A REPORT OF MINI PROJECT OF PYTHON

at

MODEL INSTITUTE OF ENGINEERING & TECHNOLOGY
(AUTONOMOUS)
(Permanently Affiliated to the University of Jammu, Accredited by NAAC with "A" Grade)

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENT FOR THE AWARD

OF THE DEGREE

BACHELOR OF ENGINEERING

(Computer Science and Engineering)

SUBMITTED BY:

NAME:Suryansh Sharma

ROLL NO.:2020A1R119

BRANCH:CSE

SEMESTER:5th A2

ACKNOWLEDGEMENT

Through this section of my report, I want to present my gratitude towards my institute MIET, from where I learned and gained an experience worth enhancing my credibility for my career. Being able to work on projects that require you to analyse and solve a problem that exists in the real world is an experience that one can achieve rarely, and I am happy and thankful to get that chance here.

I would also like to thank my mentors Prof. Ankur Gupta, Dr. Sahil Sawhney, Asst. Prof. Mekhla Sharma and Asst. Prof. Surbhi Gupta for being a guiding force throughout this semester. It was a great experience to work under your guidance as I learned a lot of new things and also got the privilege to have a good experience with practical work to do. I would also like to thank my HOD, Dr. Ashok Kumar and the CSE Department who encouraged and inspired me to learning new skills.

Lastly, I want to thank my parents and colleagues who have been a support system throughout my life and help me lessen my burden and stress of work.

Contents

PROJECT SUMMARY

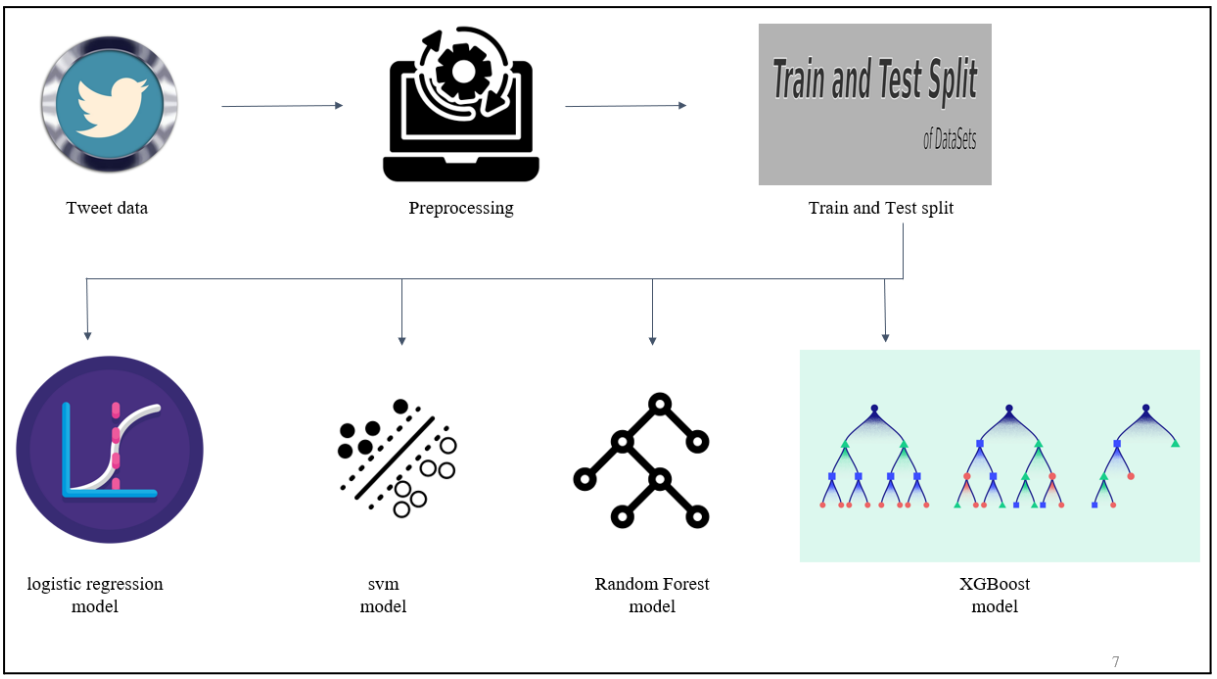
1. INTRODUCTION

- When a user wishes to voice his view on a trending topic on social media, we apply sentiment analysis to try to determine the sentiment score of that given opinion.
- Twitter is the most popular microblogging social media site, with over a billion users. Nearly 145 million people use the site on a daily basis.
- In today's world, the user tweets utilizing Hashtags, emojis, and punctuation make it difficult to examine the data and Create sentiment scores of tweets. For this project, Tweets of ukraine-russia war are used for sentiment analysis.
- In this project,we are using dataset of ukraine-russia war to train machine learning models and using the trained models to predict sentiment values for example Positive,Negative,Neutral for any tweet related to ukraine-russia war.

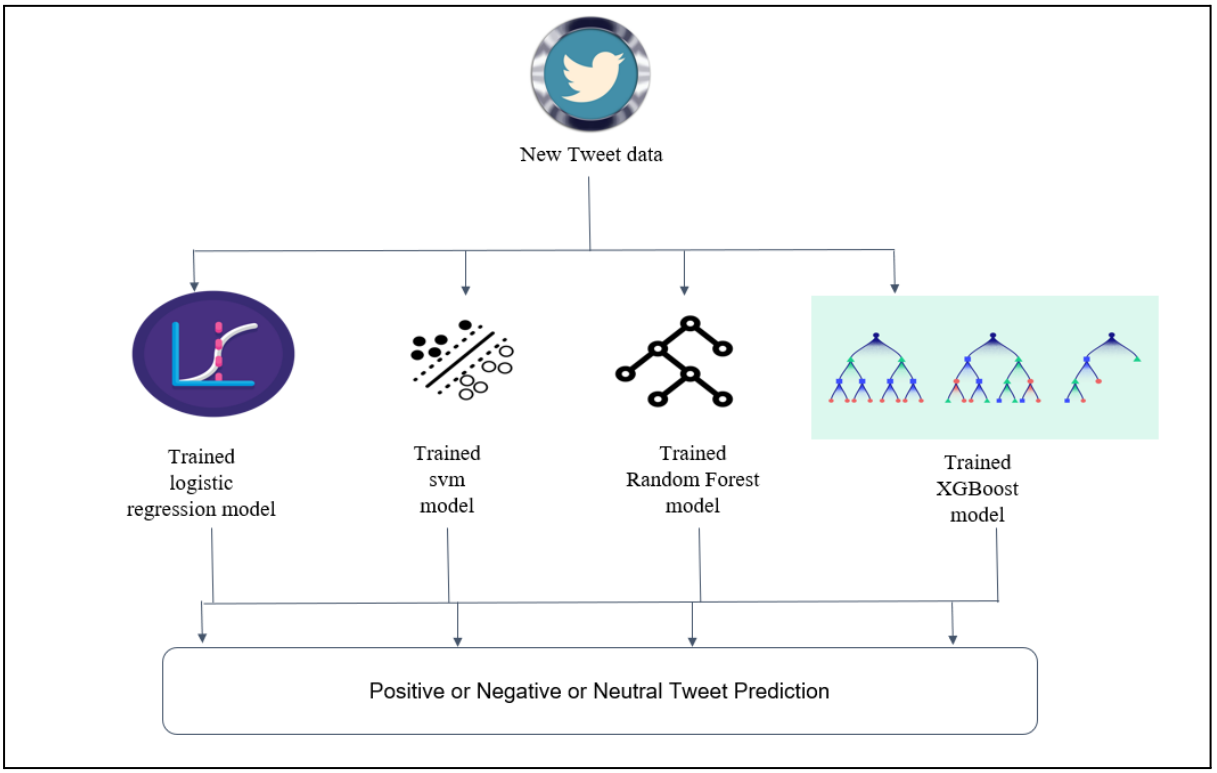
2. Workflow

We have divided the workflow into two parts:

1. Training Workflow



2. Testing Workflow



3. TECHNICAL DETAILS (Coding)

Things we are covering in this analysis are:

➢Tweets Preprocessing and Cleaning/Data Cleaning

➢Visualization from Tweets

➢Extracting Features from Cleaned Tweets from TfidfVectorizer

➢Model Building: Sentiment Analysis

1. Logistic Regression
2. Support Vector Machine
3. Random Forest
4. XGBoost

We have used libraries like numpy,pandas,matplotlib,vadersentiment for this analysis.

Here in this project we are using raw tweets from ukraine-russia war dataset and then cleaning those tweets like removing hashtags,urls etc from the raw tweets.And then using TfidfVectorizer for converting characters in tweets into numpy arrays in order to train the model.And then training logistic regression,support vector machine,random forest,xgboost models.And then testing each model prediction for a given tweet and then evaluating which model is giving highest accuracy.

(Outputs)

Prediction outputs of each machine learning model as follows:

1. Logistic Regression

In [21]:
```
input_tweet=["Ukraine army will defeat the russian army"]
predictions=vectorizer.transform(input_tweet)
result=logistic_model.predict(predictions)
if result[0]==0:
  print("Postive")
elif result[0]==1:
  print("Negative")
elif result[0]==2:
  print("Neutral")
```

```
Negative
```

2. Support Vector Machine

In [20]:
```
input_tweet=["Ukraine army will defeat the russian army"]
predictions=vectorizer.transform(input_tweet)
result=svm_model.predict(predictions)
if result[0]==0:
  print("Postive")
elif result[0]==1:
  print("Negative")
elif result[0]==2:
  print("Neutral")
```

```
Negative
```

3. Random Forest

In [18]:
```
input_tweet=["Ukraine army will defeat the russian army"]
predictions=vectorizer.transform(input_tweet)
result=randomforest_model.predict(predictions)
if result[0]==0:
  print("Postive")
elif result[0]==1:
  print("Negative")
elif result[0]==2:
  print("Neutral")
```

```
Negative
```

4. XGBoost

In [19]:
```
input_tweet=["Ukraine army will defeat the russian army"]
predictions=vectorizer.transform(input_tweet)
result=xgb_model.predict(predictions)
if result[0]==0:
  print("Postive")
elif result[0]==1:
  print("Negative")
elif result[0]==2:
  print("Neutral")
```

```
Neutral
```

Bibliography

Dataset

https://www.kaggle.com/code/samupark/ukraine-war-twitter-sentiment-analysis

Material and Content for reference

https://www.geeksforgeeks.org/

https://data-flair.training/blogs/machine-learning-tutorial/

Tools

We have used Jupyter Notebook for this sentiment analysis.

https://jupyter.org/