TWITTER

CLASSIFYING REAL VS FAKE TWEETS

ABOUT

The problem at hand is to develop a robust classification model that can accurately differentiate between real and fake disaster-related tweets. With the rise of social media usage during times of crisis, the spread of misinformation and fake news has become a growing concern. This can lead to confusion, panic, and even impede emergency response efforts. Therefore, it is crucial to develop a machine learning-based solution that can identify and filter out fake disaster tweets from real ones. The classification model should be able to handle large volumes of data, account for variations in language, and be able to generalize to different types of disasters. The solution is designed for twitter team that works with emergency responders in identifying real disasters

TABLE OF CONTENTS

ID	A UNIQUE IDENTIFIER FOR EACH TWEET
TEXT	TEXT OF THE TWEET
LOCATION	LOCATION WHERE IT IS TWEETED FROM
KEYWORD	A PARTICULAR KEYWORD FROM THE TWEET
TARGET	0 OR 1 WHETHER THE TWEET IS REAL OF FAKE

TABLE OF CONTENTS

ID	KEYWORD	LOCATION	TEXT	TARGET
1	scream	USA	Our deeds are reason for EQ	1
2	fire	CANADA	Forest fire near La Ronge Sask	1
3	attack	INDIA	Militants attack police post in Udhampur	1

TABLE OF CONTENTS

Kaggle Dataset

Training Data : 7613 rows, 5 colums

Testing Data : 3263 rows, 4 colums

Missing Data: 2533 in location and 61 keywords

TOOLS

- Python
- Regular Expressions
- Lemmatizer from NLTK
- Tweet Tokenizer from NLTK
- Spacy

Data Visulization

- Matplotlib
- Seaborn
- Wordcloud

Gridsearch

Randomized Search

ALGORITHMS

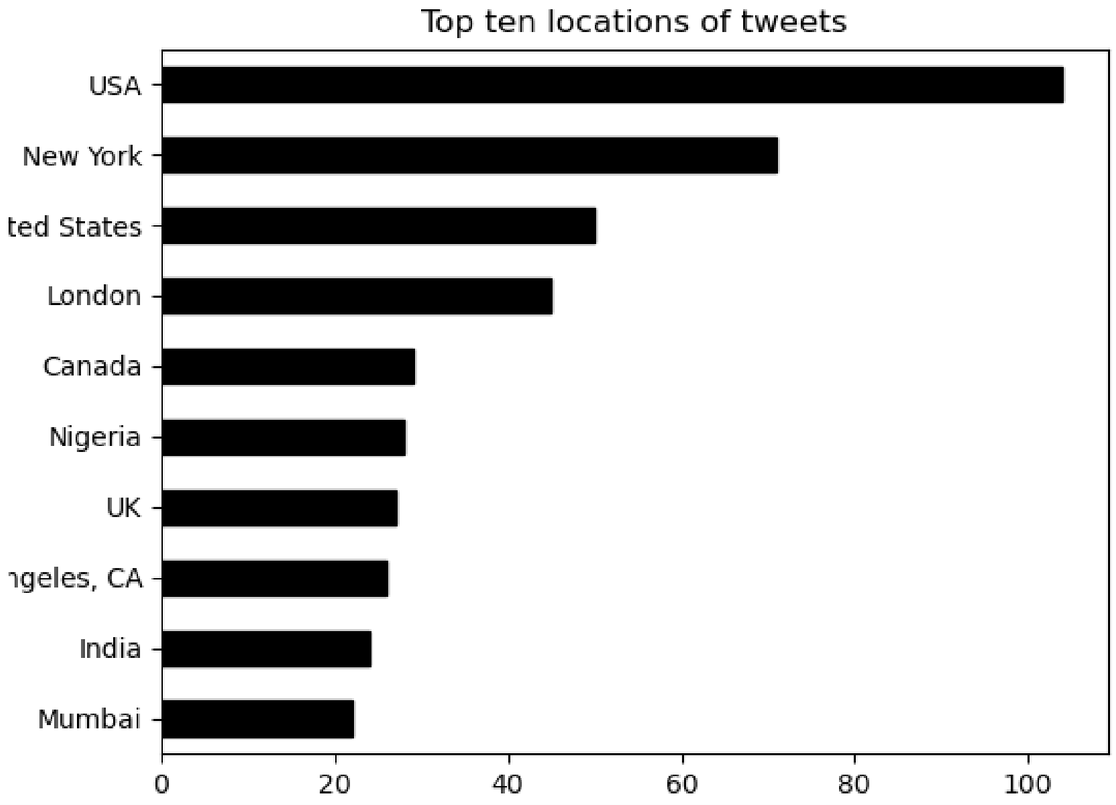
Natural Language Processing

- TFIDF Vectorizer
- Counter Vectorizer

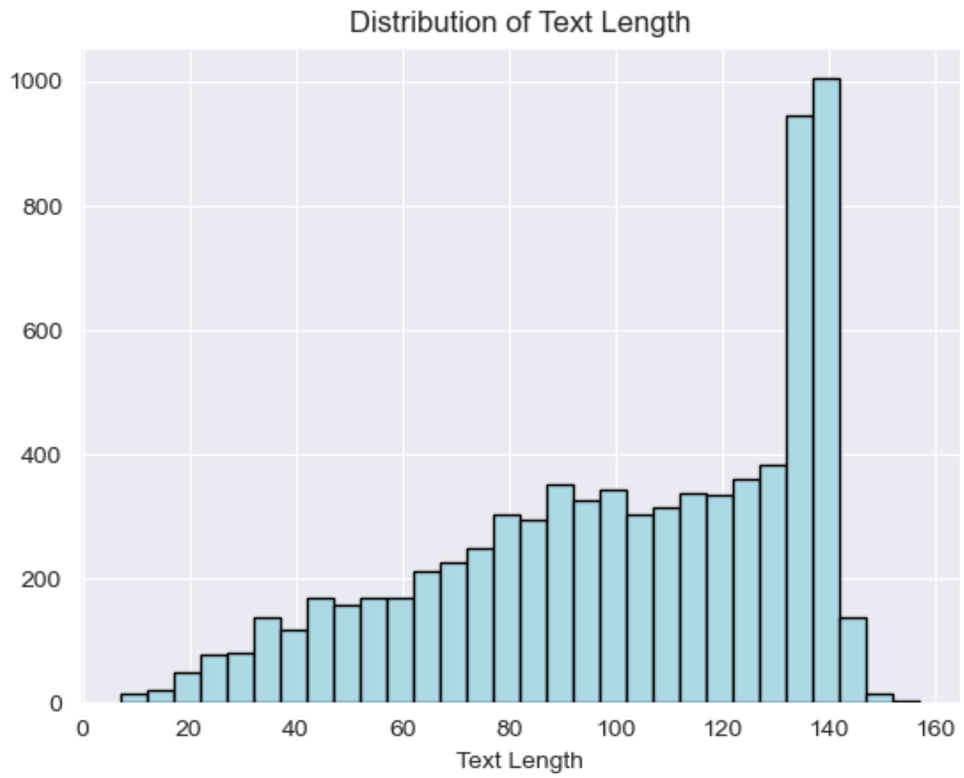
Supervised Learning

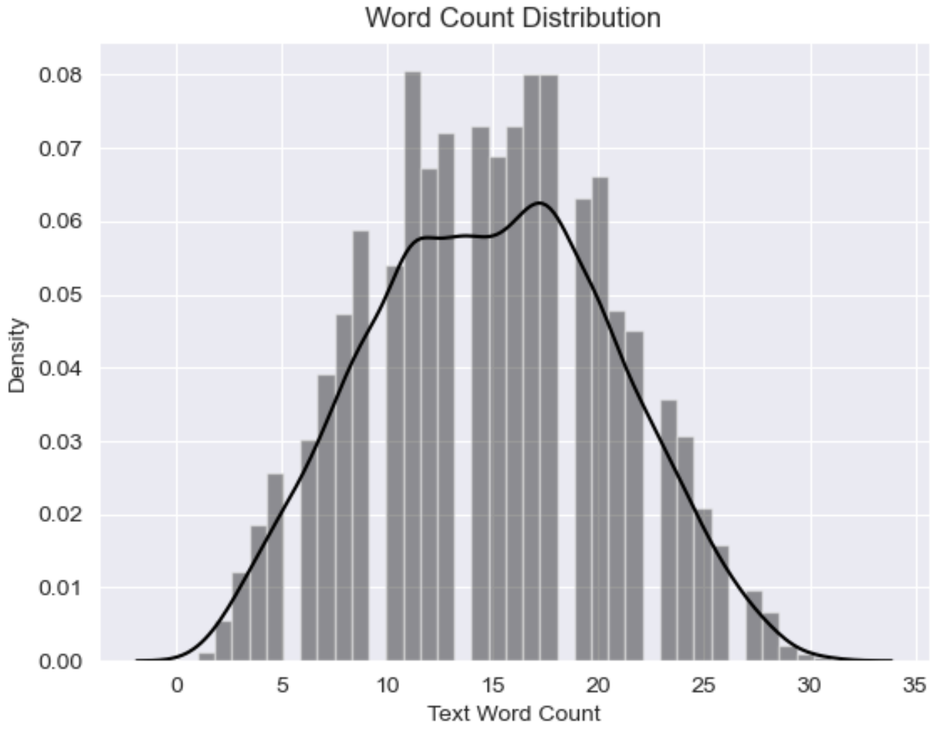
- Multinomial Naive Bayes Classifier
- Logistic Regression
- Random Forest Classifier
- XGBoost Classifier

LOCATIONS

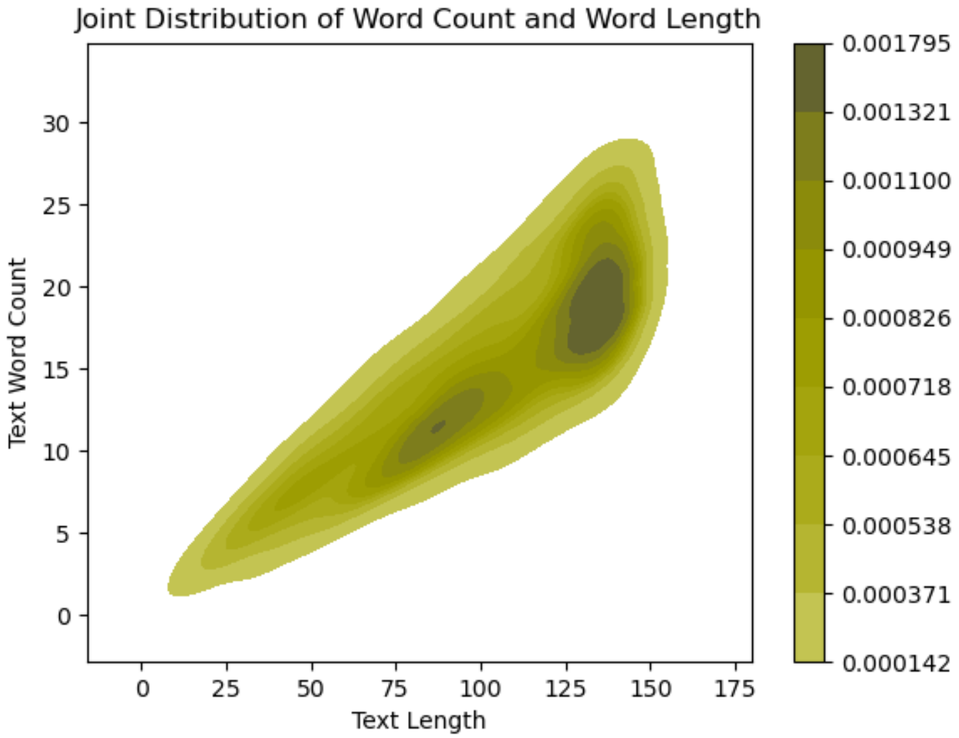

Top ten locations of tweets
USA
New York
ted States
London
Canada
Nigeria
UK
ngeles, CA
India
Mumbai
0
20
40
60
80
100

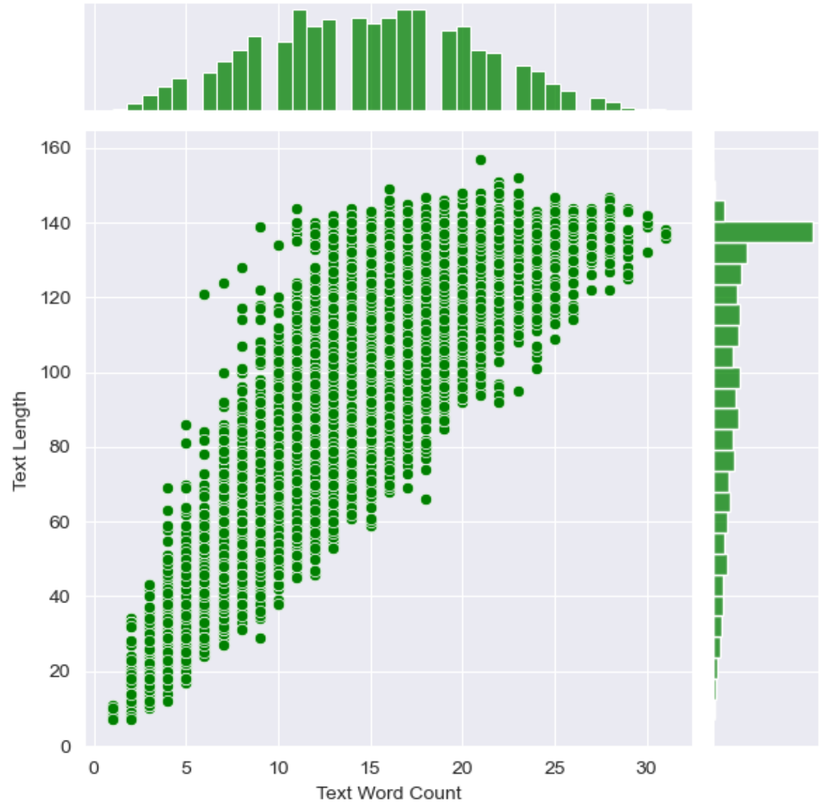
DISTRIBUTIONS



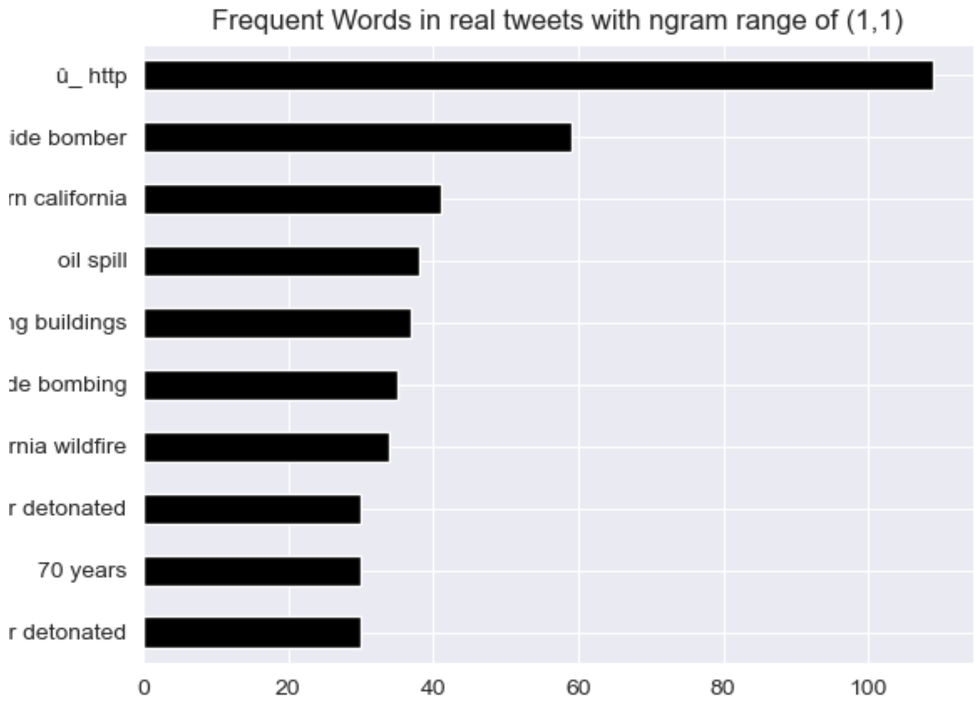


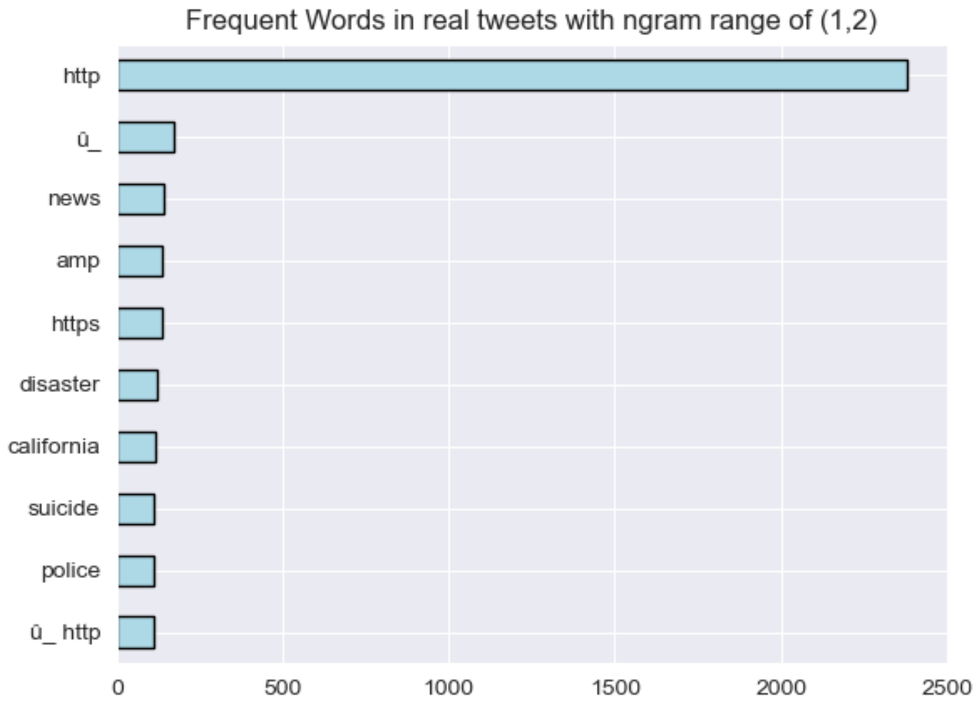
DISTRIBUTIONS



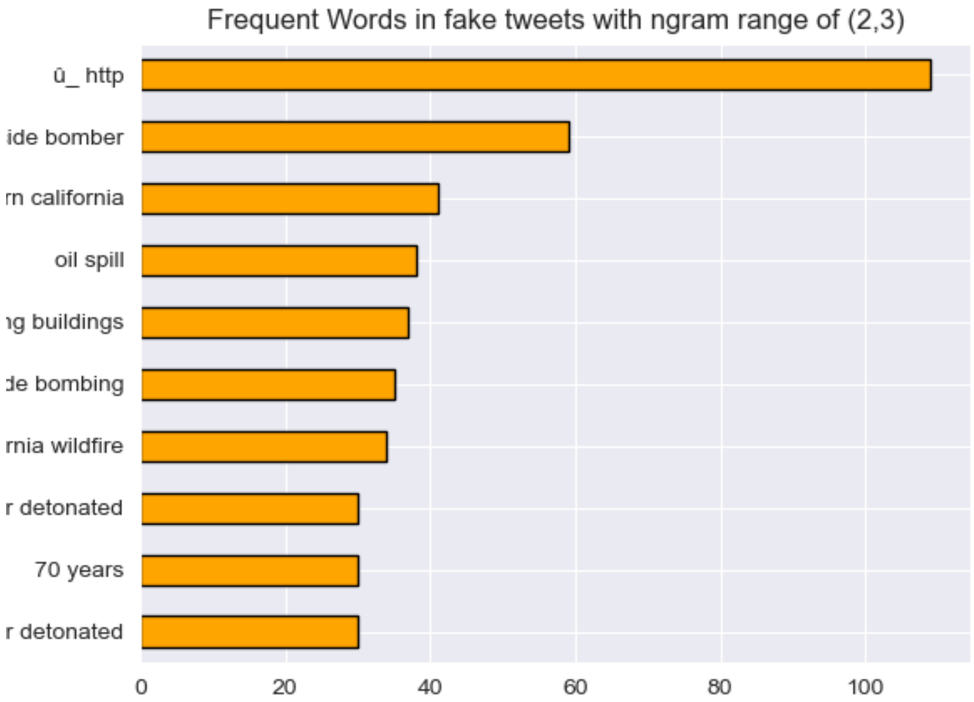


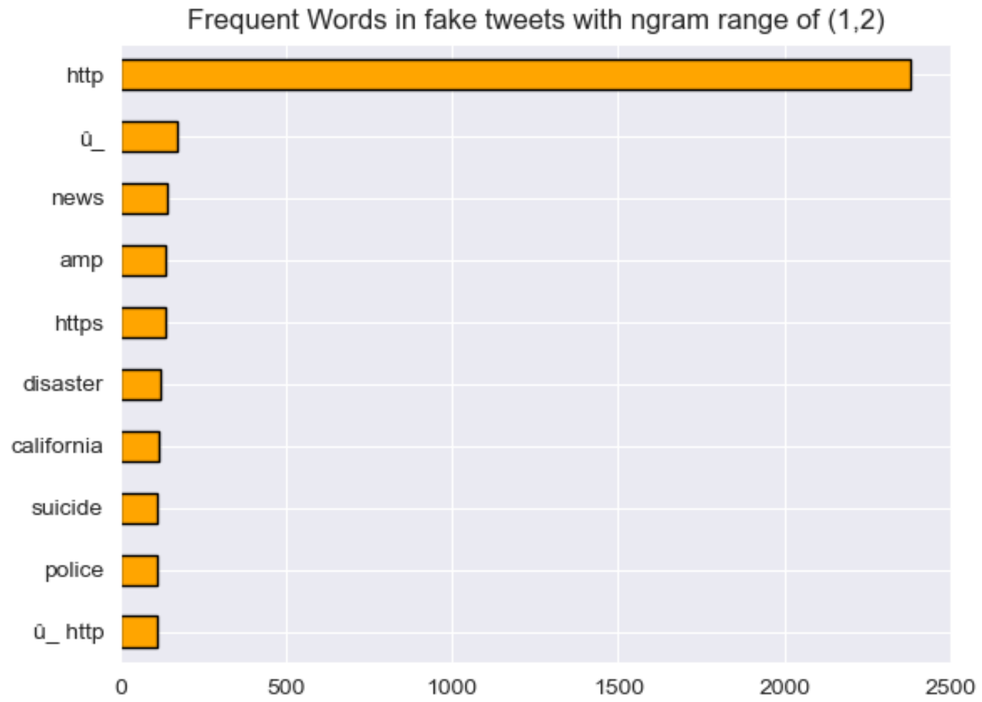
FREQUENT WORDS





FREQUENT WORDS

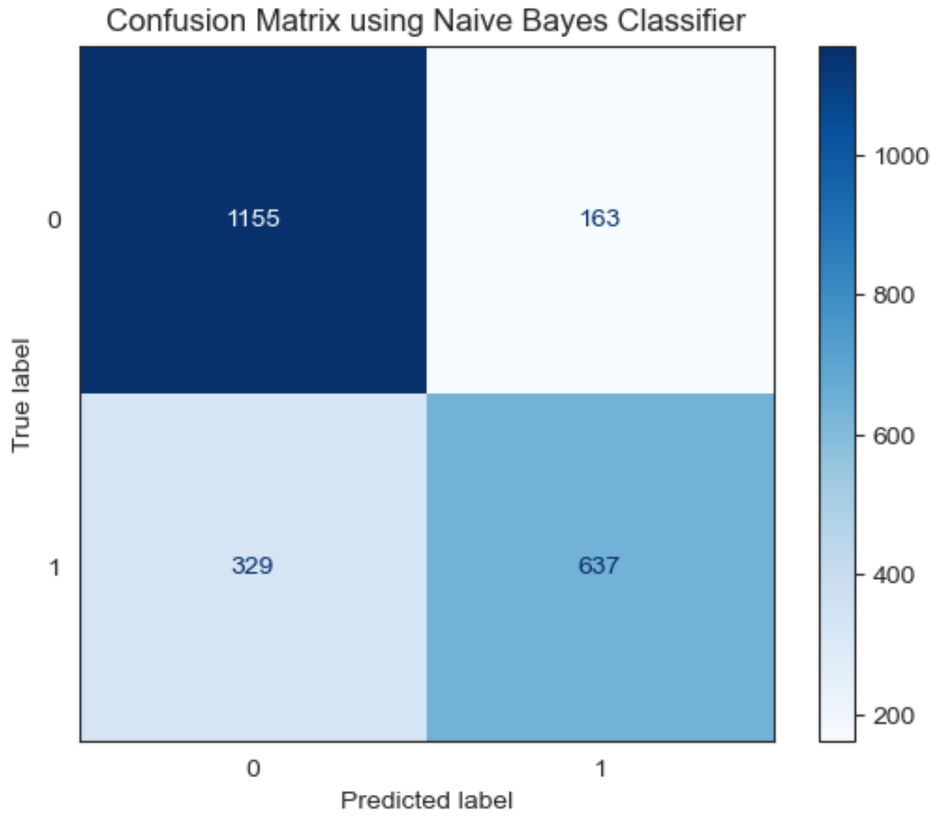


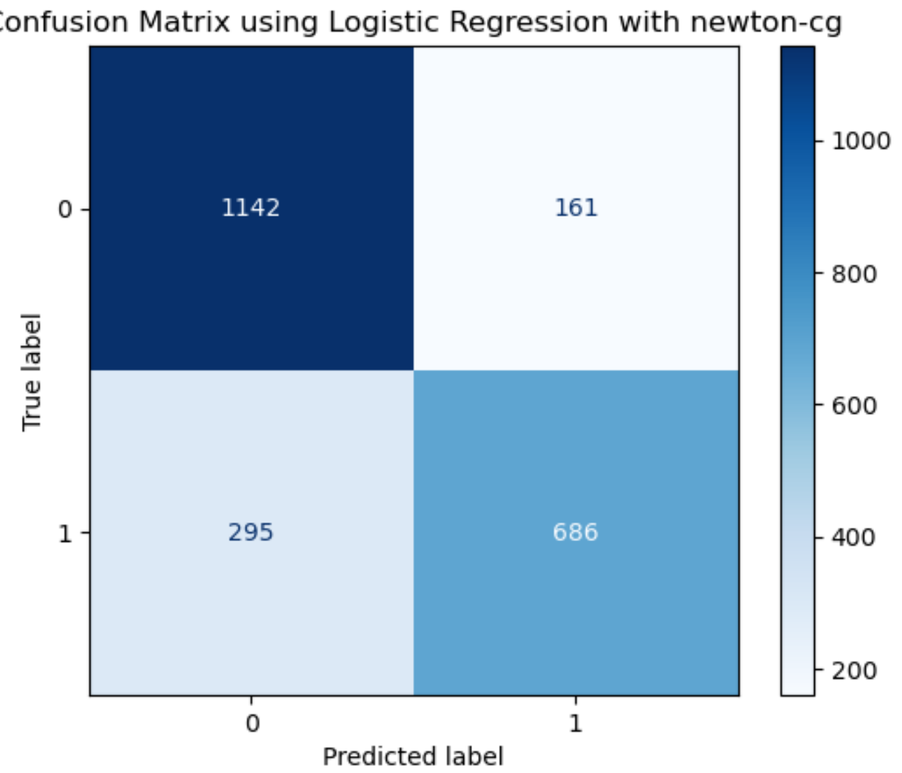


MODELS

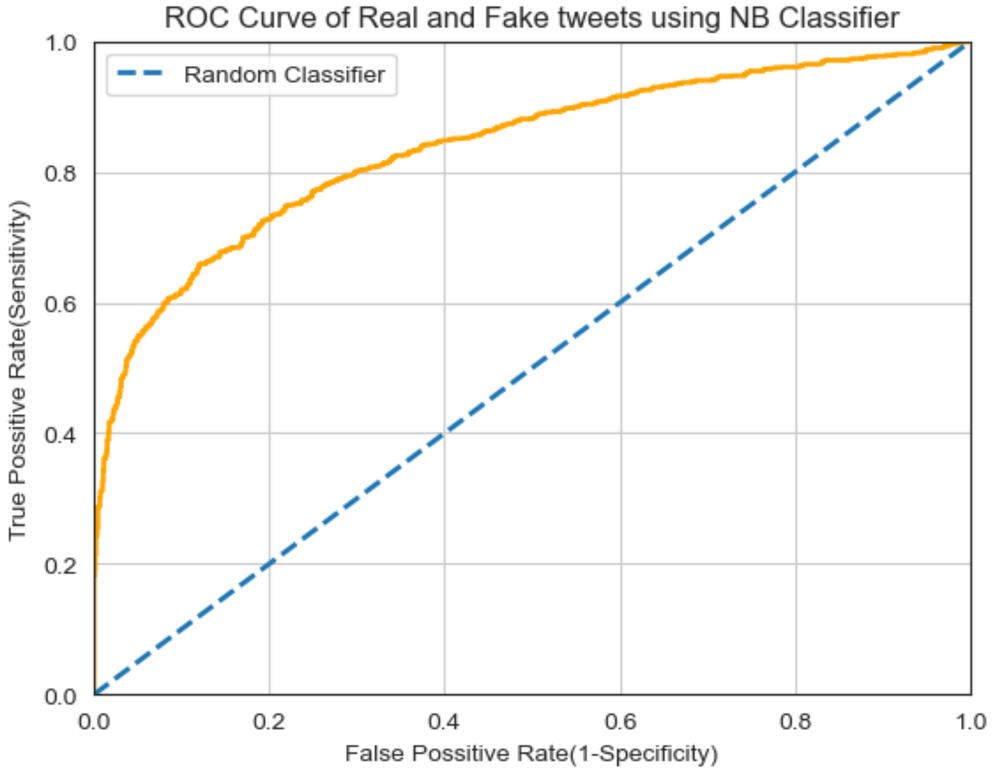
MODEL	TRAINING SCORE	TESTING SCORE
MULTINOMAIL NAIVE BAYES CLASSIFIER COUNTER VECTORIZER	81%	78%
LOGISTIC REGRESSION COUNTER VECTORIZER	85%	79%
RANDOM FOREST RANDOMSEARCHCV	84.7%	79.4%
RANDOM FOREST GRIDSEARCH	84.0%	78.5%
XGBOOST TFIDF	86.1%	78.2%
LOGISTIC WITH REGULARIZATION	83%	80%

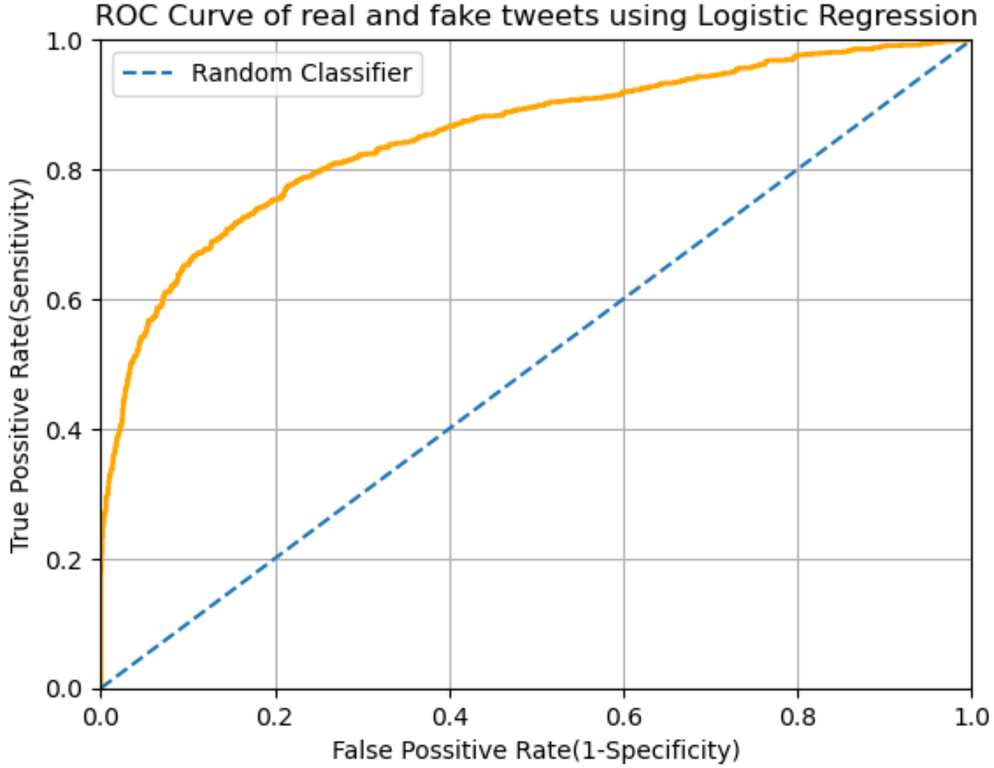
CONFUSION MATRIX





ROC CURVES


ROC Curve of Real and Fake tweets using NB Classifier
Random Classifier
True Possitive Rate(Sensitivity)
False Possitive Rate(1-Specificity)
0.0
0.2
0.4
0.6
0.8
1.0


ROC Curve of real and fake tweets using Logistic Regression
Random Classifier
True Possitive Rate(Sensitivity)
False Possitive Rate(1-Specificity)
0.0
0.2
0.4
0.6
0.8
1.0

CONCLUSION

- All the models have testing accuracy of around 78 to 80%
- Logistic Regression with L2 penalty and regularization of 1 is the best model with 80% Test Accuracy
- The model helps in mitigating the negative impact of false information and helps people and emergency responders
- Various models were trained and tested rigorously
- More data needs to be procured in order to predict more accurate models

NEXT STEPS

- **Collect more data and train and test using the models created**
- **Take feedback from emergency responders when collecting data and training new models**

Any Questions ?