Active Depression Monitoring and Alert System in Reddit through Machine Learning

GROUP 14

Ken Cheah – A0218906Y
Krishnan Ananth – A0218894M
Tommy Kangdra – A0218866N
Victor Goh – A0067229Y
Xiao Yidi – A0218962W

TABLE OF CONTENTS

01

INTRODUCTION

BT5153

INTRODUCTION

- Depression is a world-wide problem
- The amount of time of social media use is found to be significantly correlated with increased depression [1]
- Reddit is considered as a safe and anonymous space for individuals to share their feelings honestly.

BUSINESS CASE

Our solution allows social media platforms the ability to detect, monitor, intervene, and connect highly depressed, self-harming or suicidal individuals with non-profit mental health awareness / care organizations to receive the help / support they urgently need as part of the platform's Corporate Social Responsibility initiatives.

OBJECTIVES

1. Develop a classification model that accurately identifies depressed and suicidal users based on their textual postings
2. Design an intervention framework to detect individuals with severe deterioration in their mental states and to intervene where necessary

[1] De Choudhury, M., Gamon, M., Counts, S., & Horvitz, E., et all. (2013). Predicting Depression via Social Media. Proceedings of the International AAAI Conference on Web and Social Media, 7(1).

02

DATA & EDA
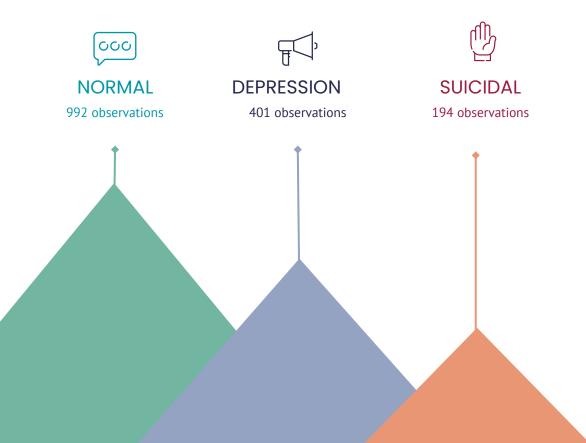
BT5153

DATA COLLECTION

reddit

Taken from depression, suicide watch, and confession subreddit forums with Pushshift API
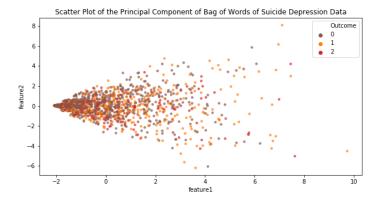
Class Labelling

60 hours of manual labelling
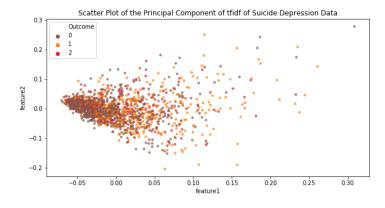
- Random Shuffled
- Three human labelers
- Majority vote

NORMAL
992 observations

DEPRESSION
401 observations

SUICIDAL
194 observations

VISUALIZING THE CLASSES

with PCA (nclass=2)

COUNT VECTORIZER

Scatter Plot of the Principal Component of Bag of Words of Suicide Depression Data

TFIDF VECTORIZER

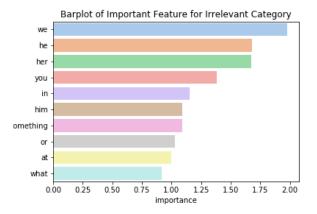Scatter Plot of the Principal Component of tfidf of Suicide Depression Data

Description:
0: Normal
1: Depression
2: Suicide

NORMAL
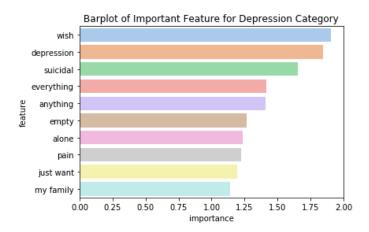
The normal class represents posts which do not contain depression or suicide language.

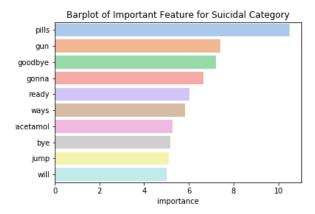Barplot of Important Feature for Irrelevant Category

DEPRESSION

The depression class represents comments which exhibit helplessness, a cry for help, depressive loneliness, lack of any motivation, or being empty and numb to their emotion.

Barplot of Important Feature for Depression Category
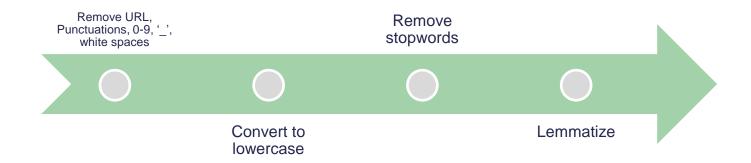
SUICIDAL

The suicidal class represents comments which exhibit self-harming tendencies, suicide planning and execution. For people to act on their plan, it took a lot of consideration and therefore only after struggling for a while that the person may be "ready" to act on the plan.

Barplot of Important Feature for Suicidal Category

03

PREPROCESSING

BT5153

Text Pre-Processing

Remove URL, Punctuations, 0-9, '_', white spaces

Remove stopwords

Convert to lowercase
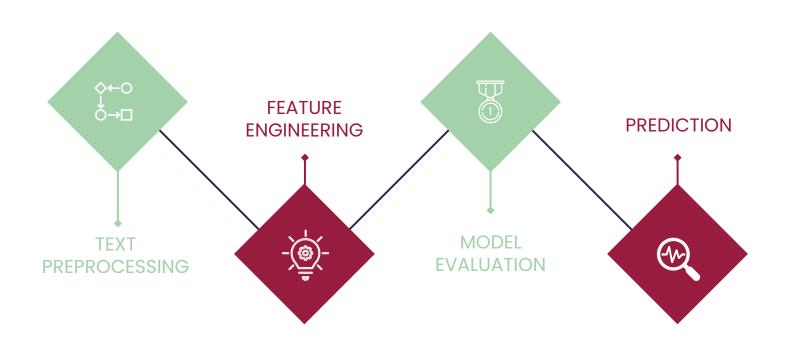
Lemmatize

Suicidal Post

```
original post:
 Easiest way to end it? I think hanging &amp; gunshot are out of the question. I've tried overdosing a few times before and eve
ry time I just end up puking. Maybe I need stronger pills? I've only tried quetiapine. I have always wondered about driving my
car full speed into a wall or something, but I don't know if it would actually work.
--------------------
Processed post:
 easiest way end think hanging amp gunshot question tried overdosing time every time end puking maybe need stronger pill tried
quetiapine always wondered driving car full speed wall something know would actually work
```

04

MODEL & METHODOLOGY

BT5153

Methodology

TEXT
PREPROCESSING

FEATURE
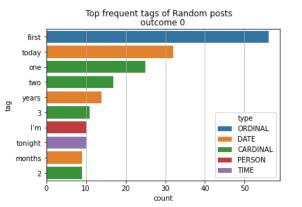ENGINEERING

MODEL
EVALUATION

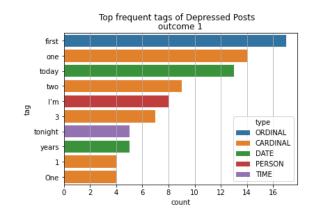PREDICTION

Feature Engineering (BoW)

Name Entity Recognition (18 Tags)

i'm pathetic and a coward just got home. walked to the bridge again. this time i was outside the railing. it took a long time before doing that. i stood there for `some minutes TIME` before walking on the side and off the bridge. when i stood there i thought it would be easy to just, do it. apparently not and i don't know why. my mind says yes but my body says no. i feel so pathetic and like a coward for not being able to do it. i called the psych hospital and talked with some lady for `30 minute-ish TIME` . when i was halfway home she said i'm on my own now to get home. was like... ok. then we hung up.
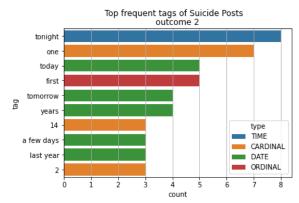
`third ORDINAL` time since `wednesday night TIME` i've been there. was there during `thursday night TIME` as well. somehow managed to walk home. i really, REALLY wanted to do it.

just...   i'm not well.

Top frequent tags of Random posts
outcome 0

Top frequent tags of Depressed Posts
outcome 1

Top frequent tags of Suicide Posts
outcome 2
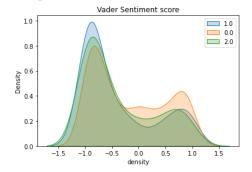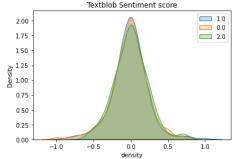
Feature Engineering (BoW)

Vader Sentiment Score

Vader Sentiment score

TextBlob Sentiment Score

Textblob Sentiment score

Suicidal Post

```
Text_processed: pathetic coward got home walked bridge time outside railing took long time stood minute walking side bridge sto
od thought would easy apparently know mind say yes body say feel pathetic like coward able called psych hospital talked lady mi
nute ish halfway home said get home like ok hung third time since wednesday night thursday night well somehow managed walk home
really really wanted       well
********************

Textblob sentiment score: -0.0697
Vader sentiment score: -0.5334
```

Feature Engineering (BoW)

Random post

'lied wanting get better ago family member asked wanted get better said yes scared angry said sometimes feel like friend would nice truth think would make thing complicated better knowing scared get better like normal know could cope normal functioning person seems draining difficult wish place could go truly abandoned people whatsoever could explore enjoy place without bothering people people bothering thanks reading'

TF-IDF (ngram=(1,2), min_df=2)

N_gram

'actually',
'actually care',
'actually feel',
'actually go',
'actually going',
'actually killing',
'add',

TF-IDF Vectors

Sparse Matrix form of data :

[[0.   0.13749421 0.   ... 0.   0.   0.   ]]

Topic Modelling (15 topics)

Depressed Topic

Topic: 0
Words: 0.012*"like" + 0.011*"depression" + 0.010*"know" + 0.010*"year" + 0.009*"feel" + 0.008*"time" + 0.008*"depressed" + 0.008*"think" + 0.008*"day" + 0.007*"get"

Topic: 1
Words: 0.013*"like" + 0.011*"go" + 0.010*"feel" + 0.010*"want" + 0.008*"day" + 0.008*"know" + 0.008*"time" + 0.007*"nothing" + 0.007*"thought" + 0.007*"one"

Suicidal / Normal Topic

Feature Engineering (DL)

Text Augmentations

Synonym Replacement
- Replaces random words / phrases based on similar words / phrases

"chronic disappointment what is the cause of chronic disappointment ways to **dealing with** disappointment"

↓

"chronic disappointment what is the cause of chronic disappointment ways to **get by** with disappointment"

Feature Engineering (DL)

Text Augmentations

Contextual word replacement
- Uses pretrained BERT embeddings to replace random words / phrases with similar words / phrases

"chronic disappointment what **is the** cause of chronic disappointment
ways to **dealing with** disappointment"

↓

"chronic disappointment what is **a use** of chronic disappointment
ways to **describe a** disappointment"

Feature Engineering (DL)

Text Augmentations

Back translation
- Translates sentences to a different language and then translates them back to English

"chronic disappointment what **is the cause of chronic disappointment ways to dealing with disappointment**"
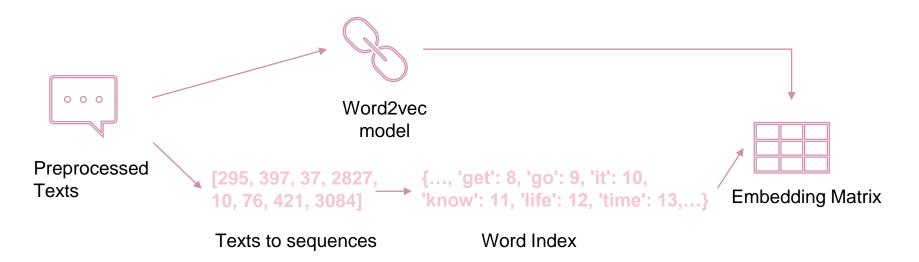
"Chronische Enttäuschung, was ist die Ursache für chronische Enttäuschung Wege, um mit Enttäuschung umzugehen"

"chronic disappointment what **causes chronic disappointment?**"

Feature Engineering (DL)

Word2vec

- Vector space built according to Semantic Similarity
- Converting words to 50-dimensional vector embedding matrix

Word2vec model

Preprocessed Texts

[295, 397, 37, 2827, 10, 76, 421, 3084]

{…, 'get': 8, 'go': 9, 'it': 10, 'know': 11, 'life': 12, 'time': 13,…}

Embedding Matrix

Texts to sequences

Word Index

1. Bag of Words Models

Logistic Regression

A simple linear classification method

Support Vector Machine
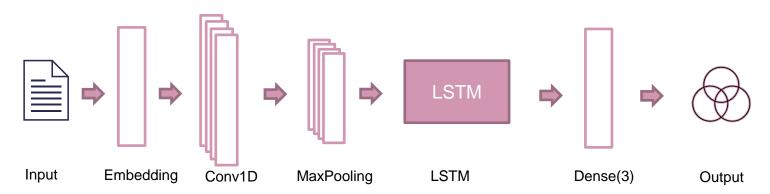
Ability to use non-linear decision boundaries with kernels

Naïve Bayes

Based on Bayes Theorem with an assumption of independence among predictors

2. CNN LSTM Model

Input — Embedding — Conv1D — MaxPooling — LSTM — Dense(3) — Output

- Convolutional Neural Network (CNN) layer for feature extraction from text data
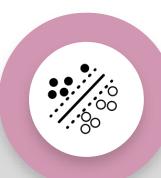- LSTM layer for tracking sequential dependencies

- Preprocessing Steps: Text to sequence, NUMBER OF WORDS = 10000
- Conv1D kernel size = 3, Optimizer = Adam, Number of Iteration = 5
- Grid Search:
  Learning rate: [0.1,0.05,0.01,0.005]
  Number of Conv1D Filter: [32, 64]
  Dimensionality of the LSTM output space: [40, 80]
  Number of Epochs: [3, 5,10]
- Learning rate = 0.01, # of filter = 32, Dimension of LSTM = 40 and number of epochs = 5 gives the best result in terms of f1 scores in depression and suicidal classes.

3. BERT

Bidirectional Encoder Representations from Transformers

Language Modeling

chef

the

Masked Language Modeling

the   chef   cooked   the   meal

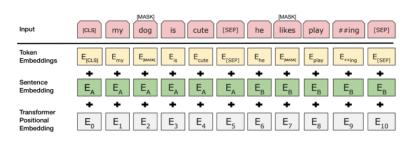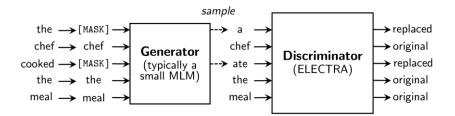- State of the art language model that is able to learn the semantics of language better than a bag-of-words model using bidirectional contextual learning with self-attention mechanisms

- Takes in three embedding vectors simultaneously as inputs

- Pretraining involves self-supervised tasks:
  - Masked Language Modeling
  - Next Sentence Prediction

Input	[CLS]	my	[MASK] dog	is	cute	[SEP]	he	[MASK] likes	play	##ing	[SEP]
Token Embeddings	$E_{[CLS]}$	E_{my}	$E_{[MASK]}$	E_{is}	E_{cute}	$E_{[SEP]}$	E_{he}	$E_{[MASK]}$	E_{play}	$E_{\#\#ing}$	$E_{[SEP]}$
	+	+	+	+	+	+	+	+	+	+	+
Sentence Embedding	E_A	E_A	E_A	E_A	E_A	E_A	E_B	E_B	E_B	E_B	E_B
	+	+	+	+	+	+	+	+	+	+	+
Transformer Positional Embedding	E_0	E_1	E_2	E_3	E_4	E_5	E_6	E_7	E_8	E_9	E_{10}
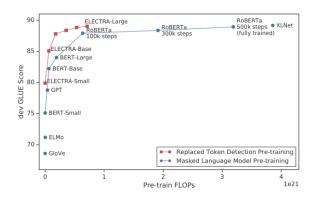
Transfer Learning using ELECTRA

Replaced Token Detection

- Unlike conventional BERT, ELECTRA trained as discriminator using Replaced Token Detection

- More efficient compared to current SOTA models (good performance at a fraction of the size)

- Finetuned ELECTRA-Small by using pretrained weights from Wikipedia and BooksCorpus (same as in the models by the original BERT authors)

05

PERFORMANCE EVALUATION

BT5153

Evaluation Criteria

F1-Score

- High F1-Score -> Good balance between precision and recall

Recall Rate

- High recall rate for each class -> Get each class predictions right

Bag of Words Models

SVM

Model: SVM		Predicted		
		Normal	Depression	Suicidal
Actual	Normal	161	35	3
	Depression	29	48	3
	Suicidal	10	21	8

```
              precision    recall  f1-score   support

         0.0       0.81      0.81      0.81       199
         1.0       0.46      0.60      0.52        80
         2.0       0.57      0.21      0.30        39

    accuracy                           0.68       318
   macro avg       0.61      0.54      0.54       318
weighted avg       0.69      0.68      0.67       318
```

NB

Model: NB		Predicted		
		Normal	Depression	Suicidal
Actual	Normal	84	83	32
	Depression	9	60	11
	Suicidal	1	20	18

```
              precision    recall  f1-score   support

         0.0       0.89      0.42      0.57       199
         1.0       0.37      0.75      0.49        80
         2.0       0.30      0.46      0.36        39

    accuracy                           0.51       318
   macro avg       0.52      0.54      0.48       318
weighted avg       0.69      0.51      0.53       318
```

LR

Model: LR		Predicted		
		Normal	Depression	Suicidal
Actual	Normal	152	38	9
	Depression	20	54	6
	Suicidal	4	20	15

```
              precision    recall  f1-score   support

         0.0       0.86      0.76      0.81       199
         1.0       0.48      0.68      0.56        80
         2.0       0.50      0.38      0.43        39

    accuracy                           0.69       318
   macro avg       0.62      0.61      0.60       318
weighted avg       0.72      0.69      0.70       318
```

Deep Learning Models

CNN-LSTM

Model: CNN-LSTM		Predicted		
		Normal	Depression	Suicidal
Actual	Normal	166	24	9
	Depression	42	25	13
	Suicidal	20	5	14

```
              precision    recall  f1-score   support

           0       0.73      0.83      0.78       199
           1       0.46      0.31      0.37        80
           2       0.39      0.36      0.37        39

    accuracy                           0.64       318
   macro avg       0.53      0.50      0.51       318
weighted avg       0.62      0.64      0.63       318
```

BERT

Model: BERT		Predicted		
		Normal	Depression	Suicidal
Actual	Normal	152	38	9
	Depression	21	55	4
	Suicidal	12	14	13

```
              precision    recall  f1-score   support

           0       0.82      0.76      0.79       199
           1       0.51      0.69      0.59        80
           2       0.50      0.33      0.40        39

    accuracy                           0.69       318
   macro avg       0.61      0.59      0.59       318
weighted avg       0.70      0.69      0.69       318
```

Modelling Summary

- Logistic Regression performed the best

- 5 Features were engineered: TF-IDF, Vader, Textblob, Topicmodel, NER

- Feature engineering improves performance

- Deep Learning models such as CNN-LSTM & BERT requires much more data to be performant

06

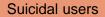# ACTIVE MONITORING APPLICATION

BT5153

Business case

Reddit Users

Social Media Platform

Suicidal users

Depressed users

Normal users

Mental Health Support Organizations

Intervene by redirecting users to mental health support organizations

Method

Highlight users with suicidal intent

Any user with recent posting acitivity showing suicidal intent will be listed as requiring immediate attention.

Flag 1

Identify users with potential suicidal intent

If any user has any depressed posts which have high levels of suicidal intent, these users will be highlighted as requiring attention.

Flag 2

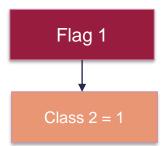## Monitor users with high amount of depressed posts

If any user is showing a significant amount of depressed posts in a recent time period, the user will be idenitified as requiring monitoring.
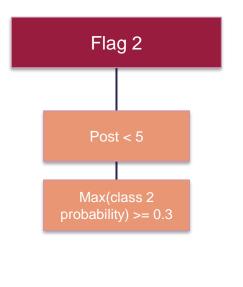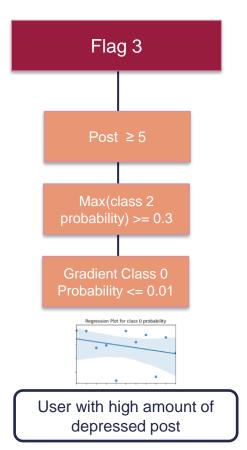
Flag 3

Logic

Flag 1

Class 2 = 1

Class 0 : Normal posts
Class 1: Depression posts
Class 2: Suicidal posts

User with suicidal intent

Flag 2

Post < 5

Max(class 2 probability) >= 0.3

User with potential suicidal intent

Flag 3

Post ≥ 5

Max(class 2 probability) >= 0.3

Gradient Class 0 Probability <= 0.01

Regression Plot for class 0 probability

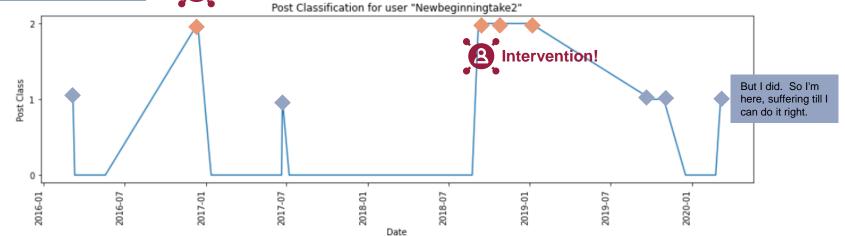User with high amount of depressed post

Case Study #1 (Flag 1)

I'm sure I'll end up going that way...Suicide that is. I can't get the thoughts out of my head. Haunted...ghosts of the past, demons that have been with me so many years. I can't fight any longer. …

…But my past week has been manic highs, suicidal thoughts, more cutting and burning ….

[story…] If it needs to be said, I've been battling depression for over 25 years, had a breakdown, multiple hospitalizations and still battle every day with urges of self harm, eating disorder, suicidal thoughts …
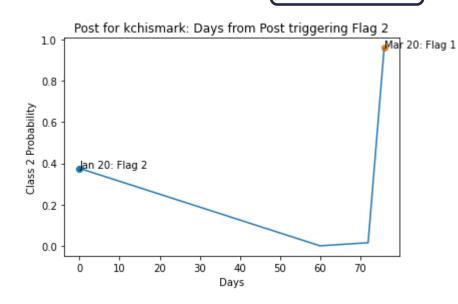
I know it's real bad, I know I should get stitches but I'm alone and I probably won't. Any advice for large gaping cuts needed!

Intervention!

Intervention!

But I did. So I'm here, suffering till I can do it right.

Post Classification for user "Newbeginningtake2"

Post Class

2

1

0

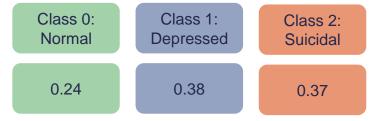2016-01 2016-07 2017-01 2017-07 2018-01 2018-07 2019-01 2019-07 2020-01

Date

Case Study #2 (Flag 2)

User with potential suicidal intent

User: kchismark

Its been two months since my last attempt and things just keep going south for me. Today would be a really inconvenient day for me to just end it, with it being the national championship and all. Ruin your night because you ruined my life. My life doesn't matter enough to you and it's okay. I don't want to be here anyway so looks like you're in luck. You never have to worry about me interfering in your life anymore.

Post for kchismark: Days from Post triggering Flag 2

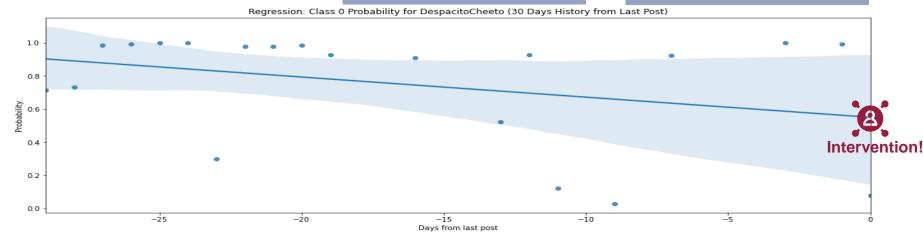| Class 0: Normal	Class 1: Depressed	Class 2: Suicidal
0.24	0.38	0.37

Case Study #3 (Flag 3)

I have only done it 2 days now and recently failed. I was too depressed to resist. I feel so isolated and alone, and the worse part is I've done it to myself. I have so little will, and I wanna know what goal I can start with for things to improve a little.

I am done…I don't even know what I want, and I don't think I want anything, except a peaceful end….but nothing means anything to me… I've had the life experience. Just need to tell someone...

I feel everytime I see a couple at school or on YouTube or anywhere I get super depressed.

I can't handle this anymore…
Tell me how to die. I BEG you. I don't want to be here and feel nothing but exhaust and sorrow

What the hell do I do? I can't live in loneliness, but I can't die.

Regression: Class 0 Probability for DespacitoCheeto (30 Days History from Last Post)

Intervention!

07

CONCLUSION

BT5153

CONCLUSION

MOST PERFORMANT MODEL

Logistic Regression with 5
Features were engineered: TF-IDF,
Vader, Textblob, Topicmodel, NER

APPLICATION & BUSINESS CASE

Out of 420 users:
- 23.57% of users are flagged with suicidal intent
- 30.00% of the users are flagged with potential suicidal intent and high amount of depressed posts

LIMITATION

DATA

More data to help models learn and distinguish between classes (especially for deep learning models)

Medical Expertise

There is no ground truth when labelling / identifying comments that have depression or suicidal elements. A dataset that is manually curated by mental health experts will lead to a more representative dataset which helps model perform better in real world settings.

THANK YOU

GROUP 14