ANATOMY OF A SUICIDE POST

Project 3: Web APIs and Classification

Alicia Tay

The Problem

"… to say that depression is the cause of all suicides is a misguided generalization repeated too often.

Someone who isn't depressed can still be at risk of suicide, and not everyone suffering from depression dies by suicide."

Samaritans of Singapore

The Task

- Build a model that sieves out suicide-related posts *(target class)* from depression-related posts, and alert moderators on priority posts for them to further investigate

- Give insights to the online behavioural differences

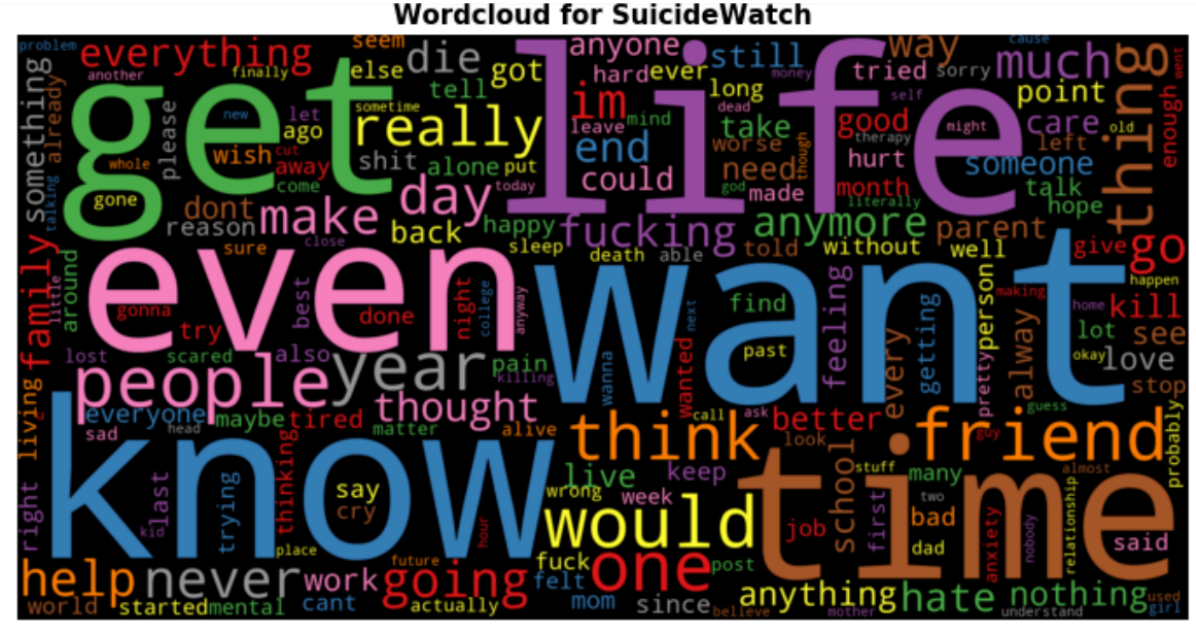
The Data

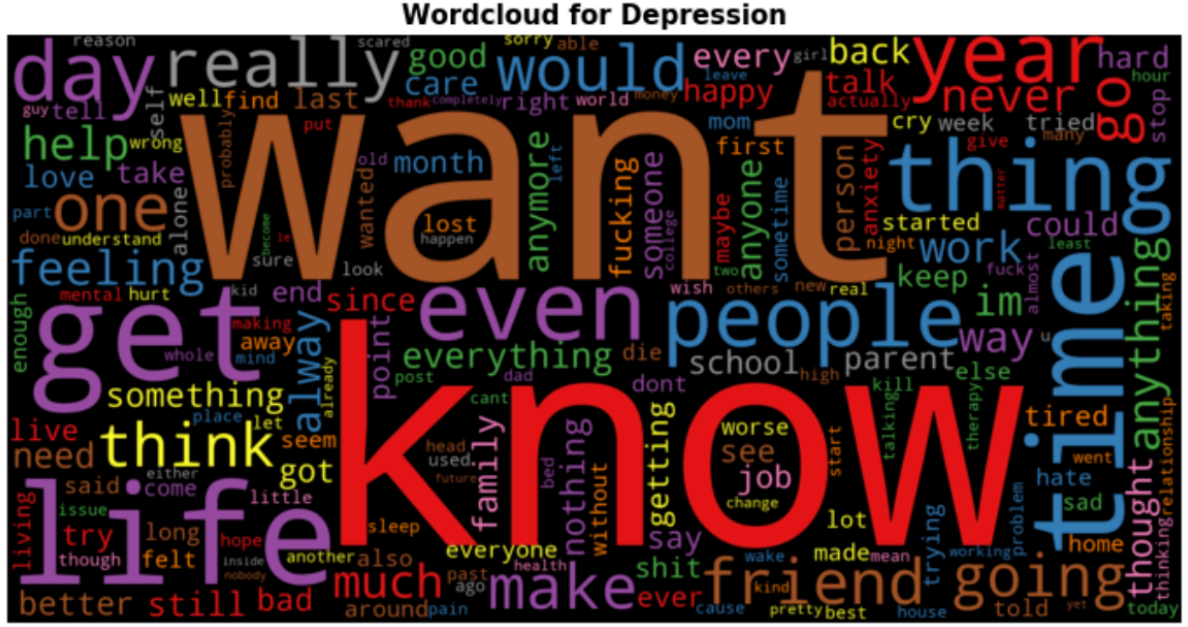
- r/SuicideWatch *(target class)*: 987 unique posts
- r/depression: 970 unique posts

INSIGHTS

from EDA

Common words appear in both subreddits, but 2-grams and 3-grams paint a fuller picture



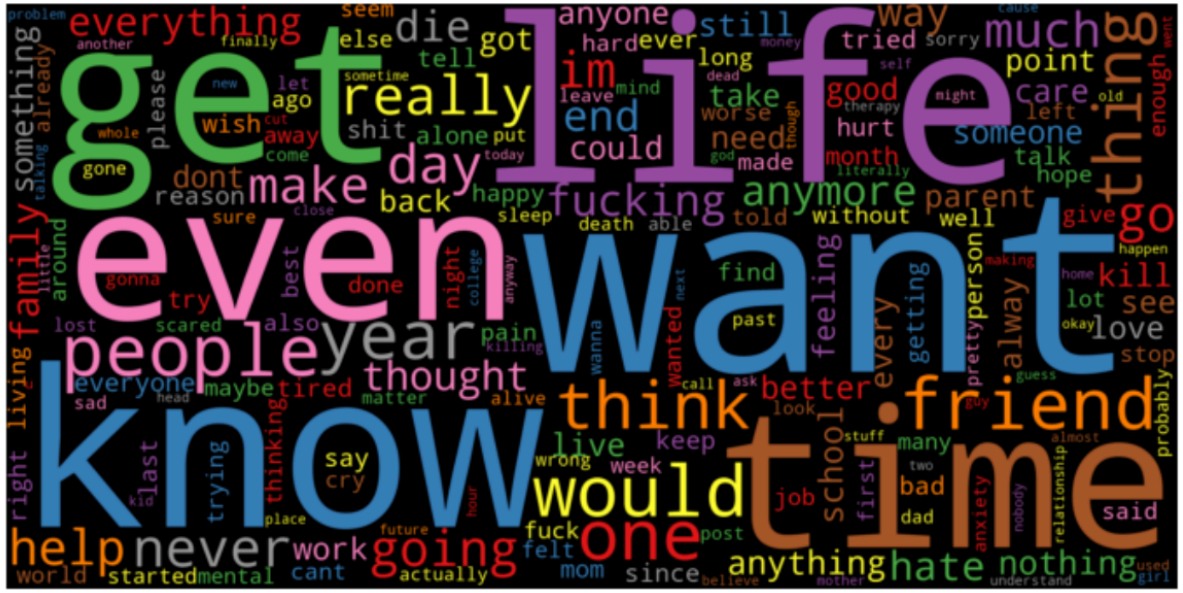


want to kill (myself)
want to end (my life)

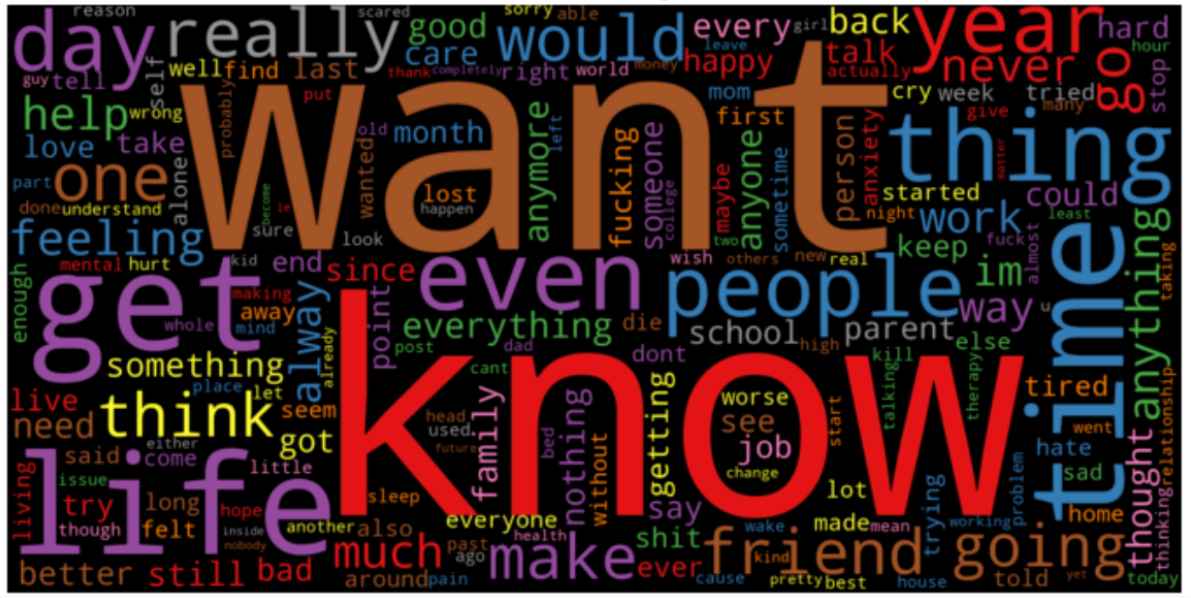
still **want** to live
want to get (better)

Both subreddits draw inferences to social ties

Wordcloud for SuicideWatch



Wordcloud for Depression



Eg: “family”, “friend”, “college”

Differences in connotations

Top 3-grams for r/SuicideWatch
- Connotes finality, more drama
- Lost of hope

nothing nothing nothing	36
want live anymore	10
thing get better	10
want die want	10
every single one	9
every single day	8
want end life	8
mental health issue	7
life worth living	6
mom dad grandma	6
anyone would care	5
even though know	5
fuck fuck fuck	5
see point living	5
dont even know	5
get work done	5
part time job	5
sleep never wake	5
nothing look forward	5

Top 3-grams for r/depression
- Contextual/feeling based,
- At tantrum with a life's moment

http open spotify	14
open spotify playlist	14
loser stupid loser	12
stupid loser stupid	12
want live anymore	11
work go home	11
alarm clock go	11
clock go work	11
go work go	10
go home bed	10
thinking ending thing	8
want live life	7
matter hard try	7
want die want	6
since high school	6
part time job	5
want get better	5
know need help	5
sleep day away	5

INSIGHTS

from Modelling

The Scoring Metric: F1 Score

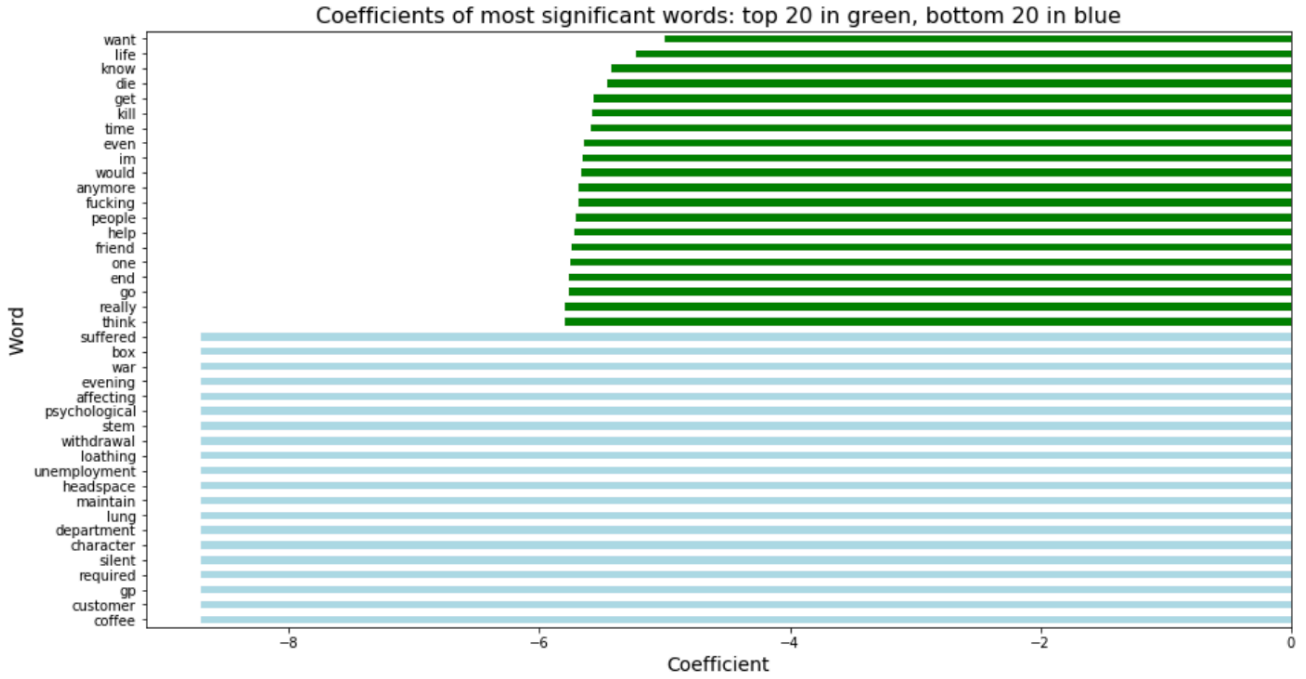
- Time-crucial element, where one requires immediate prevention
- As the cost of false positive *(predicted suicidal, but only depressed)* and false negatives *(predicted depression, but suicidal)* are very different and both needed to be minimized while maximising True positives, my models are evaluated on both Precision and Recall/Sensitivity.

- Baseline accuracy: 50.43%
- MultiNB Model with TF-IDF performs the best on test set with the highest f1score

	model	train_score	test_score	train_f1	test_f1
0	cvec + logistic regression	0.942059	0.646939	0.943968	0.653307
1	tvec + logistic regression	0.770279	0.634694	0.796620	0.680927
2	cvec + multi nb	0.813224	0.659184	0.822078	0.667992
3	tvec + multi nb	0.853442	0.673469	0.859201	0.687500
4	cvec + extra trees	0.996592	0.618367	0.996610	0.599572
5	tvec + extra trees	0.996592	0.648980	0.996610	0.629310

1 = Target Class (r/SuicideWatch)
0 = Negative Class (r/depression)

- Contextual, story-based words have the least predictive power.
- Strong extreme words that connotes action and death, have the most predictive power.



CONCLUSION

Recommendations

- More data
- Explore other models/API libraries such as Spacy or Word2Vec or AdaBoost and SVM
- Still needs a trained human moderator to come in to support and examine the classified posts

… because Misclassified texts are not (entirely) the model's fault to blame.

	selftext_title_clean	is_suicidal	pred	correctly_classified
	girlfriend attempted three time survived still badly med work spends day without sleep good since past month cut also recognise thing proud feel worthless try tell amazing often try mke understand brave going nothing work way better communicating make feel better trying best think anything	0	1	False

I can understand why my model predicted it to be suicidal as suicidal posts tend to be shorter and also express very bleak and dark emotions.

In the eg above, the post is short, bleak and narrates a suicidal scenario, but is actually posted under r/depression