r/depression and r/ForeverAlone

You are not alone

Contents.

Introduction

Problem statement and background

Data collection

Scrapping

Data cleaning

EDA and preprocessing
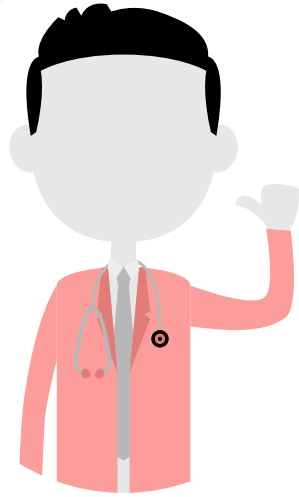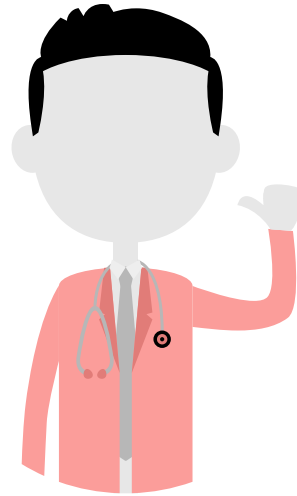
Modelling

Results and misclassifications

Conclusion

Summary and recommendations
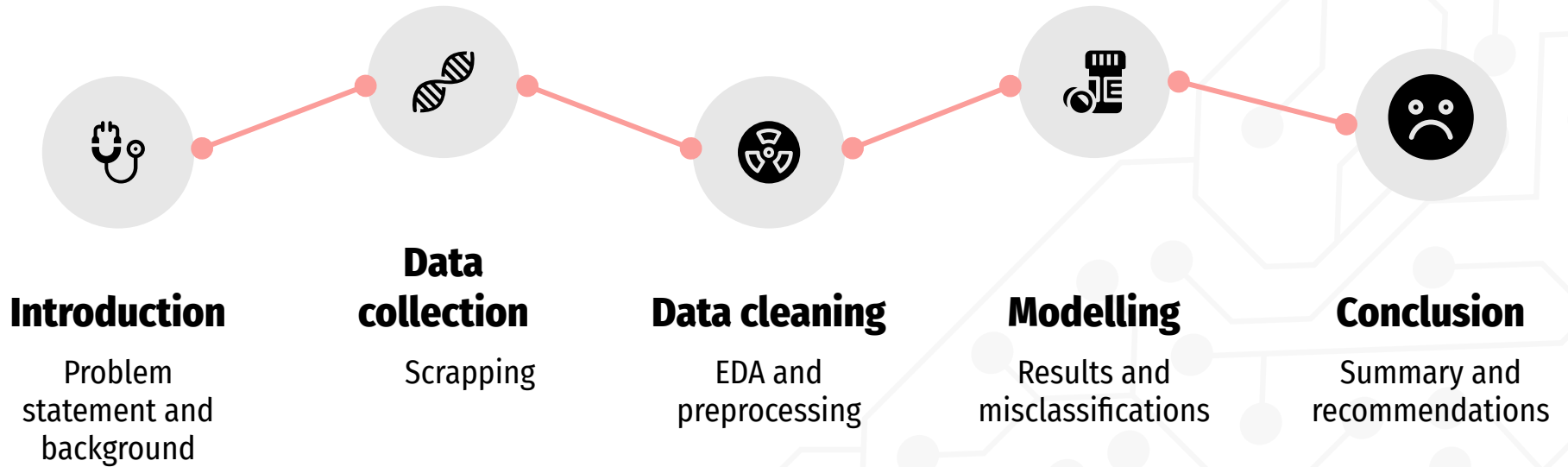
Our Team.

Edward

Mark

Chin Xia

01.

Introduction

Problem Statement

Being part of a research team in the Institute of Mental Health (IMH),This project aims to compare posts on r/depression and r/ForeverAlone using natural language processing with the following objectives:

1) To build a model that is able to correctly classify posts into the respective subreddits, which will be scored on accuracy.

2) To understand erroneous classifications especially in the cases where posts in r/ForeverAlone are wrongly labelled as r/depression posts and whether there might be a cause for concern.

This will provide valuable insights for both early detection of potential depression cases and identification of existing ones. Relevant parties can then reach out to them to offer help and assistance.

Background

/r/depression, because nobody should be alone in a dark place

r/depression

751k
Members

1.4k
Online

r/depression · Posted by u/UniqueMastodon3977 3 hours ago

I feel like I'm broken

I wish I could remember what it felt like to be happy. The only time I feel anything close to that is when I drink but my home life weighs so much on me that I just put on a face and smile so I don't upset anyone there with my problems then I come to work and listen to music and cry in between customers I don't want to feel this way anymore I want to feel loved again.

Background

Forever Alone, Together! [Join]
r/ForeverAlone

About Community ...

A subreddit for Forever Alone.

155k
Members

350
Online

Created 15 Sept 2010

r/ForeverAlone · Posted by u/hualian7 2 days ago

I know I will be forever alone

I can only have one-sided crushes and nothing more. No one will ever love me because I am an unattractive, socially awkward and mentally ill person. I knew this a long time ago but it still hurts.

Data Extraction

```
Pull Number 3, Status Code: 200
Number of post pulled : 337
Pull Number 4, Status Code: 200
Number of post pulled : 417
Pull Number 5, Status Code: 200
Number of post pulled : 497
Pull Number 6, Status Code: 200
Number of post pulled : 580
Pull Number 7, Status Code: 200
Number of post pulled : 663
```

Pushshift Reddit API

- Unique posts per pull
- Number of Posts pulled for r/depression : 2074
- Number of Posts pulled for r/ForeverAlone : 2029

Data cleaning

- Drop nulls
- Lower case
- Drop punctuation
- Drop urls
- Drop digits

Shorthand	Description
.	Any character except new line
\w	Matches alphanumeric characters: `[a-zA-Z0-9_]`
\W	Matches non-alphanumeric characters: `[^\w]`
\d	Matches digit: `[0-9]`
\D	Matches non-digit: `[^\d]`
\s	Matches whitespace character: `[\t\n\f\r\p{Z}]`
\S	Matches non-whitespace character: `[^\s]`

02.

Exploratory Data Analysis

Text Preprocessing

Distribution of word count
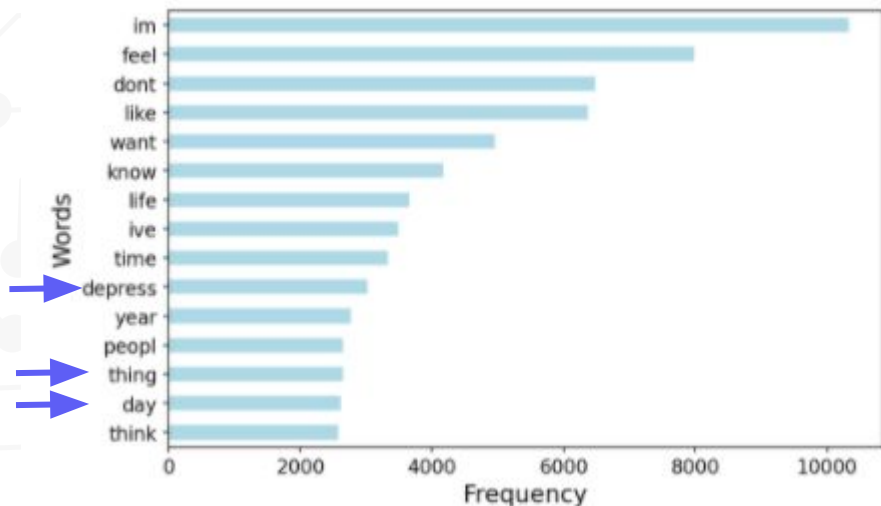
Removal of common words

Normalize the text

Data visualisation

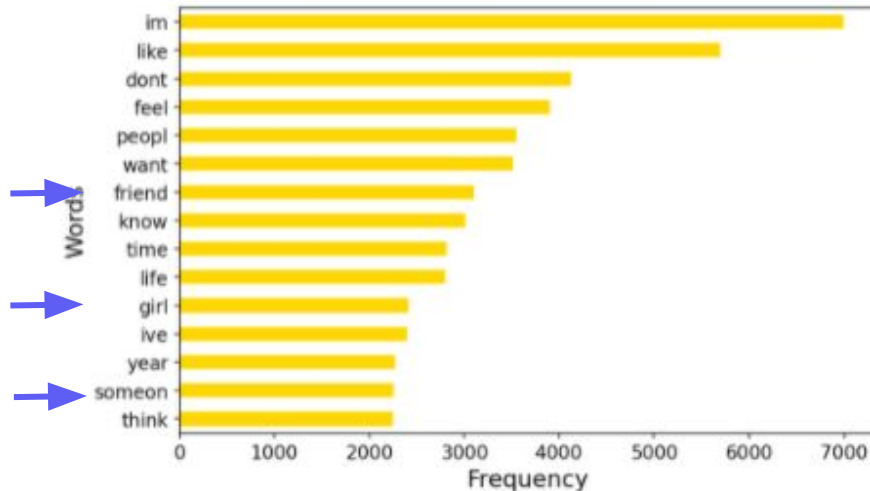# Distribution of word count in depression and foreveralone posts
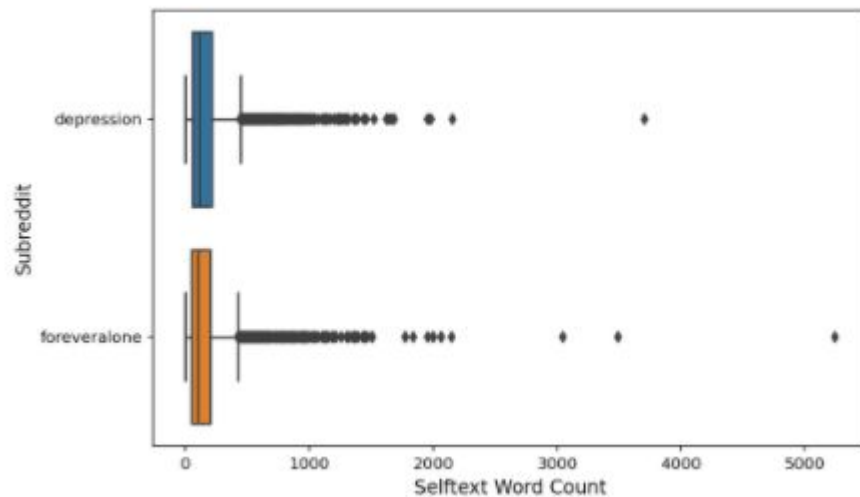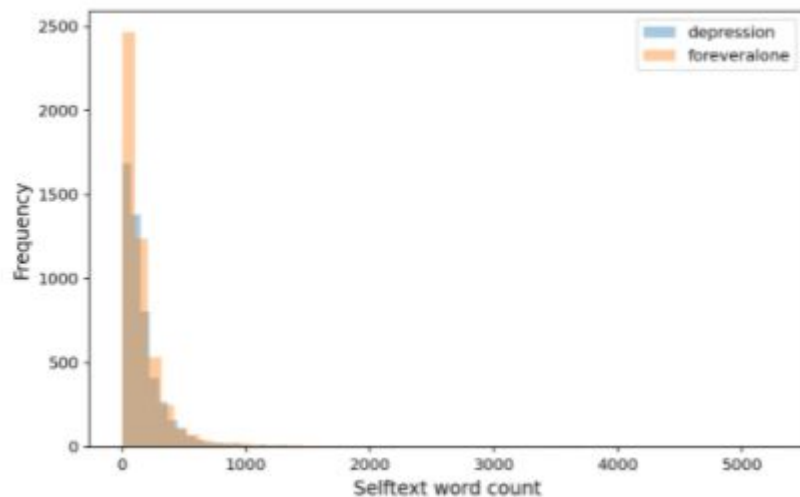
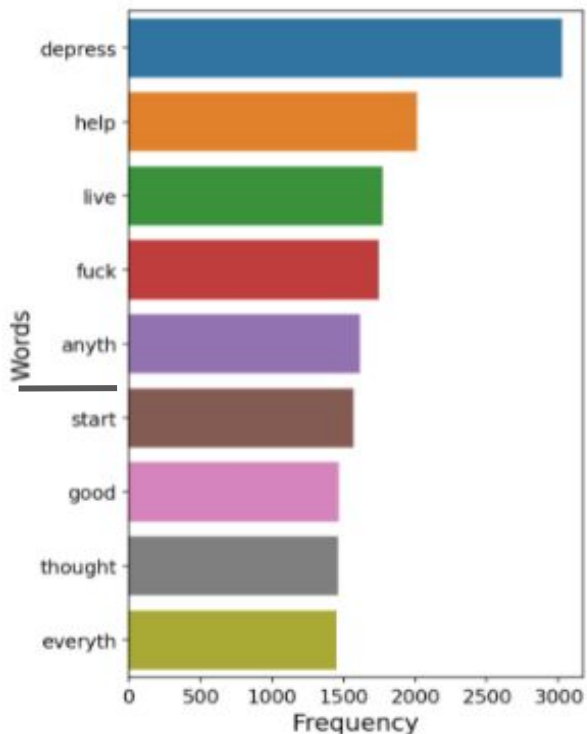# Top 15 words in depression and loneliness posts

Depression
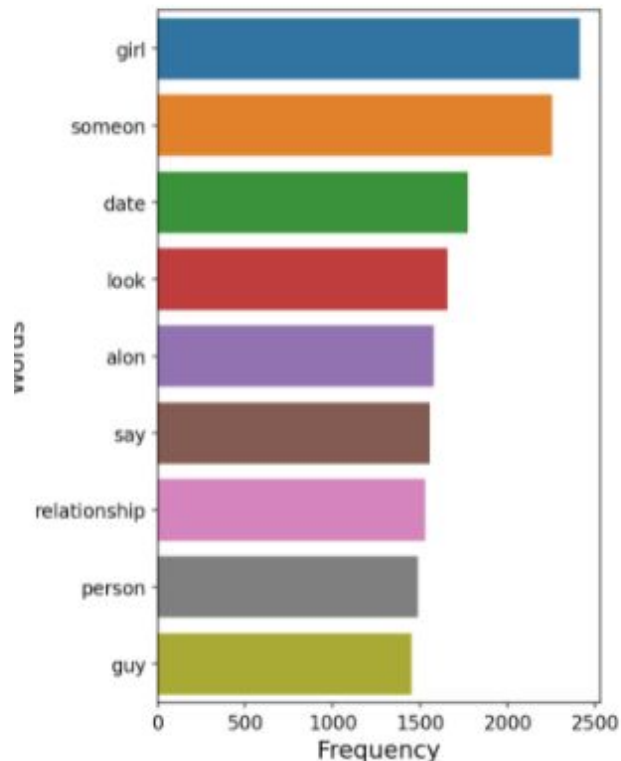
Foreveralone

Words that are specific to the forum out of the top 30 words

03.

Modeling

Building a better model.

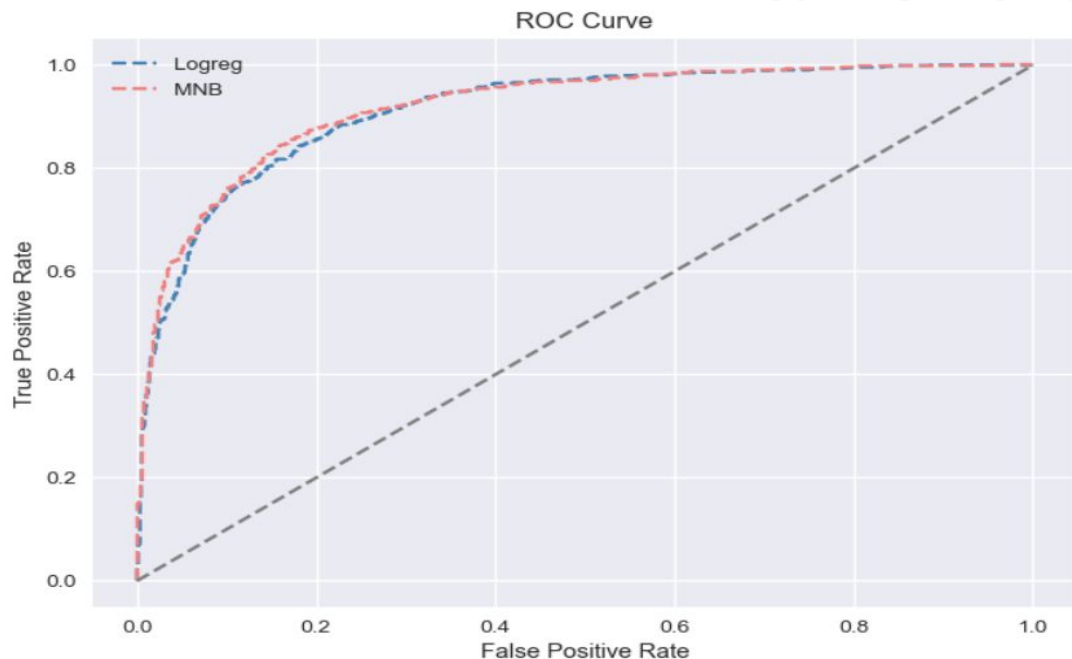| Classifiers	Accuracy (cleaned posts)	Accuracy (cleaned and processed posts)
Base model	50.7%	50.7%
K Nearest Neighbors	51.74%	51.26%
Logistic Regression	83.14%	83.57%
Multinomial NB	83.46%	83.78%

Final model performance.

Accuracy on training set: 88.8%

Accuracy on test set: 83.5%

False positivity: 86.9%
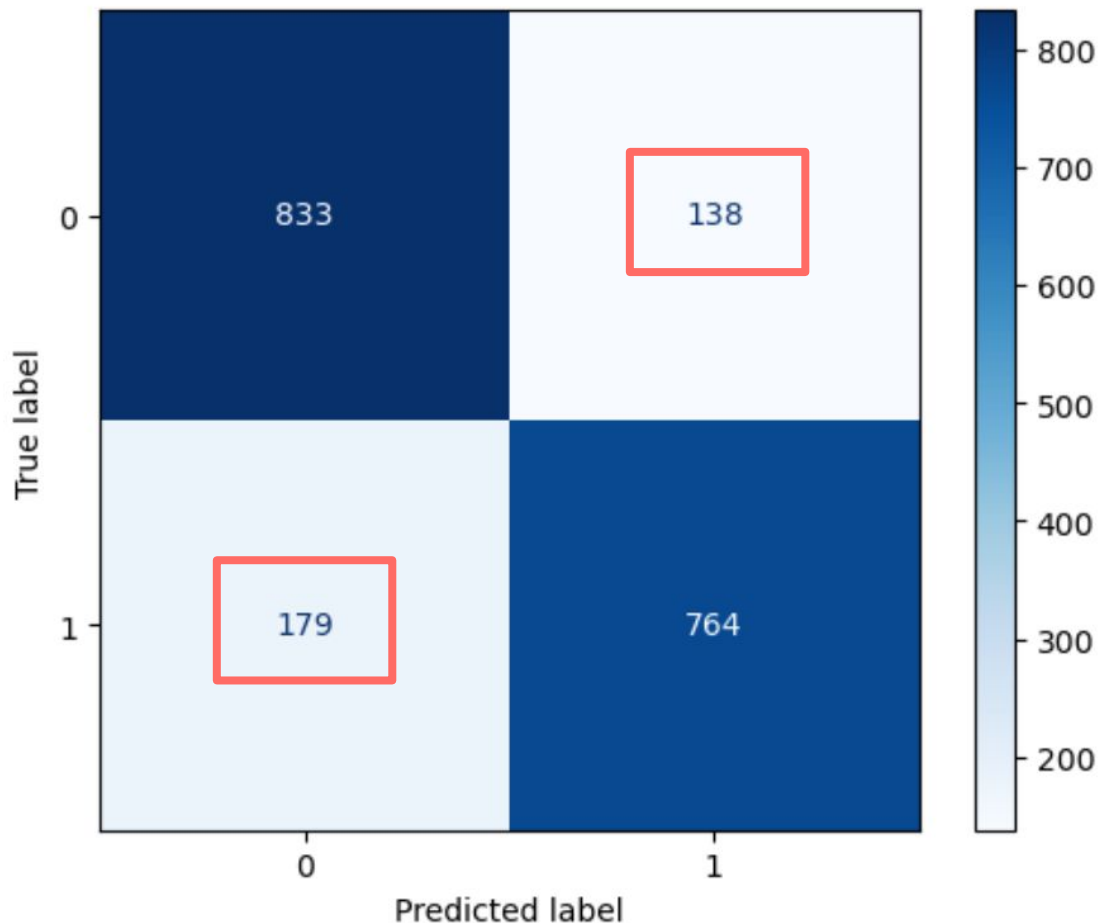
False negativity: 80.0%

AUC score: 0.918

04.

Findings and Conclusions

Examining the misclassifications

138 wrongly classified into r/ForeverAlone

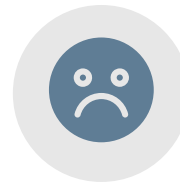**179 posts wrongly classified into r/depression**
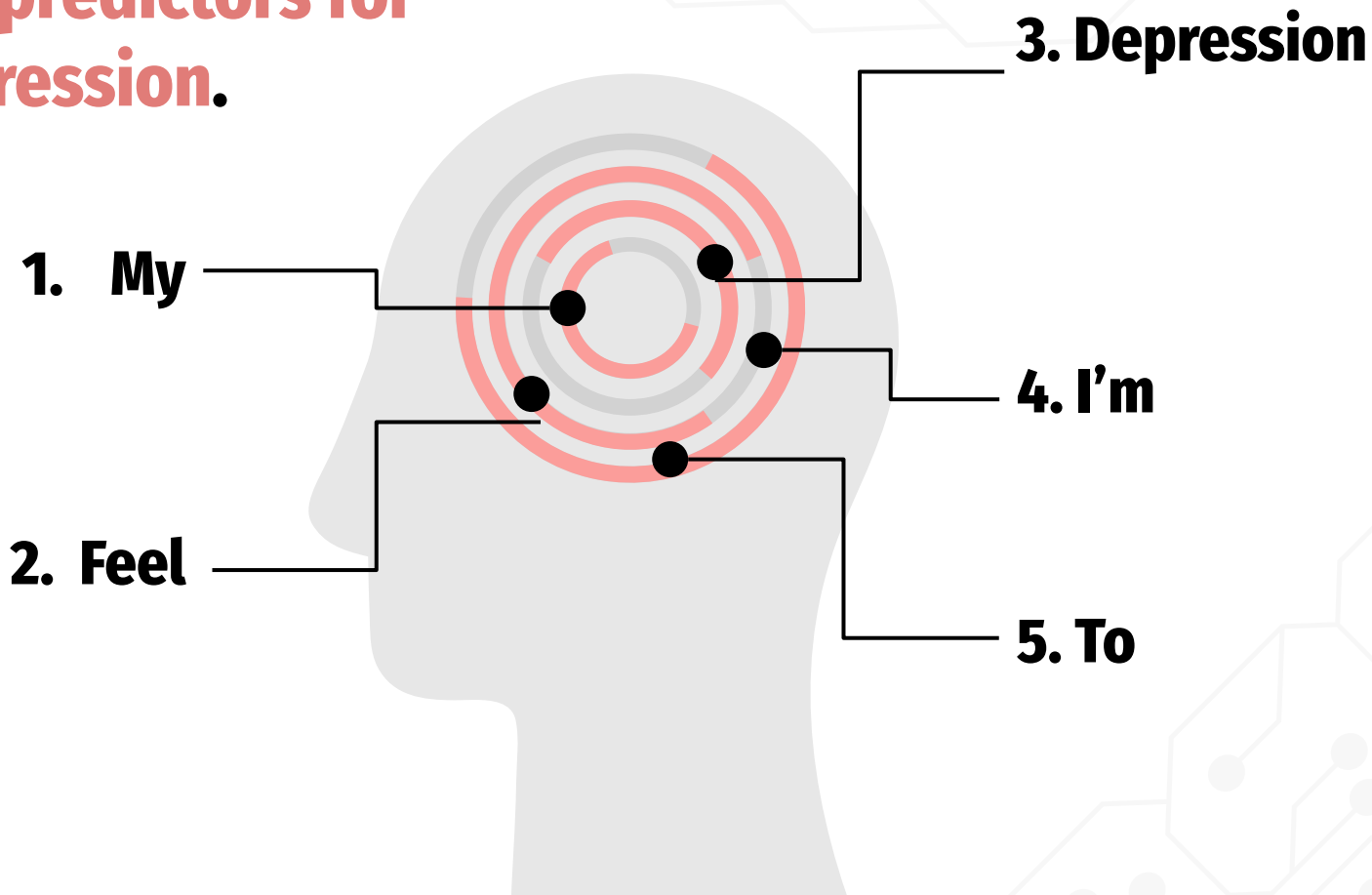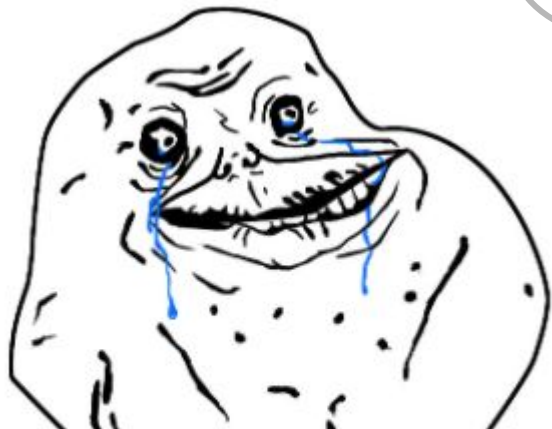
Top 5 key words predictors

My

Girl

Feel

Depression

You

Top 5 predictors for r/depression.

1. My
2. Feel
3. Depression
4. I'm
5. To

Top 5 predictors for r/ForeverAlone

1. Girl
2. You
3. Woman
4. FA
5. Guy

Conclusions #1.

We built a model that explored the topics of depression and loneliness.

- 32.8% improvement from the base model.

- New model with 83.5% accuracy.

Conclusions #2.

There are certain key words that help our model classify the submissions better.

- My
- Girl
- Feel
- Depress
- You

Conclusions #3.

Reach out to others for more social interactions

Misclassifications could be an indicator of depression.

- Excessive and persistent loneliness may lead to depression.

- Misclassifications can happen when certain keywords ('feeling, depression) are used

- Can serve as a red flag for further study on user behavior such as post frequency and age in forum

- More can also be done in terms of classification in terms of severity of depression and loneliness

Thank you! (We appreciate you)

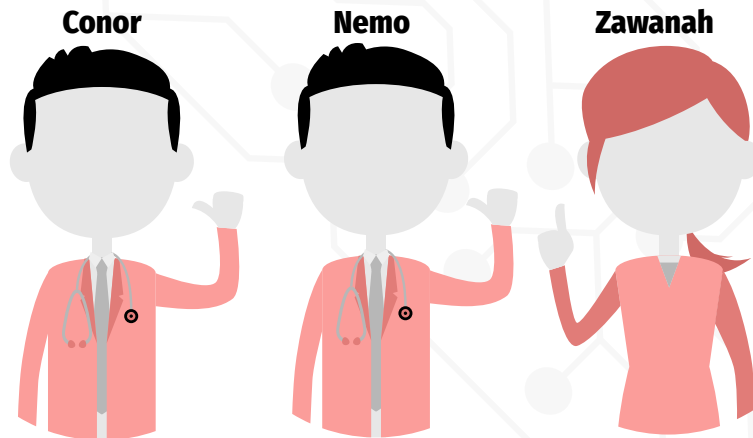**Additional Resources:**

https://www.healthhub.sg/live-healthy/87/coping_with_depression_

https://stayprepared.sg/

Acknowledgements

Conor   Nemo   Zawanah

Depression hotline:

SOS Samaritans of Singapore

1800 - 221 4444