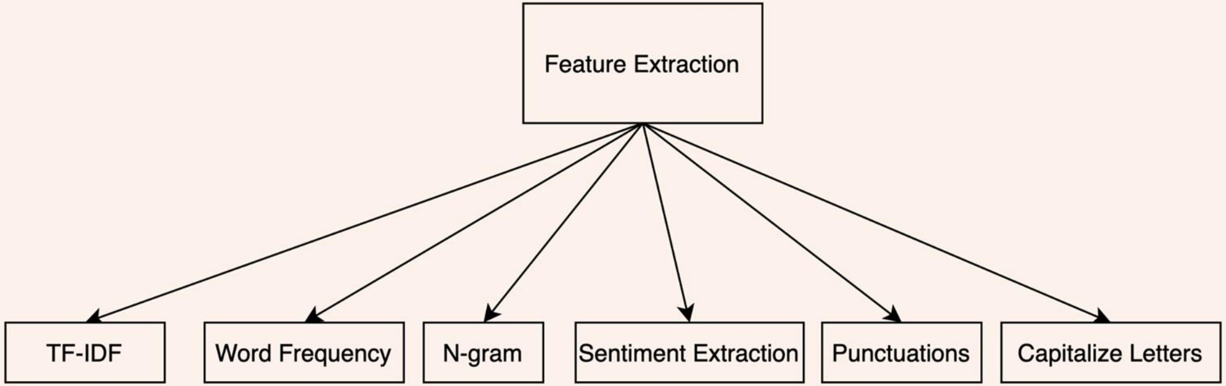
Method - Preprocessing the Dataset



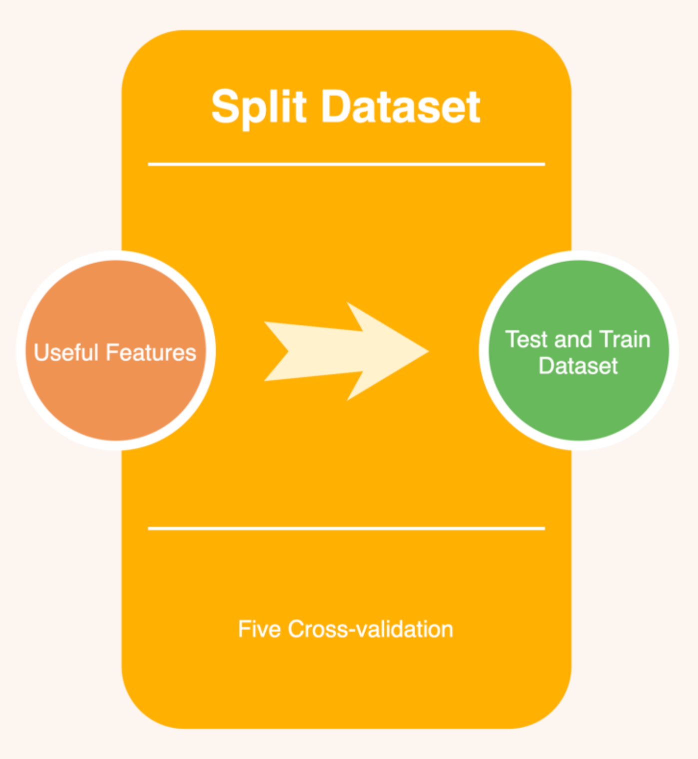


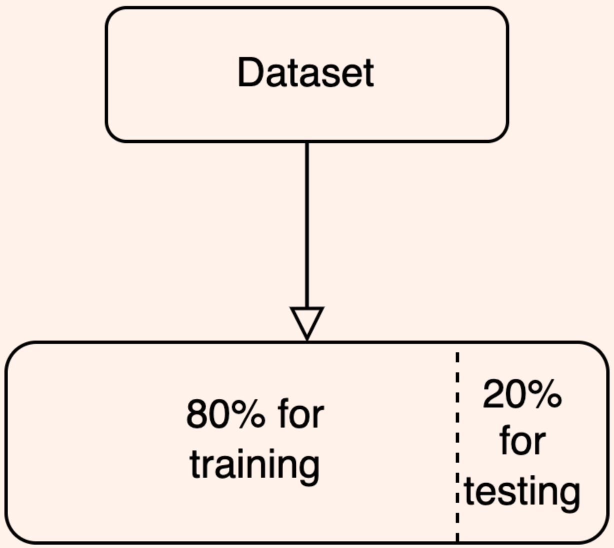
Method - Feature Extraction



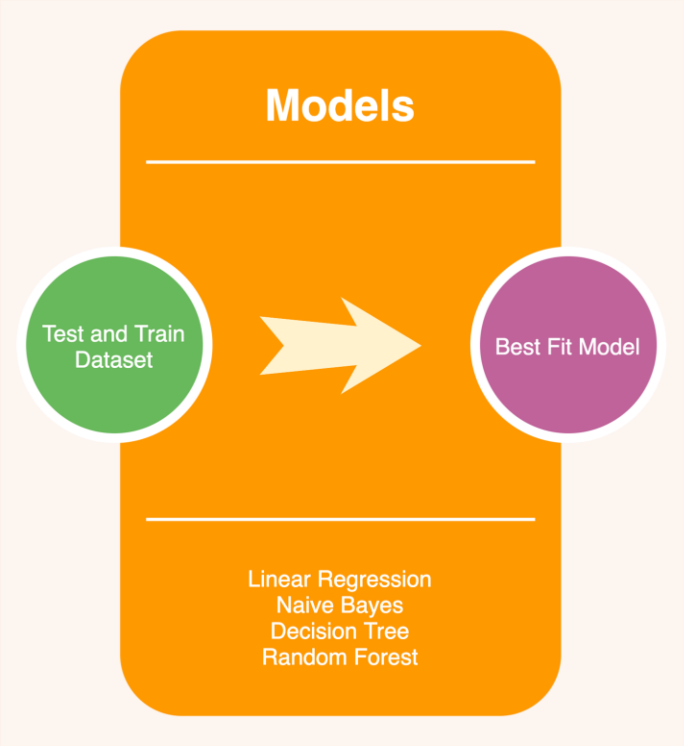


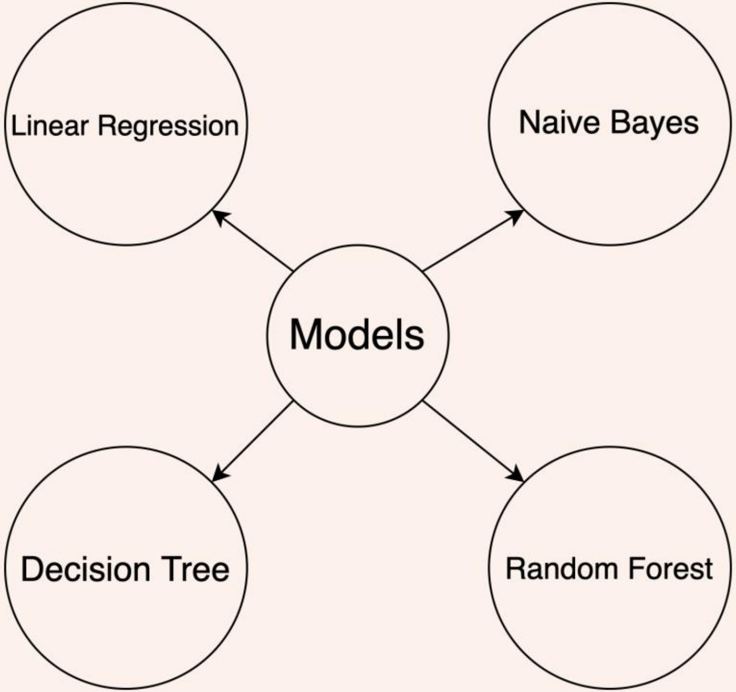
Method - Split Dataset





Method - Different Models





Method - Machine Learning

- Preprocessing the dataset: Tokenization -> Clean Data -> Lemmatization and Stemming
- Feature Extraction:
  - TF-IDF (term frequency-inverse document frequency)
  - Count Word Frequency
  - n-gram model
  - Sentiment Extraction (polarity of sentences)
  - Punctuations
  - Capitalize letters
- Split dataset: 80% training, 20% testing
- Algorithm to choose model:
  - Linear Regression
  - Naive Bayes
  - Decision Tree
  - Random Forest

Content

- Introduction
- Method
- **Result**
- Conclusion & Discussion

Results

00

Dummy classifier score:

0.23

01

PCA

Reduce dimension to 300

02

Max feature

Set to 20, 50, 100 in our model

03

Grid search

For random forest and decision tree

Table 1: using PCA **reduce dimension to 500**, no max feature, no grid search CV

Model Name	Accuracy
Linear regression	0.10
Naive Bayes	0.15
Random Forest	**0.31**
Decision Tree	0.25

Table 2: max feature=20, No PCA, random forest and decision tree used grid search CV

Model Name	Macro F1 Score	Micro F1 Score	Weighted F1 score	Best parameter
Linear regression(no grid)	0.09	0.14	0.09	/
Naive Bayes(no grid)	0.20	0.28	0.26	/
Random Forest	**0.21**	0.32	**0.27**	`criterion': 'gini', 'max_depth': 30, 'n_estimators': 38
Decision Tree	0.18	**0.33**	0.26	'criterion': 'entropy', 'max_leaf_nodes': 3, 'min_samples_split': 2

Table 3: max feature=50, No PCA, random forest and decision tree used grid search CV

Model Name	Macro F1 Score	Micro F1 Score	Weighted F1 score	Best parameter
Linear regression(no grid)	0.11	0.15	0.13	/
Naive Bayes(no grid)	0.21	0.25	0.26	/
Random Forest	**0.22**	**0.36**	**0.30**	'criterion': 'gini', 'max_depth': 5, 'n_estimators': 17
Decision Tree	0.19	0.33	0.25	'criterion': 'gini', 'max_leaf_nodes': 13, 'min_samples_split': 2

Table 4: max feature=100, No PCA, random forest and decision tree used grid search CV

Model Name	Macro F1 Score	Micro F1 Score	Weighted F1 score	Best parameter
Linear regression(no grid)	0.13	0.19	0.18	/
Naive Bayes(no grid)	0.21	0.24	0.25	/
Random Forest	**0.24**	**0.37**	**0.30**	'criterion': 'entropy', 'max_depth': 55, 'n_estimators': 31
Decision Tree	0.22	0.33	0.28	'criterion': 'gini', 'max_leaf_nodes': 4, 'min_samples_split': 2

Results

Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 189 (2021) 368–373

Procedia Computer Science

www.elsevier.com/locate/procedia

5th International Conference on AI in Computational Linguistics

Suicidal risk identification in social media

Ashok Kumar J[a,*], Tina Esther Trueman[a], Abinesh A K[b]

[a]Department of Information Science and Technology, Anna University, Chennai-600025, India
[b]Department of Journalism, Madras Christian College, Chennai-600059, India

Abstract

Social media influences people to express their mental health issues such as depression and anxiety. Specifically, depression is one of the biggest risk factors for suicidal ideation and attempts. Therefore, we propose a multiplicative attention-based bidirectional gated recurrent unit to identify the suicidal risk factors of social media users. The proposed model captures the local context in input sequences. Our experimental results indicate that the proposed model outperforms the state-of-the-art models in the multiclass classification task.

© 2021 The Authors. Published by Elsevier B.V.
This is an open access article under the CC BY-NC-ND license (https://creativecommons.org/licenses/by-nc-nd/4.0)
Peer-review under responsibility of the scientific committee of the 5th International Conference on AI in Computational Linguistics.

Keywords: Behavioral monitoring; suicidal ideation; deep learning; gated recurrent unit; attention mechanism.

1. Introduction

In modern society, suicide is rapidly increasing due to mental health problems such as anxiety and depression [1]. Anxiety is a normal feeling or an emotion where the brain reacts to the stress and alerts potential danger ahead. For instance, problems at work, making an important decision, and fear of an activity or a situation. Depression is associated with a feeling of sadness, loss of interest, or anger of an individual. In particular, the World Health Organization indicated that suicide is the second largest cause of death worldwide among teenagers [2, 3]. Nowadays, online social media influence individual users to express their feelings or emotions in the form of posts or comments [4, 5]. These personal feelings help us to identify suicidal risk factors, namely, ideation, indicator, behavior, attempt, and supportive [6]. First, the suicidal ideation category defines a suicidal thought of a user due to the loss of a strong relationship, loss of a job, mental illness, substance abuse, or chronic diseases. Second, the suicidal behavior involves actively planning to commit suicide, self-harm activity, using blunt force violence, or actions of death. Third, the attempt category is defined as a complete attempt, changed their mind, or writing a good-bye message. Fourth, suicidal indicator category involves at-risk language from acute symptoms, engagement in a supportive manner, history of divorce, chronic ill-

* Corresponding author: Ashok Kumar J
E-mail address: jashokkumar83@auist.net

1877-0509 © 2021 The Authors. Published by Elsevier B.V.
This is an open access article under the CC BY-NC-ND license (https://creativecommons.org/licenses/by-nc-nd/4.0)
Peer-review under responsibility of the scientific committee of the 5th International Conference on AI in Computational Linguistics.
10.1016/j.procs.2021.05.106

Results

Table 1. Model performance

Models	Macro F1	Micro F1	Weighted F1
NB	**0.1951**	0.2360	0.2169
LR	0.1850	0.2060	0.2081
SVM	0.1640	0.2720	0.2288
LSTM_Attn	0.1261	0.2700	0.1797
BiLSTM_Attn	0.1410	0.2960	0.1968
GRU_Attn	0.1633	0.2920	0.2236
BiGRU_Attn	0.1661	0.2960	0.2203
LSTM_Mattn	0.1608	0.2980	0.2187
BiLSTM_Mattn	0.1534	0.2873	0.2106
GRU_Mattn	0.1601	0.2930	0.2197
BiGRU_Mattn	0.1914	**0.3000**	**0.2437**

Content

- Introduction
- Method
- Result
- **Conclusion**

Conclusion - Accomplishment

1. Data preprocessing

Select the dataset, clean the dataset, stemming & lemmatization

2. Feature extraction

TFIDF, N-gram, Sentiment, strong emotions

3. Train the 4 model

Split the data to train and test

4.Evaluate & Finalize model

Choose one model from 4 model

Conclusion - Blockers & Solutions

Condition 03

Reduce dimension

PCA

Max Feature

Condition 02

Accuracy

Initial evaluate method is wrong

Use F1 score and other methods

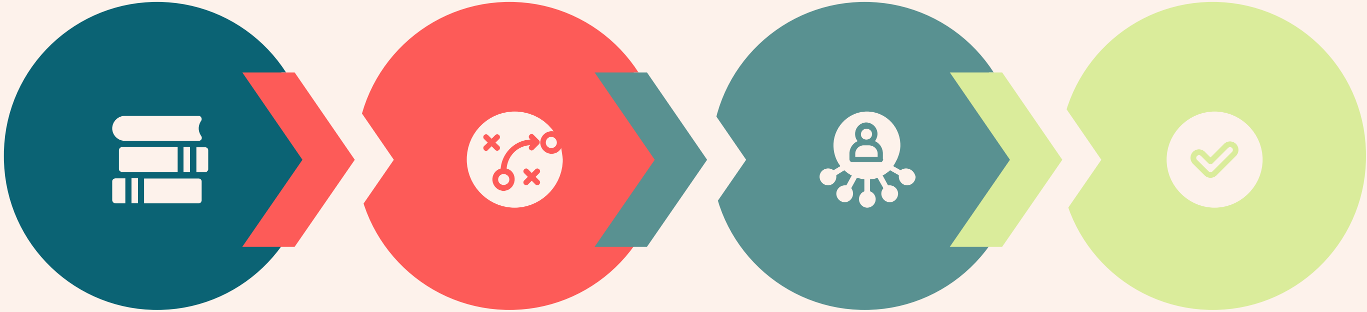
Condition 01

Feature

Not enough features

Brainstorm together

Conclusion - Experience



Educate

Learn knowledge and theories from school

Explore

Explore more information like coding on Internet

Engage

Engage in project, Combine coding & knowledge together

Execute

Get a final model from machine learning

Conclusion - Future

Support diagnose

1. Increase awareness

User could test their posts by themselves, and get suggestion

2. Support doctor

Doctor could use this model to make more structured decision

Conclusion

Thank you!

Q&A

Any Question?