

The ReJeX

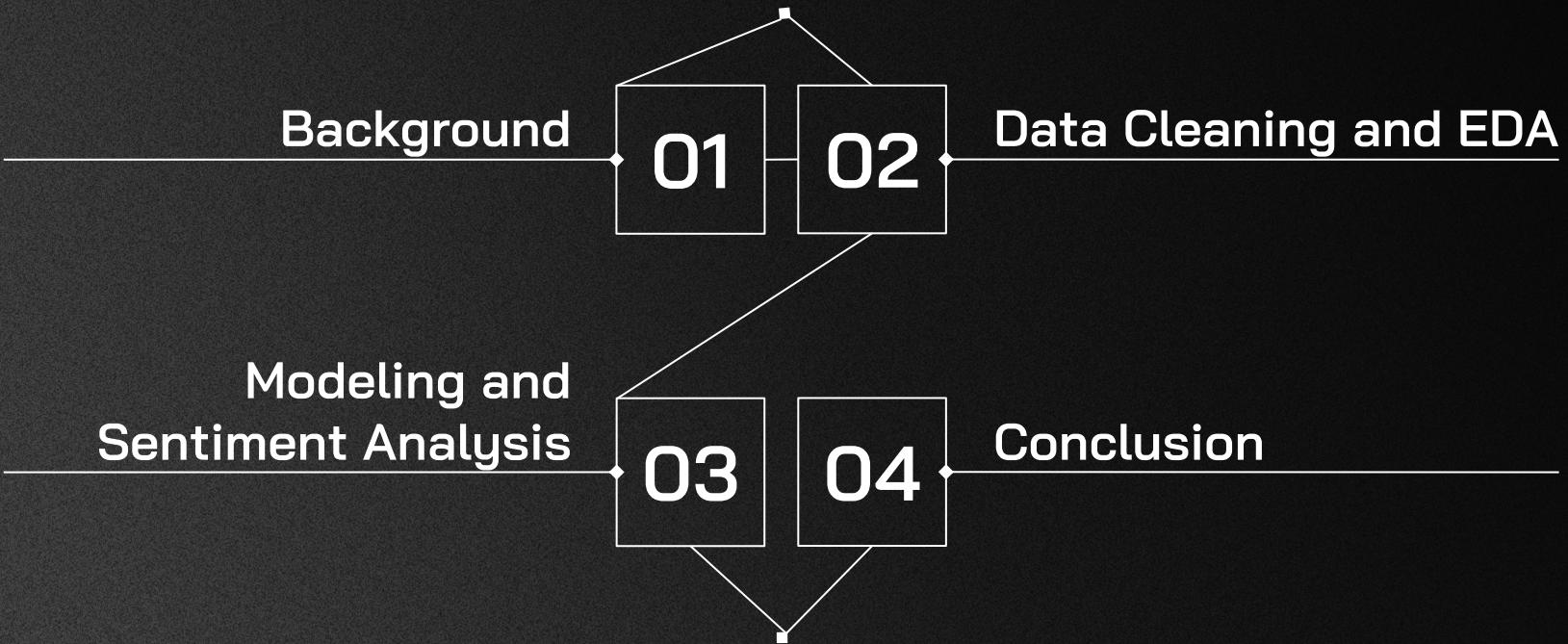


Subsidiary of RoJAK Data Consultancy

Richelle-Joy Chia, Jonathan Soon, Xue Feng Fung



Table of contents

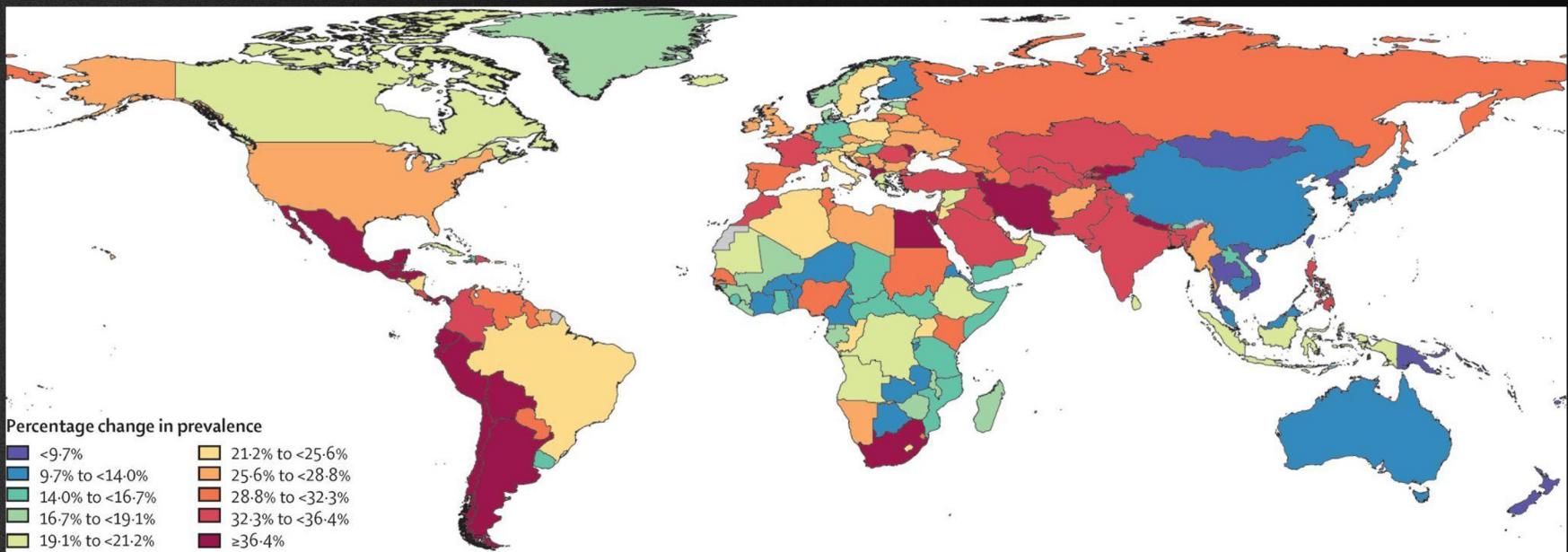


01

Background



**1 in every 8 people in the world live with a mental disorder,
with depression and anxiety being the top 2**



Source: [https://www.thelancet.com/journals/lancet/article/P11S0140-6736\(21\)02143-7/fulltext](https://www.thelancet.com/journals/lancet/article/P11S0140-6736(21)02143-7/fulltext)

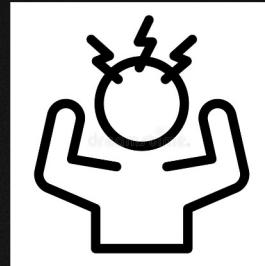
Definitions



Depression

Persistent feelings of sadness and loss of interest

Ranked 1st in the world



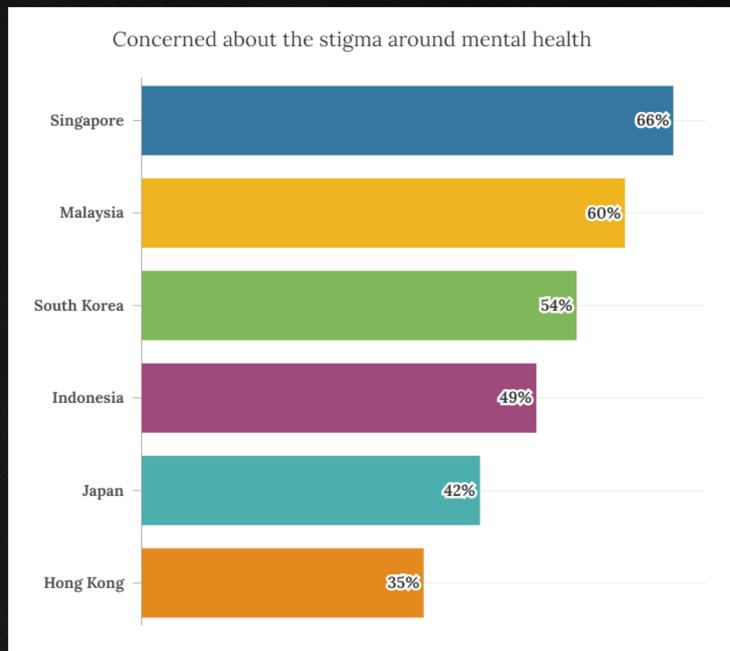
Anxiety

Persistent feelings of uneasiness and worry

Ranked 2nd in the world

Why should we care?

- Depression and anxiety are psychiatric disorders that are observed in many areas of everyday life
 - These disorders manifest somewhat frequently in texts written by non diagnosed users on social media
- Social stigma caused people to not speak out and turn to other means (eg. social platforms)



Source:
<https://www.channelnewsasia.com/interactives/mental-health-survey-2965916>

Problem statement

- How can we correctly classify posts based on the texts used by people who may be depressed or anxious?
- How can sentiment analysis be utilized to detect certain root emotions?



02

Data Cleaning and EDA



Data Scraping



(23rd August 2022 - 3rd October 2022)

Approximately 30,000 posts

pushshift/api

Pushshift API



3

Contributors

90

Issues

1k

Stars

97

Forks



r/Anxiety

Community • 573k members



r/depression

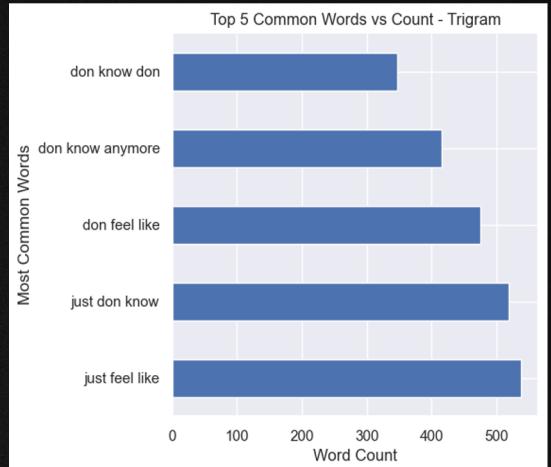
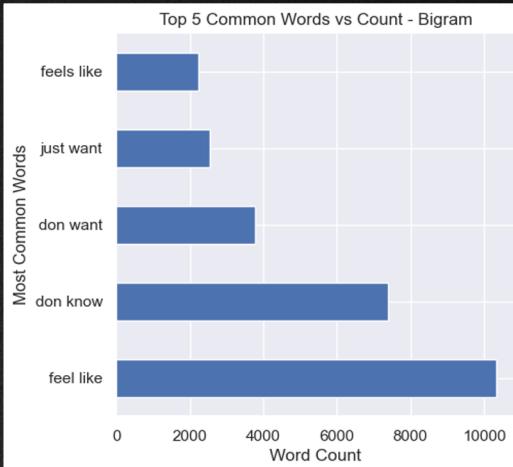
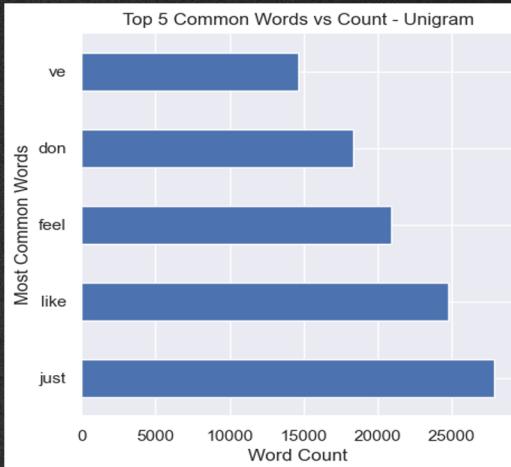
Community • 907k members

Preprocessing

Regex {characters only}

Removing duplicate rows

Remove stop words
[Deleted], [Removed], Anxiety, Anxious,
Depression, Depressed



Vectorizing

Regex



Removes symbols
Removes digits

Tokenization



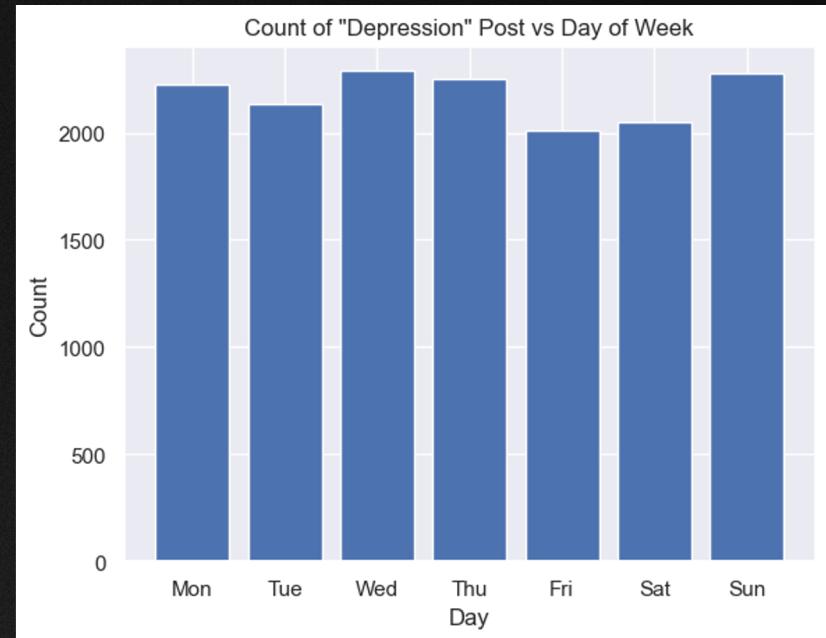
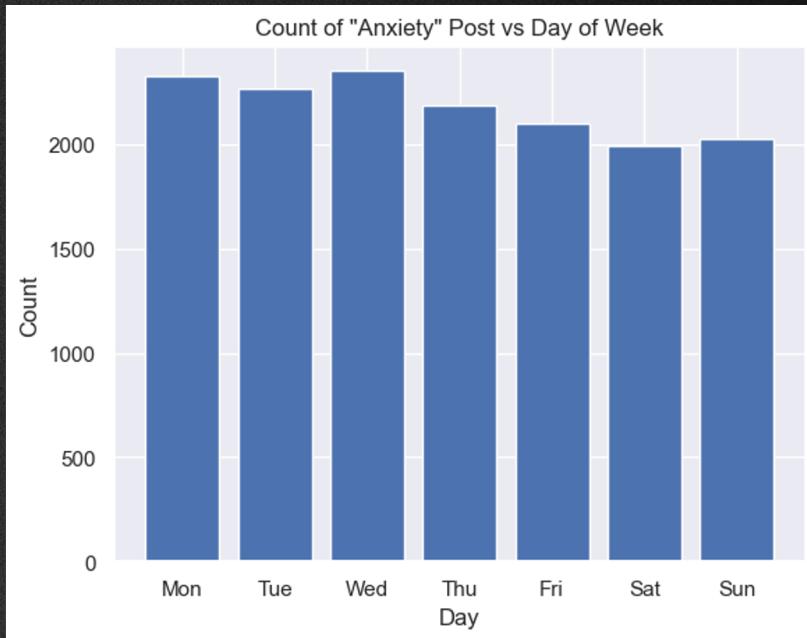
Lower case
Stemming
Unigrams + Bigram

Limit words

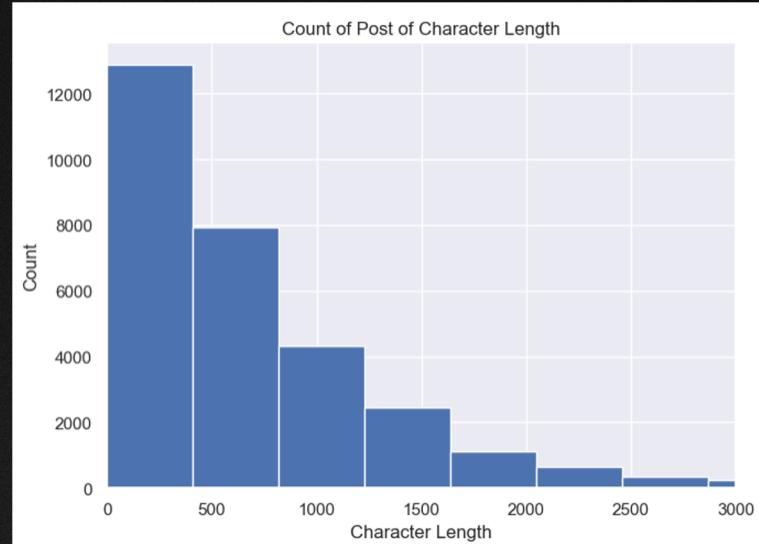
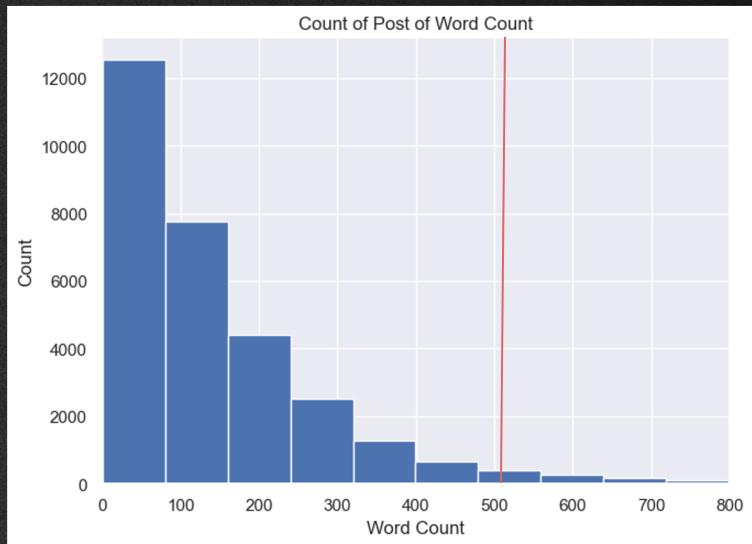


Max_df
Min_df
Max_features

Distribution of Posts by Day



Word/Character Length of Posts



03

Modeling & Sentiment Analysis



Model Comparisons

Pipeline	Vectorizer	Model	Accuracy (Train/Test)
1	CountVectorizer	<u>LogReg</u> /kNN/NB	0.93/0.89
2	TFIDF	<u>LogReg</u> /kNN/NB	0.93/0.90
3	TFIDF	RandomForest	0.88/0.87
PyCaret	TFIDF	<u>Logreg</u> (20 other models)	0.91/0.90



Sentiment Analysis

Hugging Face



**“can’t motivation continuing degree the reason i
good pressure family software industry just i got
good mark good teacher friend i idea continue it’s
getting complicated harder...”**



“can't motivation continuing degree the
reason i good pressure family software
industry just i got good mark good
teacher friend i idea continue it's getting
complicated harder...”

Model 1

Emotions	Scores
Surprise	0.435
Sadness	0.302

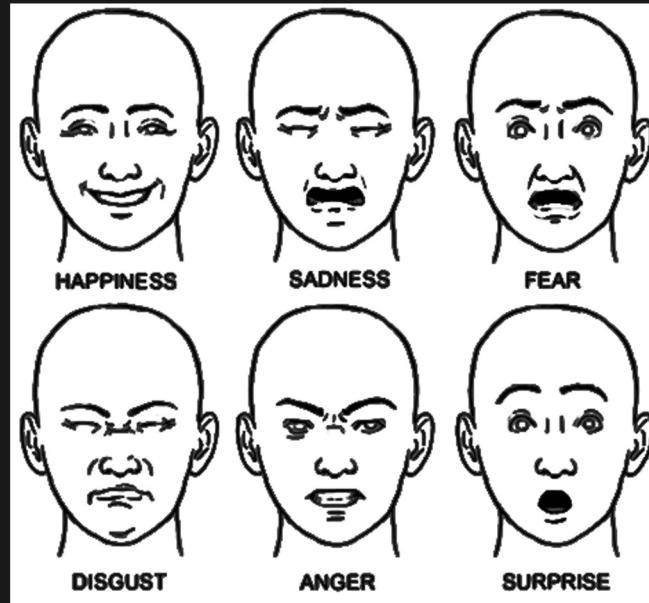
Model 2

Emotions	Scores
Neutral	0.588
Desire	0.158

Model 1 - Primary emotions

Source: [j-hartmann/emotion-english-distilroberta-base](#)

Definition: Initial reaction to external events or stimuli

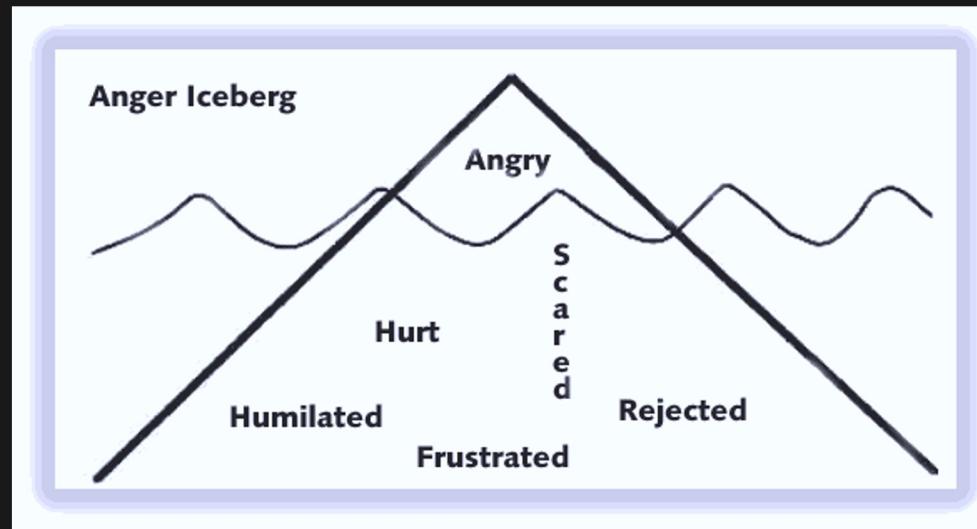


+ neutral

Model 2 - Primary and Secondary Emotions

Source: arpanghoshal/EmoRoBERTa

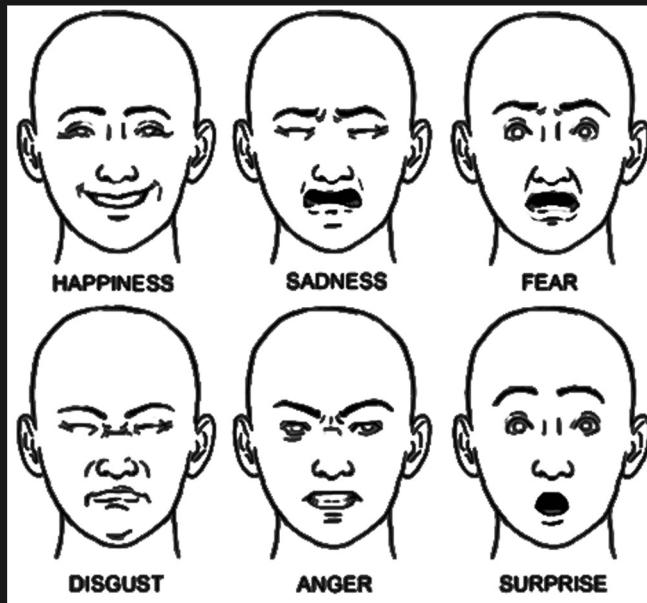
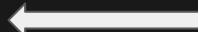
Secondary emotions definition: Emotional reactions we have to other emotions



Model 2 - Primary and Secondary Emotions

Source: [arpanghoshal/EmoRoBERTa](#)

Amusement,
Excitement
Joy,
Optimism,



Annoyance,
Disapproval

*Amusement, anger, approval,
caring, confusion, curiosity, desire,
disappointment, disgust,
embarrassment, excitement, fear,
gratitude, grief, joy, love,
nervousness, optimism, pride,
realization, relief, remorse, sadness,
surprise + neutral*

**“can’t motivation continuing degree the
reason i good pressure family software
industry just i got good mark good
teacher friend i idea continue it’s getting
complicated harder...”**

Model 1

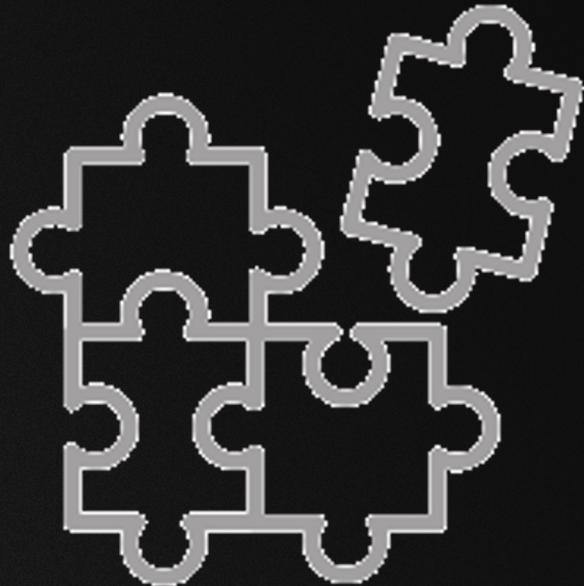
Emotions	Scores
Surprise	0.435
Sadness	0.302

Model 2

Emotions	Scores
Neutral	0.588
Desire	0.158

04

Conclusion



Conclusion



Accuracy

Able to accurately classify topics based on texts using our model



Further insights

Provide insights into the underlying feelings behind people's mood disorders

Limitations



Data retrieved
during a specific
time period



Only texts (lower
case) were
included



Only English texts



Unknown
geographical
location

What's Next?



Identify key words

System will be able to pick up certain emotions and provide support



Dashboard for moderators

Show the top emotions and possibly allow one to pin a response message



Examine links with suicide ideation

Flag out potential posts that match certain emotions

Google

how to kill myself

All Videos Images Shopping News More

About 317,000,000 results (0.52 seconds)

Help is available
Speak with someone today

Samaritans of Singapore
Hours: Available 24 hours. [Learn more](#)

1-767

[Call](#) [Official website](#)

Thank You!



there
is
always
hope.

