Suicide Analysis

By Richard Twitty

The Problem

700000 people die from suicide every year. Some of these usually send a message to social media before they commit the act. They usually send messages on boards like Reddit suicide watch or Facebook suicide group. These messages go undetected and the individual proceeds with the dreadful act.

How can we catch these suicide content?
How can we monitor the individual history for any signs of suicidal thinking?

Who Might Be Interested

Twitter

Facebook

Reddit

TicToc

What Can Affect the suicide content

Some people can joking

Some people can be sarcastic

Some people might not be suicidal but wrote suicidal content

Data Information

Number of rows: 232074

Number of field: 3

Data: Suicide text and classification

Steps for Suicide Analysis

Build a world cloud for suicide content

Analyze the sentiment of the suicide content

Generate bigrams and trigrams of suicide content

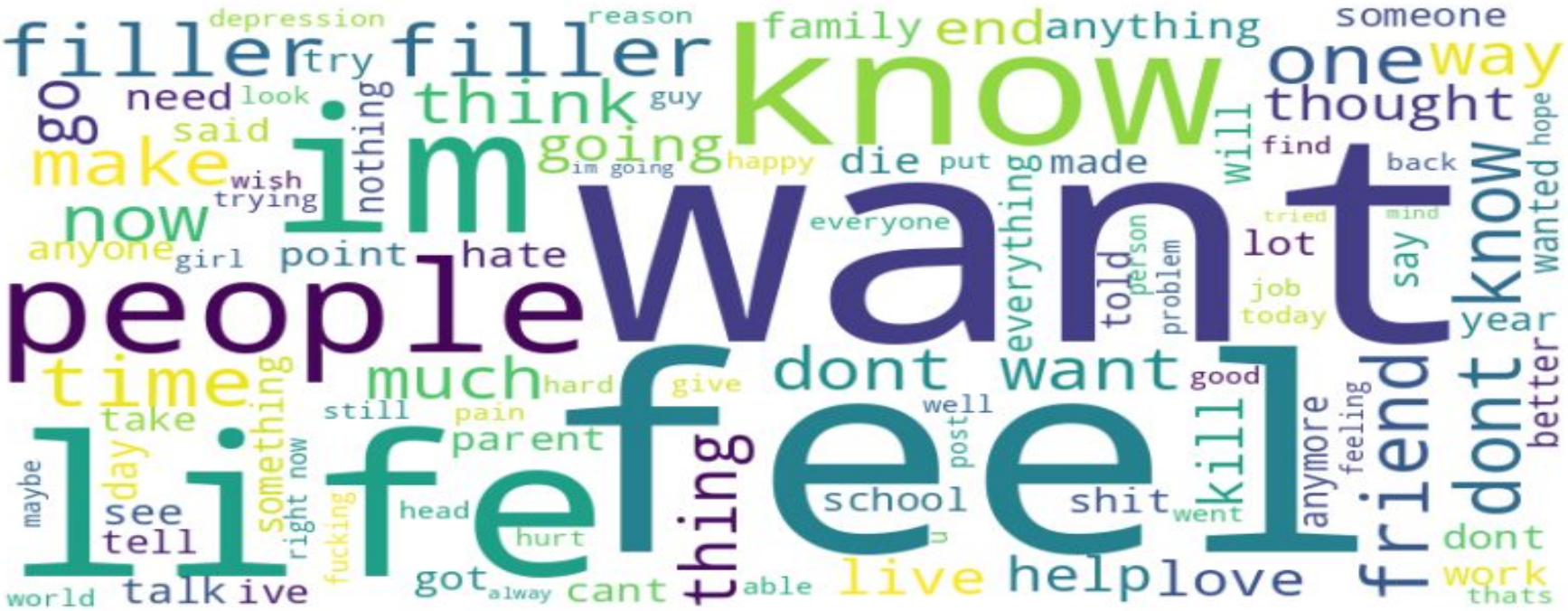
Generate a bar graph of the most common words

Generate model without grid search

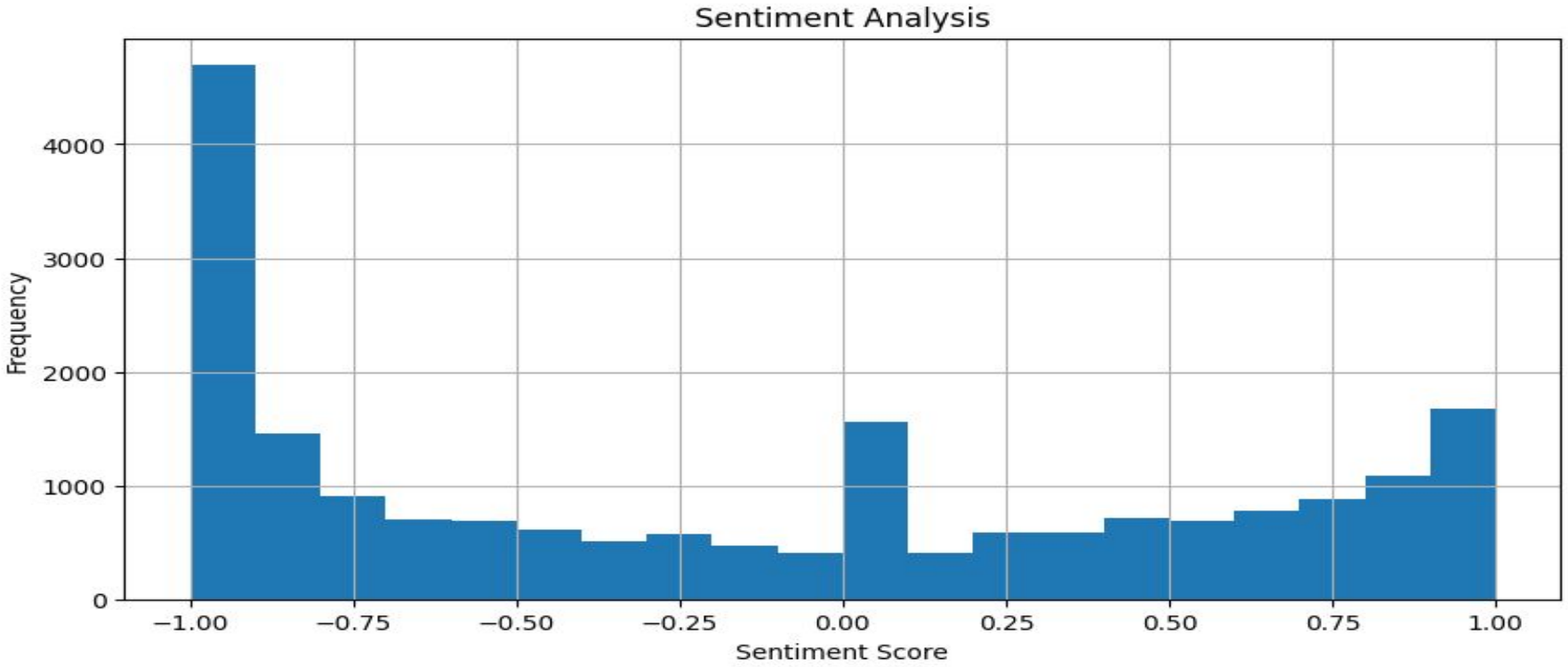
Generate models with grid search

Analyze results using graphs

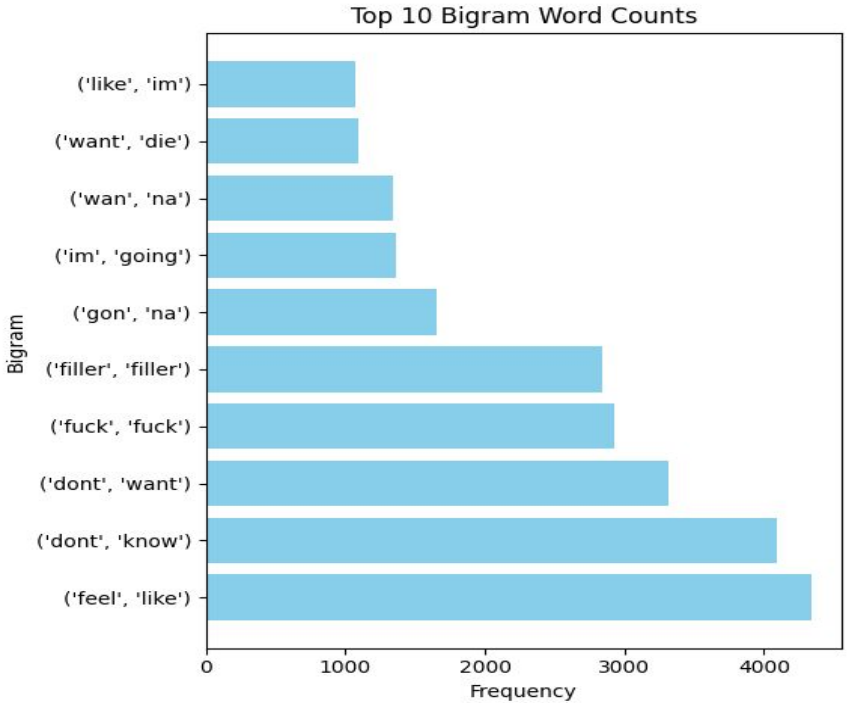
Data Exploration

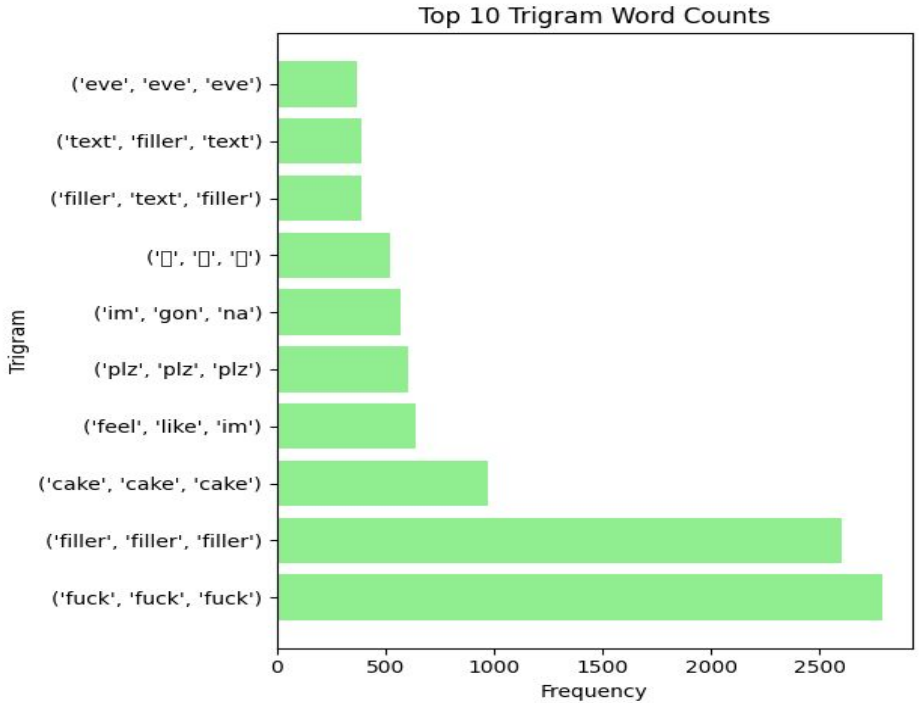


Data Exploration

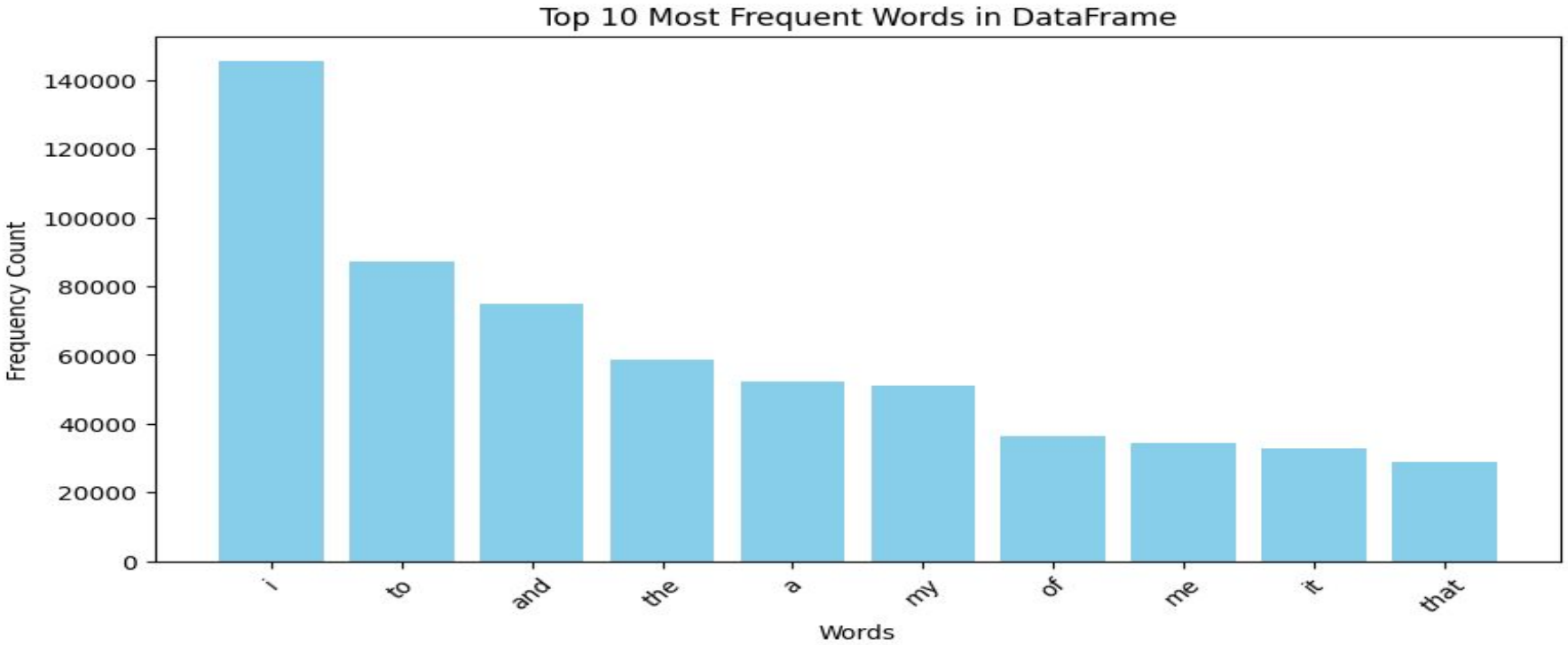


Data Exploration





Data Exploration



Model Analysis Precision Score (Without Grid Search)

Decision Trees Classifier: 0.8145

Logistic Regression: 0.88625

Support Vector Classifier: 0.90225

Gradient Boosting Classifier: 0.89325

Gaussian Naive Bayes: 0.8375

Random Forest Classifier: 0.89575

Model Analysis Precision Score (With Grid Search)

Decision Trees Classifier: 0.849

Logistic Regression: 0.883

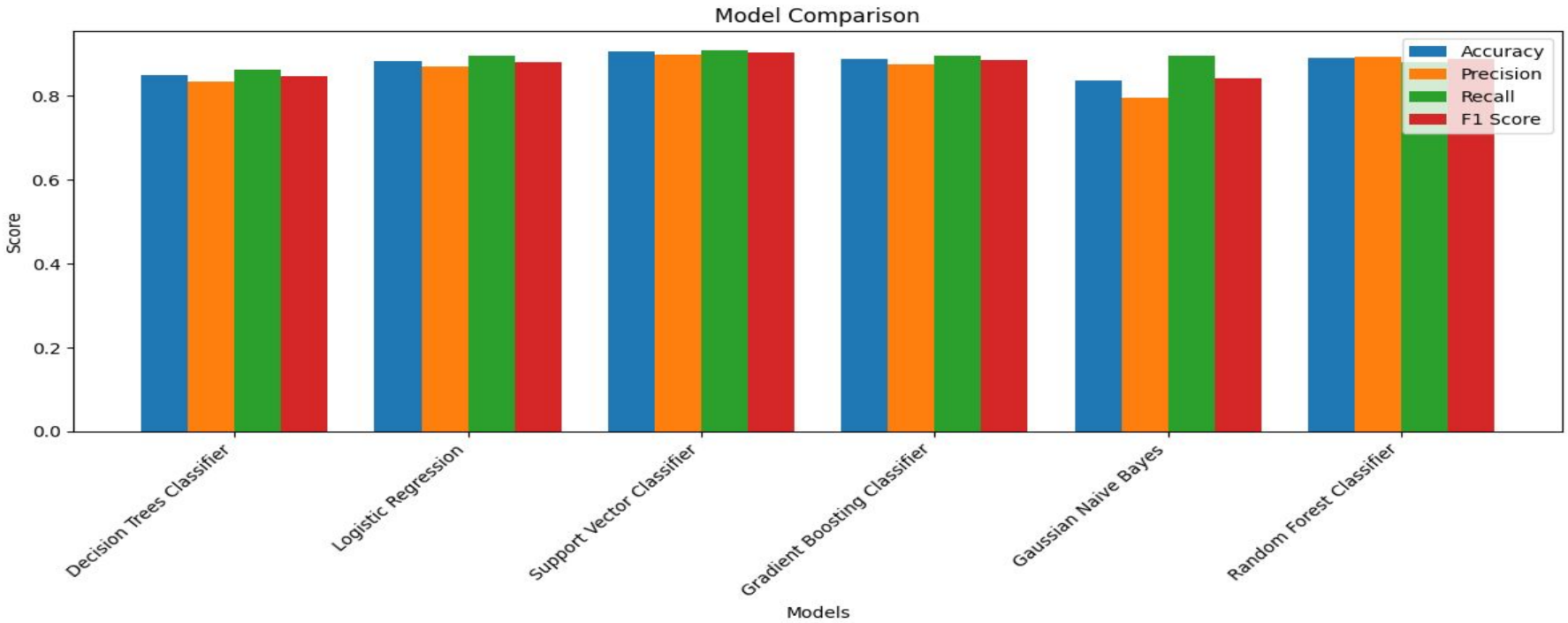
Support Vector Classifier: 0.90575

Gradient Boosting Classifier: 0.888

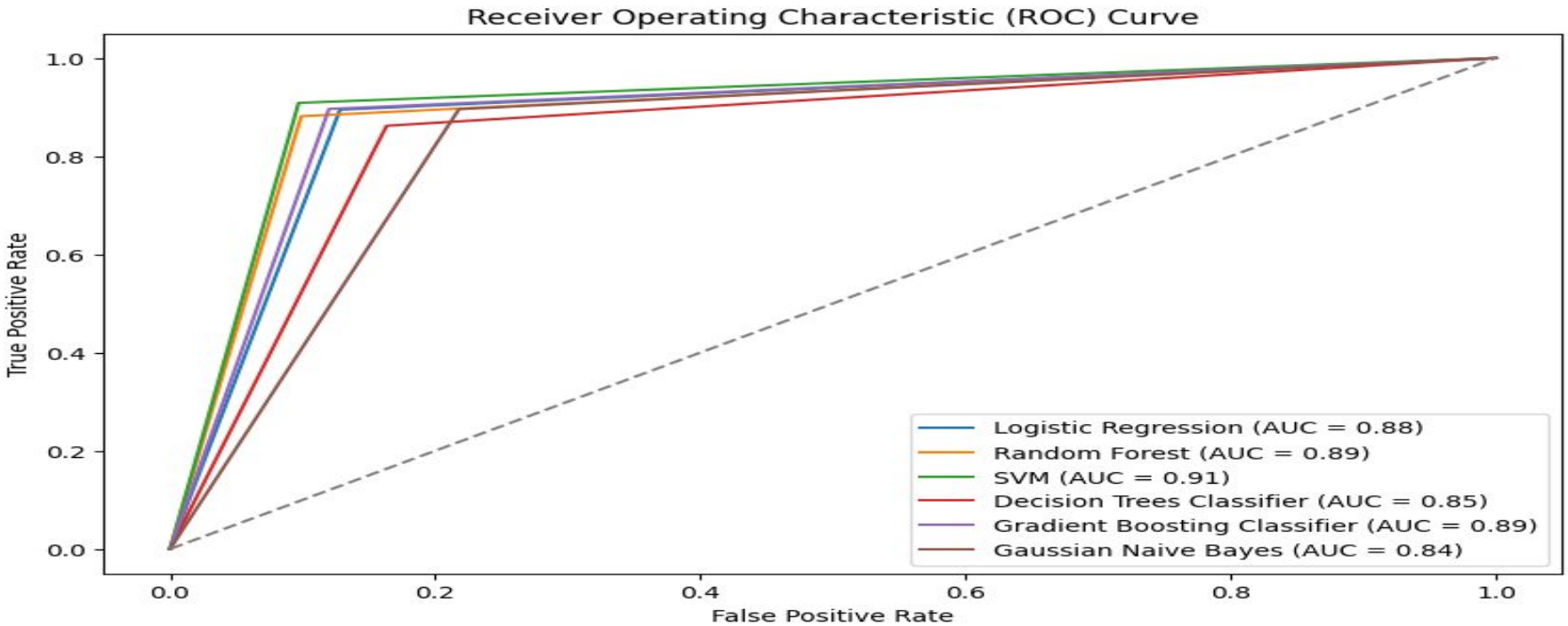
Gaussian Naive Bayes: 0.8375

Random Forest Classifier: 0.8915

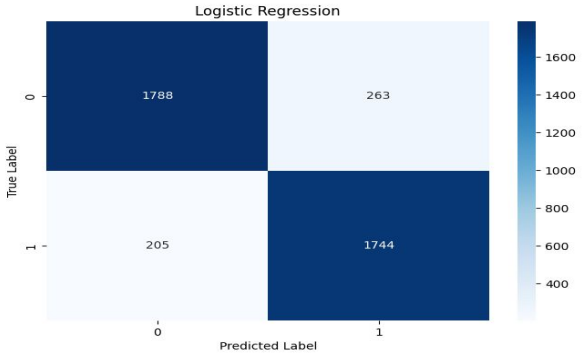
Model Analysis

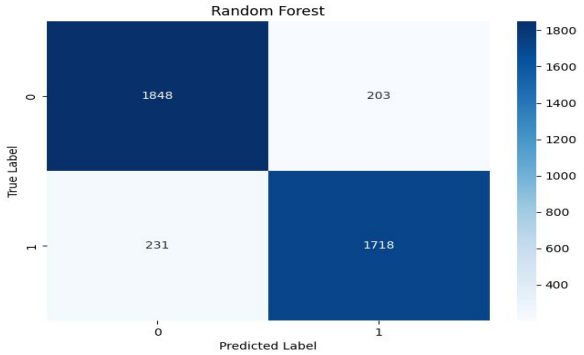


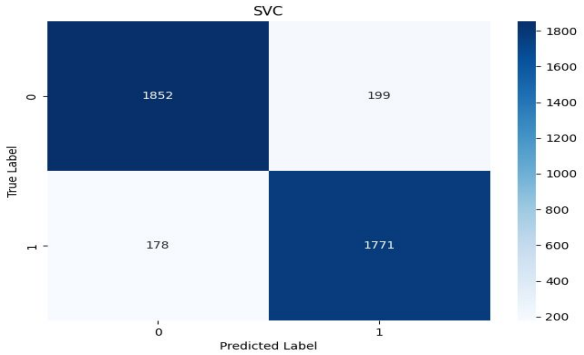
Model Analysis

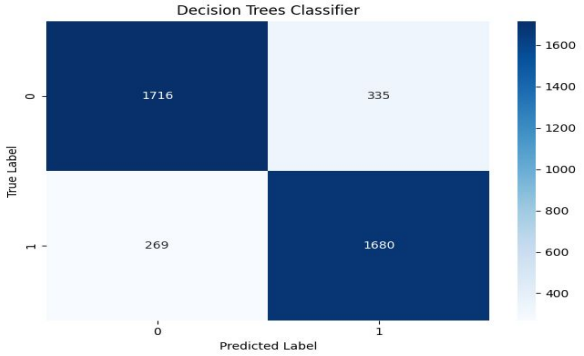


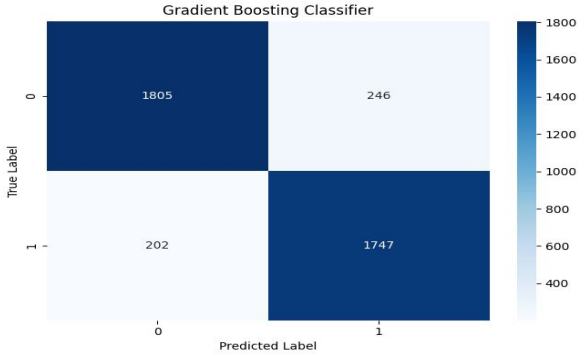
Model Analysis

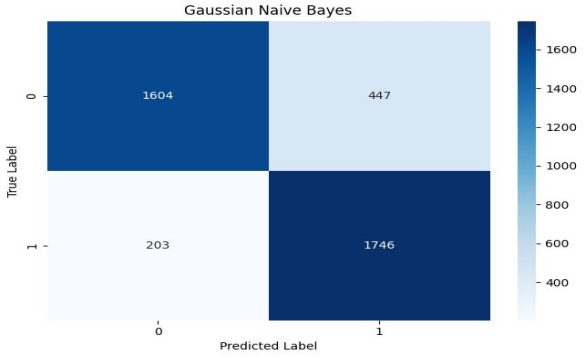












Conclusion

Support vector classification is the best model in with grid search and without grid search.

Future Research

Using sentiment as feature variable

Using grammatical structure as feature variable

Reference

https://www.kaggle.com/datasets/nikhileswarkomati/suicide-watch