# INSTITUTE OF ENGINEERING & MANAGEMENT

## Department of Computer Science & Engineering

## Project Title: Twitter Sentiment Analysis

| Name | Enrollment Number | Roll Number |
|---|---|---|
| Oindrila Sengupta | 12019002002113 | 78 |
| Aritra Ray | 12019002002114 | 79 |
| Tiyasha Nag | 12019002002135 | 99 |

**Guided By: Prof. Bavrabi Ghosh**

# Acknowledgement

We thank the Director of our esteemed Institution, Mr. Satyajit Chakrabarti for giving us this opportunity to work on this project. We extend our heartfelt thanks to Professor Bavrabi Ghosh and Professor Soma Das  who guided us throughout the project and always kept us motivated. We also thank our parents and friends who have been a constant support.

# Table of Contents

# Abstract:

In today's era, the impact of social networking media such as Facebook, Google Plus, YouTube, Blogs, and Twitter is increasing rapidly day by day. Millions of people are connected with each other on social networking sites and express their sentiments and opinions through tweets, and comments. This motivates the automatic mining and classification of views, emotions, opinions, and feelings of people on social networking websites. Sentiment analysis is the method of automatic mining of the sentiments or opinions of a text unit. In today's world, people express their feelings, opinions on twitter about certain things i.e. event, topic or person. The proposed work deals with mining sentiments or emotions of tweets using Support Vector Machine. Unigram and TF-IDF are used as feature extractors and the performance of the proposed approach is measured in terms of accuracy, precision, recall, and f-measure.

Keywords: Sentiment Analysis, SVM, Twitter, Ukraine-Russia War

# Introduction:

The Russo-Ukrainian War is an ongoing war between Russia (together with pro-Russian separatist forces) and Ukraine. It began in February 2014 following the Ukrainian Revolution of Dignity, and initially focused on the status of Crimea and the Donbas, internationally recognised as part of Ukraine. We had to  create a dataset of approx 5000 relevant tweets. Annotation can be done manually or automatically. We had to apply sentiment analysis to this dataset to find how many tweets are in support of Ukraine and how many are in support of Russia. Also,we had to  do an analysis to find how many tweets are in support of the war and how many tweets are against the war. We had to show accuracy of our model and specify which machine learning algorithm has been used and justification for using the same. We have to add the justification as a comment in the notebook where you developed the code. We collected our tweets based on the war of Ukraine and Russia and stored it as a database. This process is referred to as Web Scraping. We manually annotated the tweets in order to know the sentiment of the user writing it. We have used the SVM Model to predict the sentiment of the user writing the tweet.

# Steps of collecting the data and building the dataset:

Step 1: We have imported the package named snscrape.modules.twitter as sntwitter.
Step 2: We have taken a query which will filter the tweets having connection with the Ukraine and Russia war.
Step 3: We have stored the specified tweets in an array named tweet[] which has a limit of 3000.
Step 4: Under the for loop, we have appended the 3000 tweets in our tweet[] array.
Step 5: We have created a dataframe and stored our tweets in that.
Step 6: After having printed the dataframe, we converted it to a csv file named tweets.

**Output:**

Columns: index,Tweets,Support,War

```
0,So is #USA waging a proxy war against Russia in Ukraine? This can only be answered with
cheap humour,Russia,No
1,I am so confused. At this moment nobody talks about the Ukraine/ Russia war just the
media. Where are the pacifists of USA ? Do they support this war ? Many protests against
Afghanistan and Irak war and zero on this war . A war where Biden is interested canâ€™t not
be good.,Ukraine,No
2,"""Women are ""the biggest victims of Putin's war"" and ""the easiest target for a Russian
soldier's cruelty,""""
""Russia is not only waging war against Ukraine, Russia is also waging war against
women,""
@euobs #EverydaySexism
https://t.co/ucUSVA9Zrc",Ukraine,No
3," Are there US Airforce and troops, Command and Control in Ukraine? Russia supported
Armenia while Israel and Turkey supported Azerbaijan in the Nagorno-Karabakh war but this
doesn't mean those countries were at war against each other. You guys are mixing issues
here",Russia,No
4,"How would the international community respond to a Russian declaration of war against
Ukraine? Going forward, what would that mean for countries that supply Ukraine with
military aid? #Ukraine #Russia",Ukraine,No
```

# Algorithm or flowchart explaining steps to develop the project:

**Web Scraping and dataframe conversion:**

- Gather data with the help of **snscrape.modules.twitter**
- Manually annotate the data with two columns holding the following values: Support [‘Russia’, ‘Ukraine’] and War [‘Yes’, ‘No’]
- Upload the file on to cloud/Google Colab for further operations
- We will read the dataset into a DataFrame named: df=pd.read_csv(io.BytesIO(uploaded['filename.csv']))

**Pre-processing the data:**

- Download necessary packages like pandas, nltk etc.
- Now, proceed on to cleaning the tweets by removing the following: @mentions, hashtags, stopwords, extract stem words, retweets, hyperlinks etc.
- Perform Subjectivity and Polarity functions

**Perform EDA**

- Import nltk and download 'vader_lexicon'
- Using nltk.sentiment.vader , import SentimentIntensityAnalyzer
- Create a new columns named “Positive”, “Negative” containing the sentiment intensity of individual tweets from our dataset using SentimentIntensityAnalyzer

**Data Visualisation**

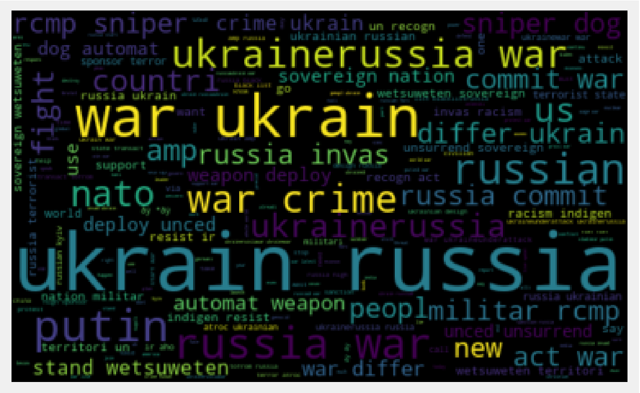
- We display all the words, positive and negative tweets in word cloud for data visualisation
- Then we count all the tweets in support of Russia/ Ukraine and tweets in support of War and No war. Sort the tweets accordingly
- Plot a scatter graph in accordance with Subjectivity and Polarity
- Plot and visualise the counts in bar graphs

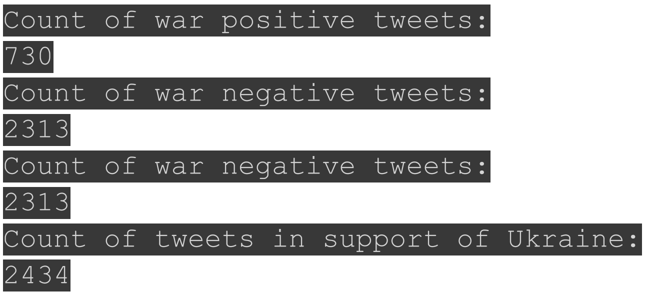
**Prediction using Machine Learning Algorithm**

- Feature extraction and vectorisation
- from sklearn.feature_extraction.text import CountVectorizer and from sklearn.feature_extraction.text import TfidfTransformer
- Here, we will take two cases, one with a target as 'Support' column and the other with 'War' with ‘Tweets’

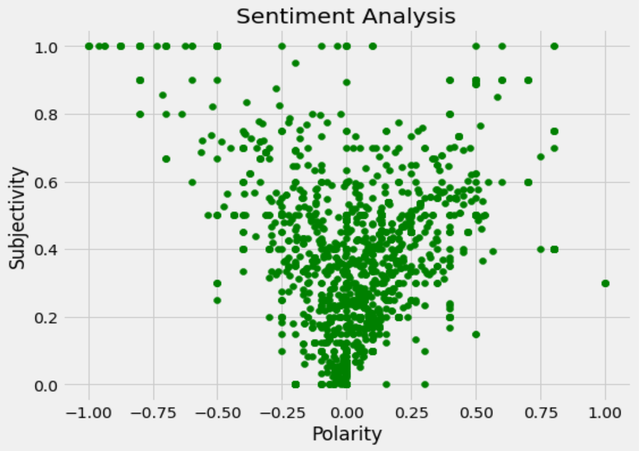
- Data Feeding to the model is the next step: in the following step, we feed the selected data to our model, SVM. SDG Classifier by default holds the following characteristics:
  The model it fits can be controlled with the loss parameter..The loss function defaults to ‘hinge’, which gives a linear SVM. SGDC is an approximation algorithm like taking single points and as the number of point increases it converts more to the optimal solution. Therefore, it is mostly used when the dataset is large. The regularizer is a penalty added to the loss function that shrinks model parameters towards the zero vector. We used SVM specifically over Multinomial Naive Bayes or other models due to better performance and accuracy achieved. It gives an excellent result for text categorization tasks such as sentiment analysis By using the correct kernel and setting an optimum set of parameters.
- We tally the score of the model via **Accuracy and F1 score**
- Checking accuracy with other models by using the following model list: [LGBMClassifier(objective='binary'), RandomForestClassifier(random_state=42,n_jobs=-1), GaussianNB(),KNeighborsClassifier(n_jobs=-1)]
- As we see, results with SVM provides the best accuracy based on the data provided

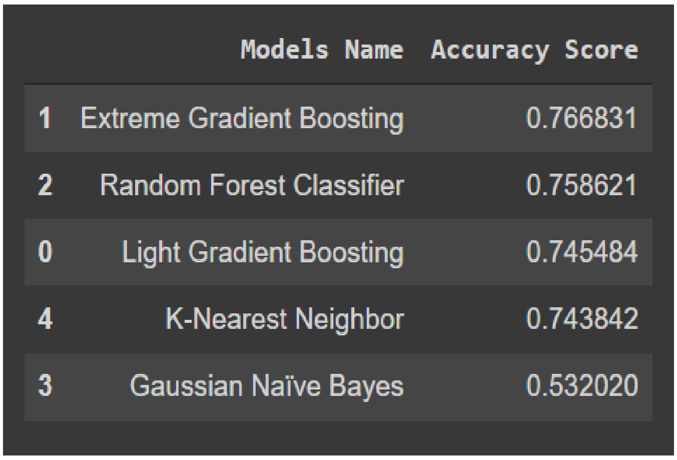
## Screenshot of outputs:



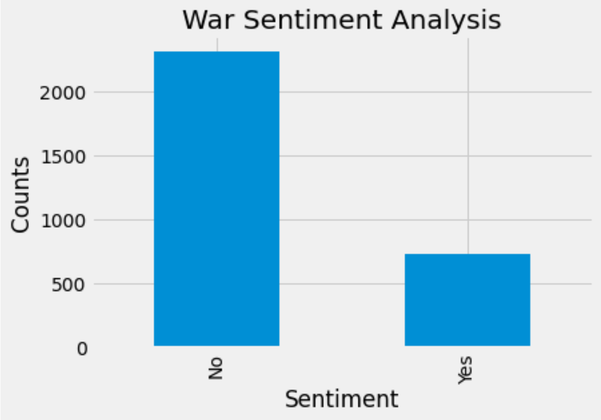


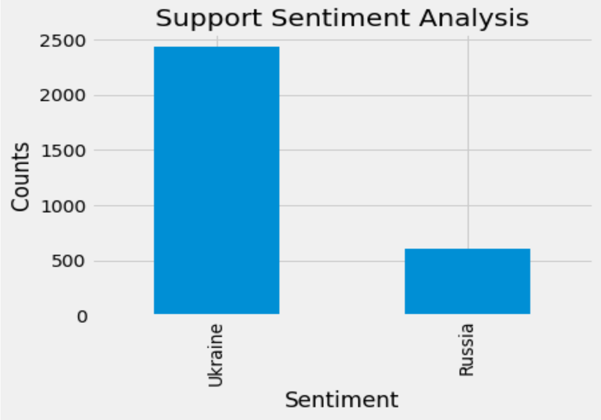
```
Count of war positive tweets:
730
Count of war negative tweets:
2313
Count of war negative tweets:
2313
Count of tweets in support of Ukraine:
2434
```

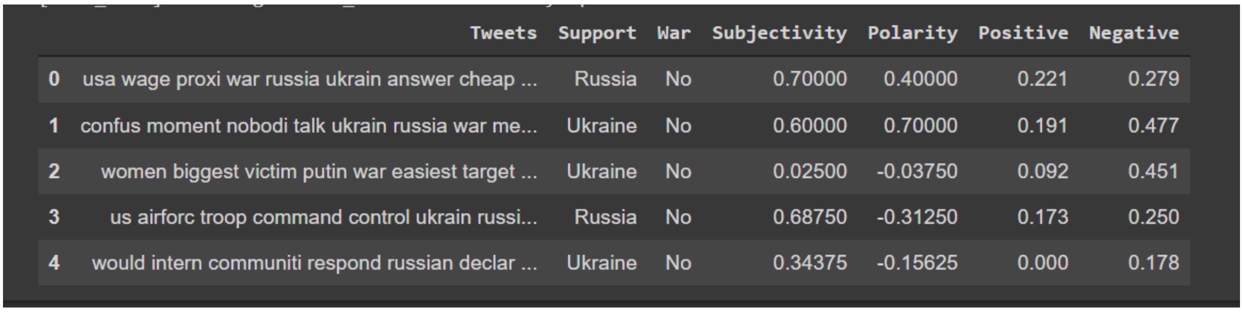




| | Models Name | Accuracy Score |
|---|---|---|
| 1 | Extreme Gradient Boosting | 0.766831 |
| 2 | Random Forest Classifier | 0.758621 |
| 0 | Light Gradient Boosting | 0.745484 |
| 4 | K-Nearest Neighbor | 0.743842 |
| 3 | Gaussian Naïve Bayes | 0.532020 |







| | Tweets | Support | War | Subjectivity | Polarity | Positive | Negative |
|---|---|---|---|---|---|---|---|
| 0 | usa wage proxi war russia ukrain answer cheap ... | Russia | No | 0.70000 | 0.40000 | 0.221 | 0.279 |
| 1 | confus moment nobodi talk ukrain russia war me... | Ukraine | No | 0.60000 | 0.70000 | 0.191 | 0.477 |
| 2 | women biggest victim putin war easiest target ... | Ukraine | No | 0.02500 | -0.03750 | 0.092 | 0.451 |
| 3 | us airforc troop command control ukrain russi... | Russia | No | 0.68750 | -0.31250 | 0.173 | 0.250 |
| 4 | would intern communiti respond russian declar ... | Ukraine | No | 0.34375 | -0.15625 | 0.000 | 0.178 |

# Result - Accuracy measurement:

### Tweets - War

```
from sklearn import metrics
print("\nF1 Score: {:.2f}".format(f1_score(y, predicted,
average='micro') * 100))
print("Accuracy: {:.2f}%".format(accuracy_score(y, predicted) * 100))
F1 Score: 76.01
Accuracy: 76.01%
```

### Tweets - Support

```
print("\nF1 Score: {:.2f}".format(f1_score(y, predicted,
average='micro') * 100))
print("Accuracy: {:.2f}%".format(accuracy_score(y, predicted) * 100))
F1 Score: 79.99
Accuracy: 79.99%
```

# Conclusion:

Sentiment analysis is the process of categorising a person‟s opinion and emotions expressed in the form of text. In today's scenario, social networking sites are full of people's opinions about product reviews, movie reviews, politics or a particular topic, etc. In this paper, Kernel SVM are applied for classification of the sentiments of about the ongoing Ukraine-Russia War. The tweets of people are collected from twitter snscrape. In addition to it, feature selection is also applied here. The performance of both SVM is analysed on various measures, i.e, accuracy, precision, recall, and f-measure. It is analysed that linear SVM performs better than other verified models in this case. This can be tallied by observing the accuracy/ f1 scores in the output section. We hope that this war soon comes to an end for the good of humanity.

# References

- scikit-learn.org: https://scikit-learn.org/stable/tutorial/text_analytics/working_with_text_data.html
- thecleverprogrammer.com: https://thecleverprogrammer.com/2022/03/15/ukraine-russia-war-twitter-sentiment-analysis-using-python/