Suicide Prediction

Juhee Sung-Schenck

TABLE OF CONTENTS

01

PROBLEM

Increasing suicides

02

OBJECTIVES

- Impact of COVID19
- Suicide Prediction Model

03

PROCESS

Overall steps of this project

04

DATA

Exploratory Data Analysis

05

FINDINGS

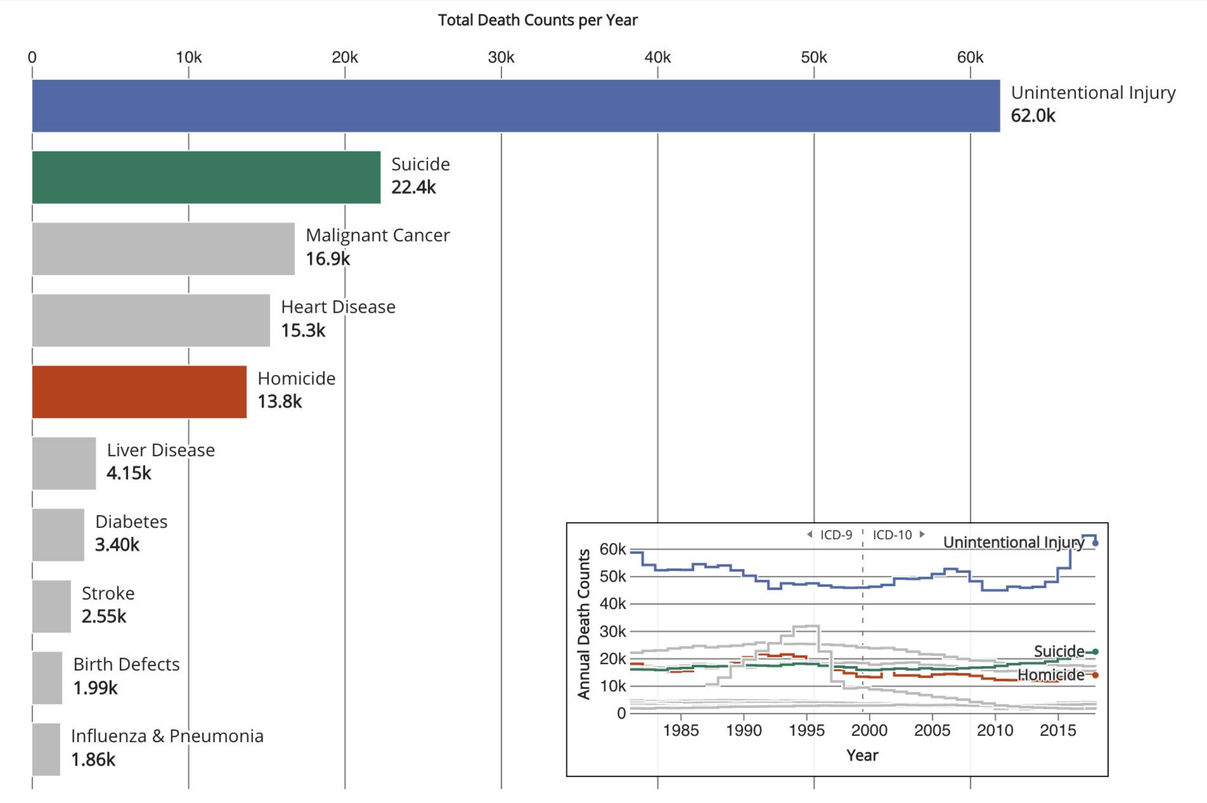
Modeling result

06

TAKEAWAY

Conclusion and resources

PROBLEM



Age 1 to 44 group, 2018

Source:
CDC website

OBJECTIVES

IMPACT OF COVID19

Find how COVID19 has changed our daily lives and how we, general public, feel everyday

SUICIDE PREDICTION MODEL

Using TF-IDF vectorizer, create a label with semi-supervied learning method, then build a model to predict whether the post is suggestive of suicide

PROCESS

01

DATA COLLECTION

02

DATA CLEANING

03

FEATURE ENGINEERING

04

DATA ANALYSIS

05

UNSUPERVISED LEARNING

06

SEMI-SUPERVISED LEARNING

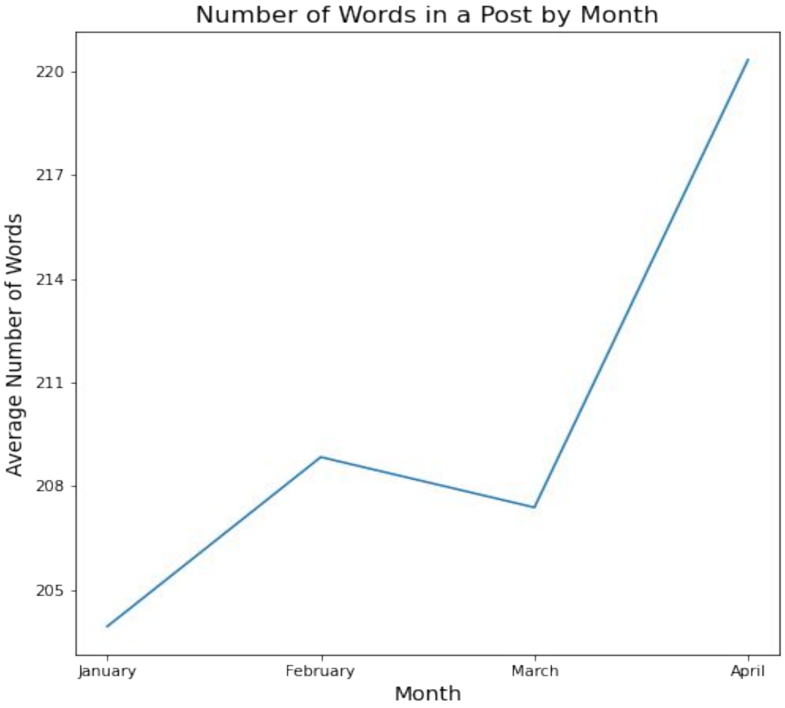
07

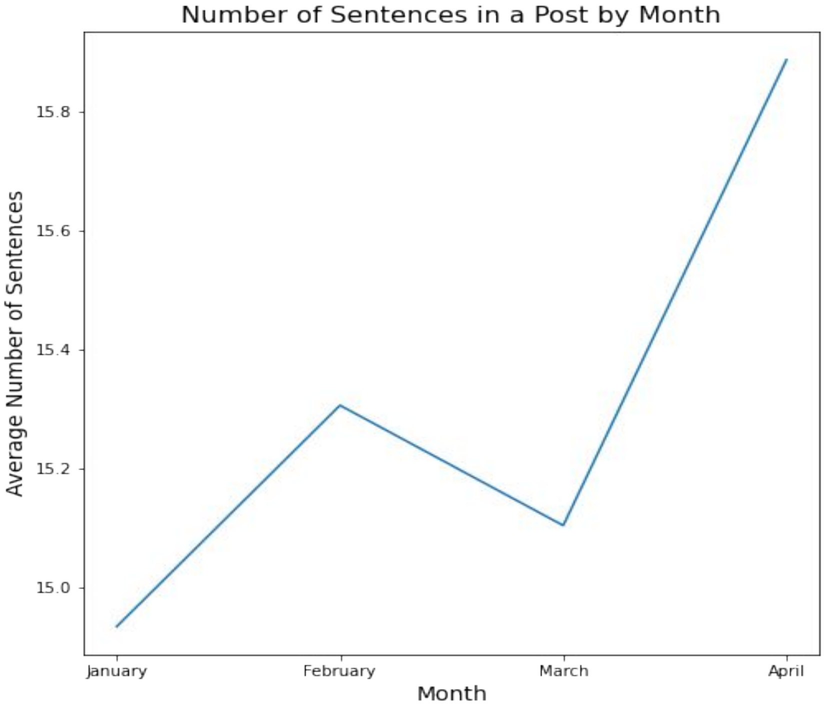
MODELING

08

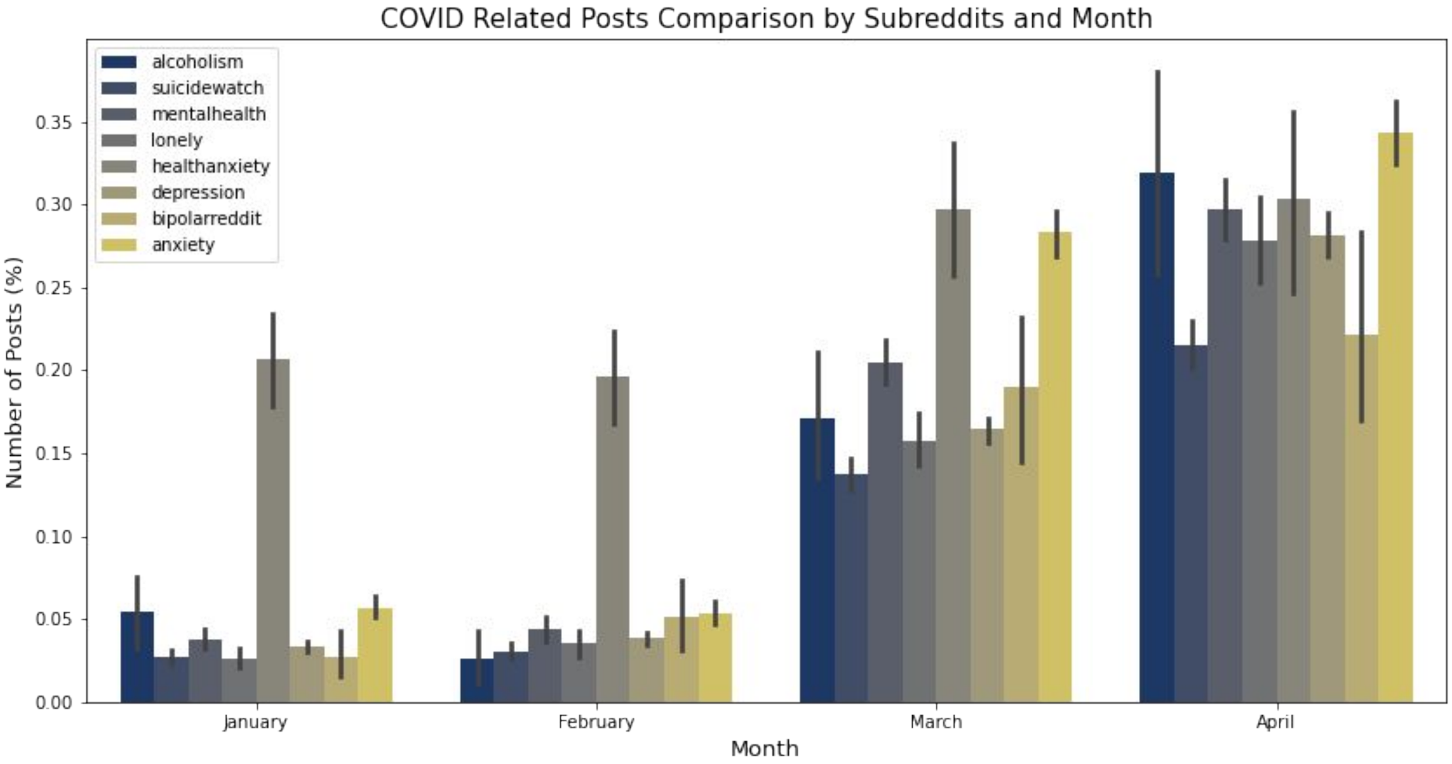
MODEL SELECTION

DATA ANALYSIS

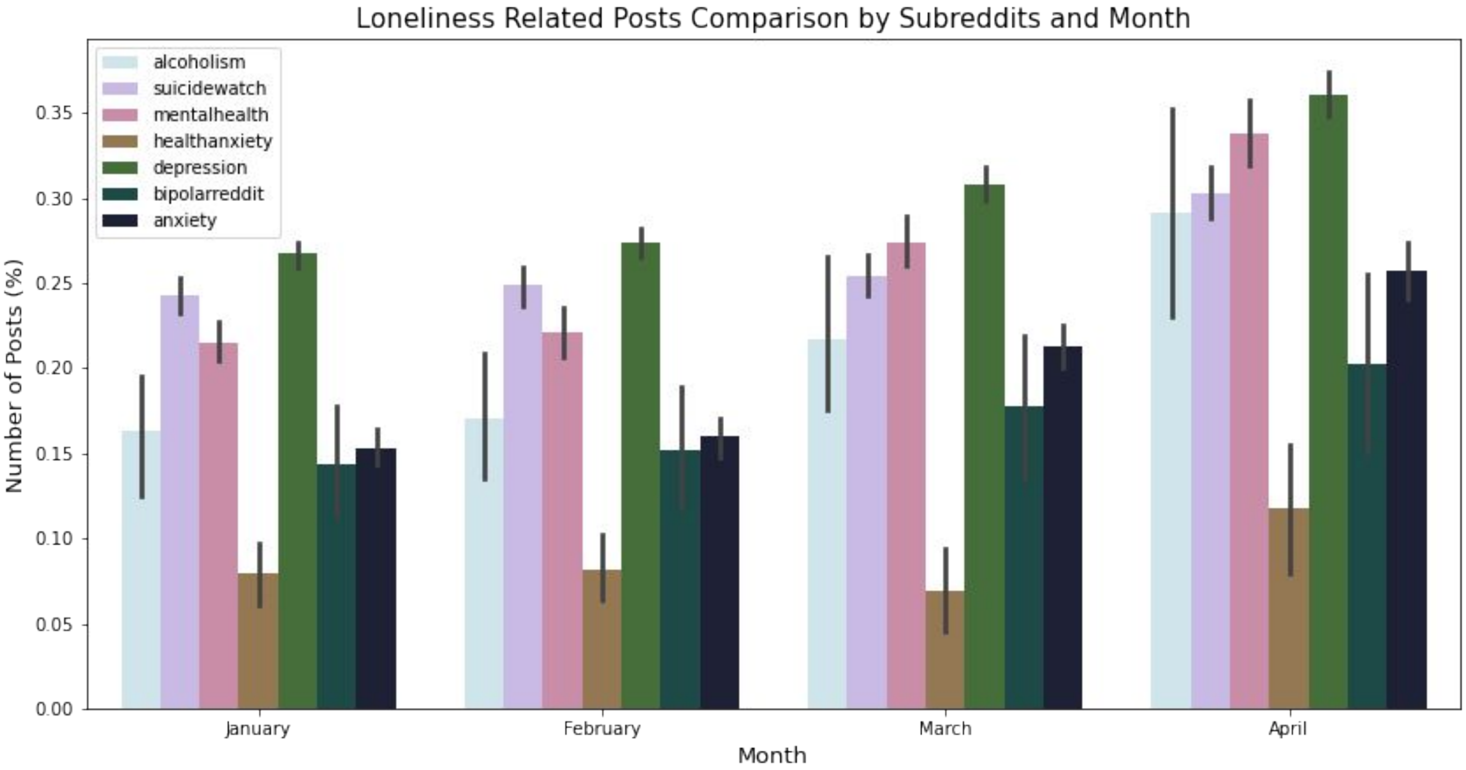




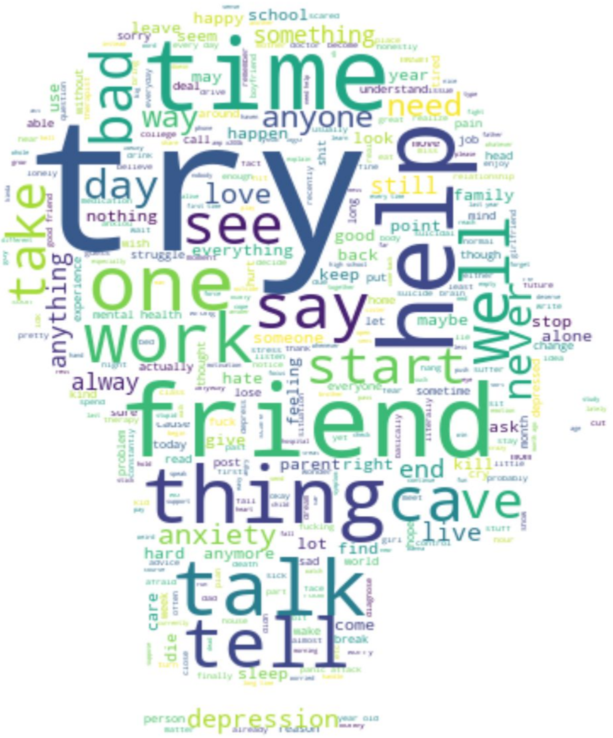
DATA ANALYSIS

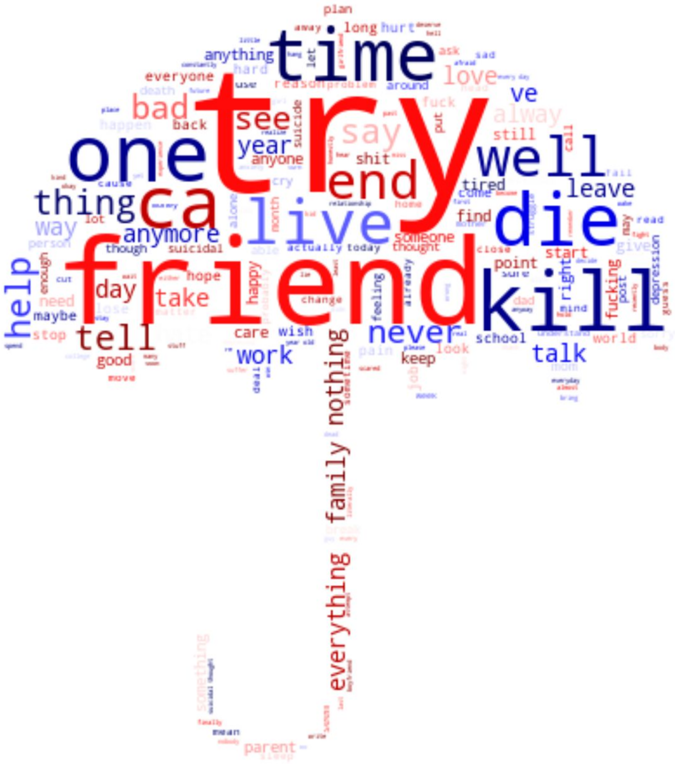


DATA ANALYSIS

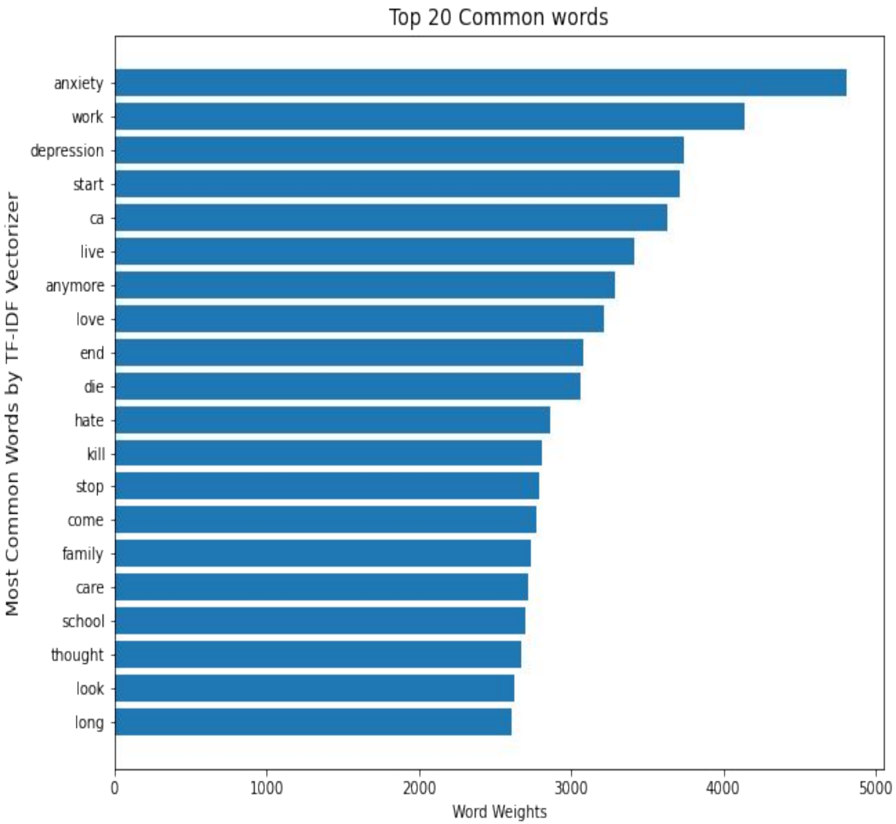


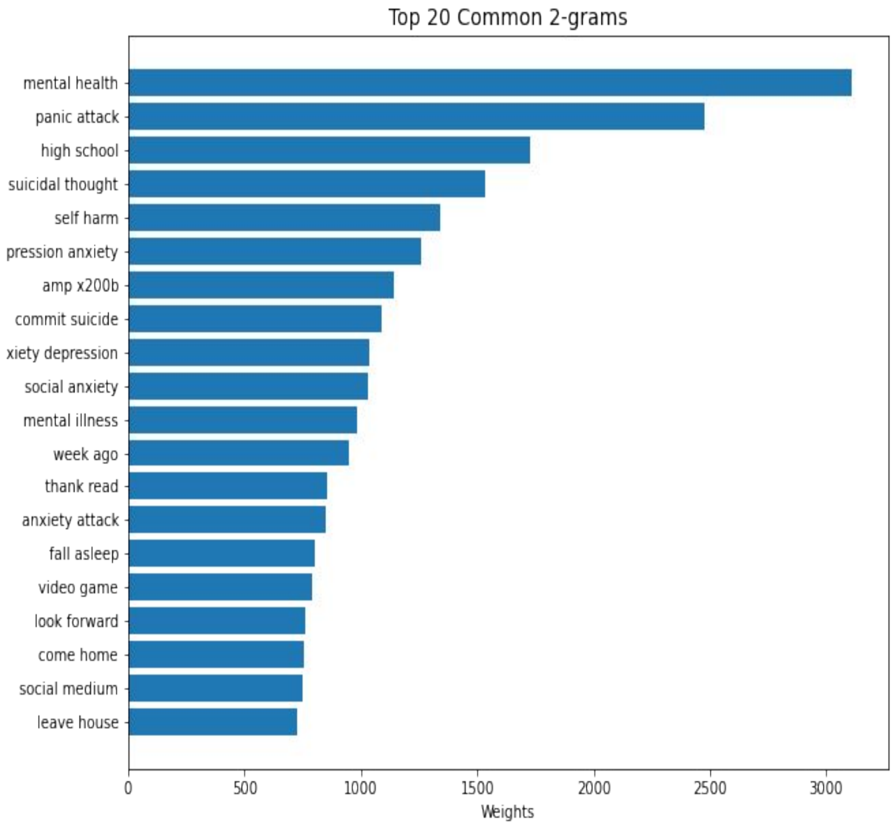
DATA ANALYSIS


Most Common Words


Most Common Words in Suicidewatch

PREPROCESSING





UNSUPERVISED LEARNING

- t-SNE: No definitive clusters detected
- KMeans: Maximum silhouette score of 0.94
  - → 3 clusters, unsuccessful
- Feature Agglomerative Clustering: Grouping every single vector

» Homogeneous data

SEMI-SUPERVISED

Label Spreading

Subreddit	Label
Alcoholism	0.06285
Anxiety	0.107574
Bipolar	0.074561
Depression	0.217601
Health Anxiety	0.055624
Lonely	0.175487
Mental Health	0.151621
Suicide Watch	0.339623

Semi-Supervised Learning:

Uses a combined dataset of a small amount of labeled data with a large amount of unlabeled data during training to create pseudo labels.

Label Spreading:

Minimizes a loss function that has regularization properties, as such it is often **robust to noise**. The algorithm iterates on a modified version of the original graph and normalizes the edge weights by computing the normalized graph Laplacian matrix.

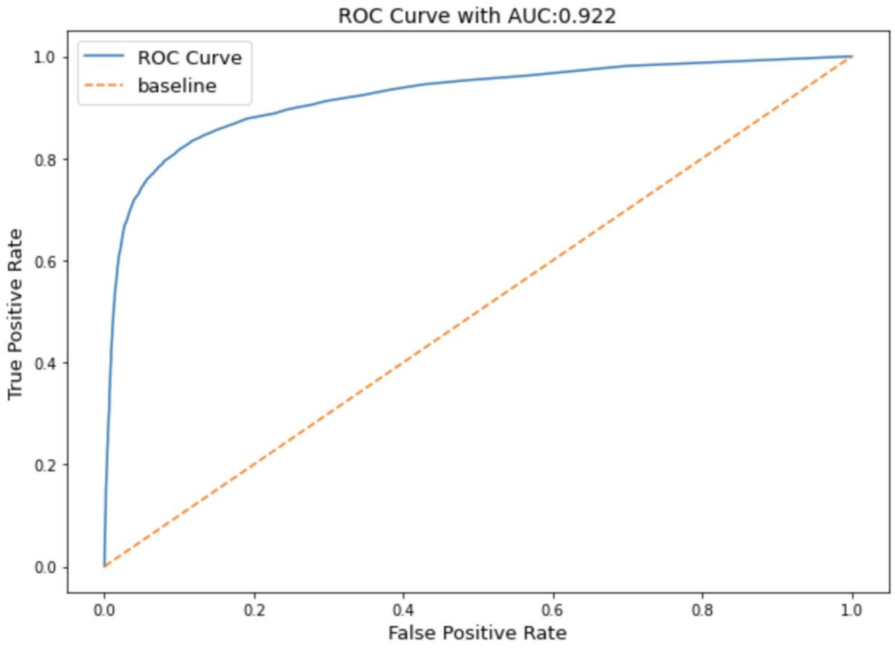
MODEL PERFORMANCE

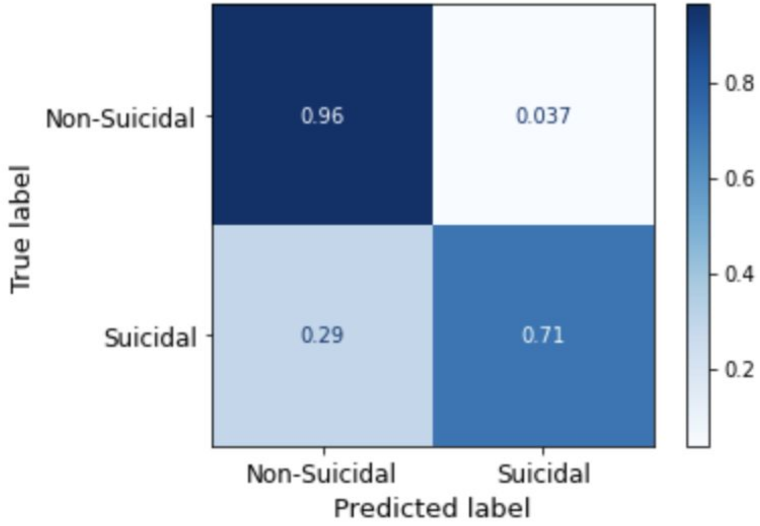
NEURAL NETWORK	Bidirectional LSTM	CNN LSTM
Train Loss	0.329	0.322
Test Loss	0.377	0.400
Train Accuracy	0.857	0.860
Test Accuracy	0.835	0.830

MODEL PERFORMANCE

CLASSIC CLASSIFIERS	Support Vector Machine		XGBoost	
	Non-Suicidal	Suicidal	Non-Suicidal	Suicidal
Precision	0.93	0.85	0.93	0.83
Recall	0.97	0.73	0.96	0.71
F1 Score	0.95	0.79	0.95	0.73

BEST MODEL


ROC Curve with AUC:0.922
ROC Curve
baseline
True Positive Rate
False Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0
0.0
0.2
0.4
0.6
0.8
1.0


Non-Suicidal
Suicidal
True label
0.96
0.037
0.29
0.71
Non-Suicidal
Suicidal
Predicted label
0.8
0.6
0.4
0.2

TAKEAWAY

- Pandemic’s impact is greater for those with mental health issues
- Mental illness is not a character defect
- Need more data to interpret historical trend, especially seasonality, as well as other posts from subreddits that are not related to mental Health to look at the impact of Pandemic
- Suicidal detection is possible with enough data and model training:
  still room for improvement - trying other word embeddings and adding layers or tuning parameters further would be worth experimenting
- Things to consider: privacy, ethics, and etc.

MENTAL HEALTH RESOURCES

Social Work License Map (60 Digital Resources for Mental Health)

Mental Health Gov.

National Institute of Mental Health

National Help Line (SAMHSA: Substance Abuse and Mental Health Services Administration)

Mental Health First Aid

Suicide Prevention Lifeline at 1-800-273-TALK (8255)

CITATION

Data obtained from

Low, Daniel M., Rumker, Laurie, Talker, Tanya, Torous, John, Cecchi, Guillermo, & Ghosh, Satrajit S. (2020). Reddit Mental Health Dataset (Version 01)

Reddit Mental Health Dataset

Inspiration for this project

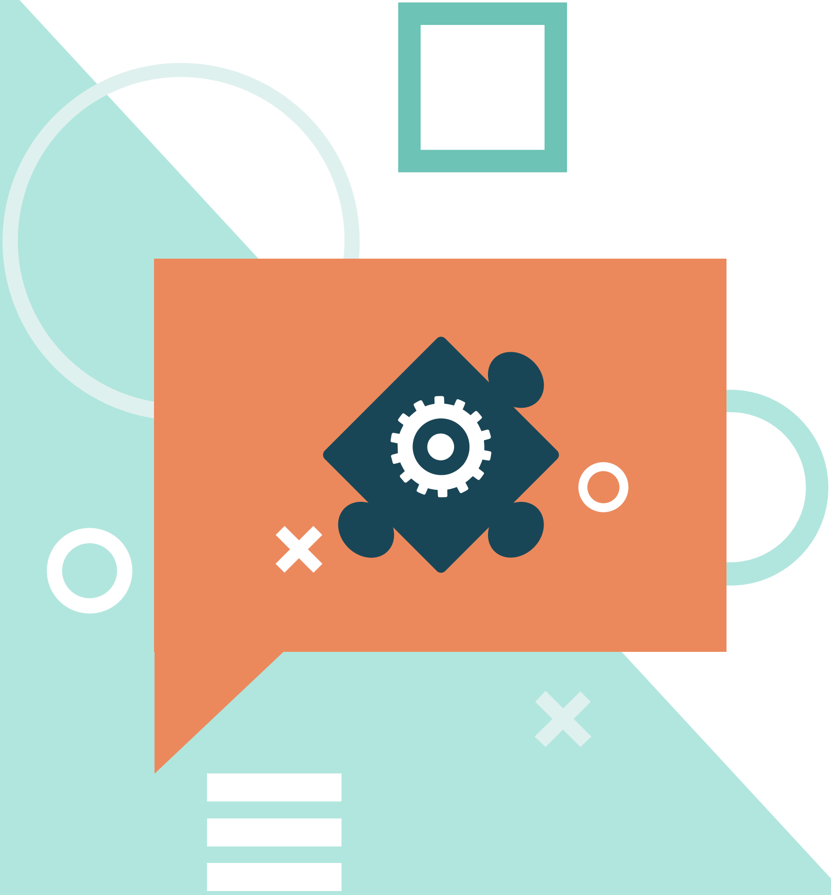
COVID-19 could lead to an epidemic of clinical depression, and the health care system isn't ready for that, either

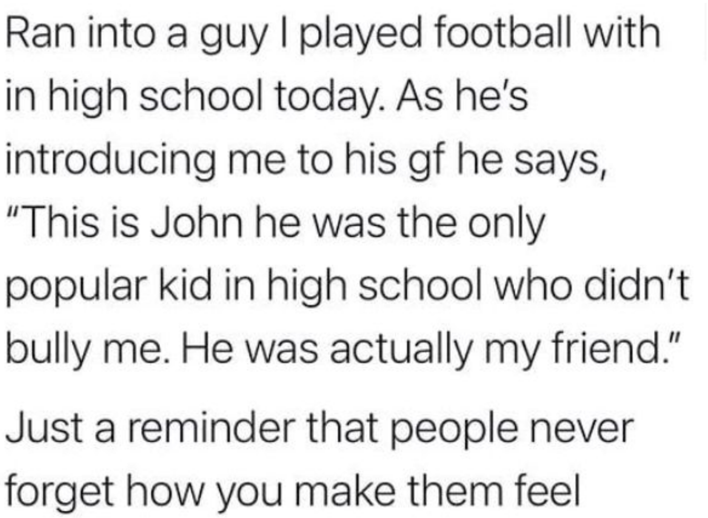
Articles about semi-supervised learning

Rediscovering Semi-Supervised Learning

More information about semi-supervised learning

sklearn.semi_supervised





Ran into a guy I played football with in high school today. As he's introducing me to his gf he says, "This is John he was the only popular kid in high school who didn't bully me. He was actually my friend."

Just a reminder that people never forget how you make them feel

THANKS

Any questions?

CREDITS: This presentation template was created by Slidesgo, including icons by Flaticon, and infographics & images by Freepik.

Please keep this slide for attribution.