A PROJECT REPORT

on

"Ukraine vs. Russia: Twitter Sentiment Analysis"

Submitted to

KIIT Deemed to be University

In Partial Fulfillment of the Requirement for the Award of

BACHELOR'S DEGREE IN
COMPUTER SCIENCE AND ENGINEERING

BY

ANUSKA MISHRA	21051717
BHUVAN PAGILLA	21051748
UJJWAL KASHYAP	21051780
AADARSH VERMA	21051701

UNDER THE GUIDANCE OF
Prof. Deependra Singh

SCHOOL OF COMPUTER ENGINEERING

KALINGA INSTITUTE OF INDUSTRIAL TECHNOLOGY

BHUBANESWAR, ODISHA - 751024

April 2024

A PROJECT REPORT

on

"Ukraine vs. Russia: Twitter Sentiment Analysis"

Submitted to

KIIT Deemed to be University

In Partial Fulfilment of the Requirement for the Award of

BACHELOR'S DEGREE IN COMPUTER SCIENCE AND ENGINEERING

BY

ANUSKA MISHRA	21051717
BHUVAN PAGILLA	21051748
UJJWAL KASHYAP	21051780
AADARSH VERMA	21051701

UNDER THE GUIDANCE OF

Prof. Deependra Singh

SCHOOL OF COMPUTER ENGINEERING

KALINGA INSTITUTE OF INDUSTRIAL TECHNOLOGY

BHUBANESWAR, ODISHA -751024

April 2024

KIIT Deemed to be University

School of Computer Engineering
Bhubaneswar, ODISHA 751024

CERTIFICATE

This is certify that the project entitled

"Ukraine vs. Russia: Twitter Sentiment Analysis"

submitted by

ANUSKA MISHRA	21051717
BHUVAN PAGILLA	21051748
UJJWAL KASHYAP	21051780
AADARSH VERMA	21051701

is a record of bonafide work carried out by them, in the partial fulfilment of the requirement for the award of Degree of Bachelor of Engineering (Computer Science & Engineering OR Information Technology) at KIIT Deemed to be university, Bhubaneswar. This work is done during year 2022-2023, under our guidance.

Date: 18/03/2024

(Prof. Deependra Singh)
Project Guide

Acknowledgements

We are profoundly grateful to **Prof. Deependra Singh** of **KIIT School of Computer Engineering** for his expert guidance and continuous encouragement throughout to see that this project reaches its target since its commencement to its completion.

ANUSKA MISHRA
BHUVAN PAGILLA
UJJWAL KASHYAP
AADARSH VERMA

ABSTRACT

This project dives into the cosmos of social media to analyze the sentiments of people regarding the Russia and Ukraine war. It uses the power of python to help extract the data and analyze their sentiments from the tweets on X (formerly known as Twitter).

Keywords in tweets which were public, were collected, related to the ongoing war using pre-existing datasets. Irrelevant and useless information was cleared out of the datasets such as hashtag, special characters/symbols, URLs, etc. A sentiment analysis library was used to classify the tweets into different categories like positive, negative and neutral. The overall sentiment analysis is done and calculated and later on plotted for visualization purposes.

Keywords: Russia, Ukraine, war, sanction, crisis, bomb, weapons, Kyiv, Kharkiv, Donetsk, Putin, peace, ban, oil, genocide, kill, hope, strength, Zelensky, crime, die, pray, children, border, halt, not, no, humanitarian, etc.

Contents

List of Figures

Chapter 1

Introduction

This project reflects on the thinking as well as the feelings of millions of people regarding a genocidal war i.e. Russia vs. Ukraine which has been dated back to decades.

It helps in understanding public opinion. Social media, whether it be Facebook, Instagram or Twitter, it gives us a glimpse into the mindset of the person who typed out the captions/tweets. It helps analyze the emotional side of the conflict.

The project can be used to track data such as how public opinion varies over time. It will show the support for Ukraine either increasing or else decreasing and the same for Russia in the statistical analysis. It will also show the events of outrage and whether things have cooled down or not. Analyzing provides valuable insights.

The narratives will be identified. This project will let us know if we are being fooled by the Russian government sending out too many bot-messages to twist the real incidents and give the public as well as the media a fake narrative by twisting the truth.

By capturing the datasets and turning them into a sentiment analysis structure, we can keep records of the history of X (twitter) and public opinions which will be extremely valuable later on in the future.

To put this simply, this project is beyond just the "who is siding with which country?" . It will give us insights into the emotional turmoil of conflicting opinions and nerve-wrecking truth, which can affect the decision making process of the people understanding this concept and trying to wrap their minds around the realness of the situation. It will also affect the historical understanding and the communication strategies.

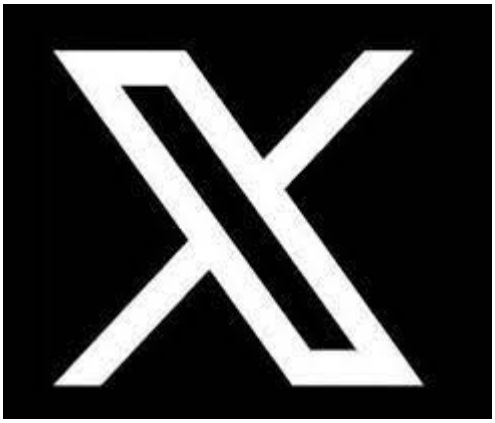


Fig 1.1: Data collection from X

TOP 10 HASHTAGS AND MENTIONS

hashtags	mentions
Ukraine	ZelenskyyUa
Russia	NATO
Putin	POTUS
UkraineRussiaWar	Ukraine
Russian	UN
Kyiv	vonderleyen
Ukrainian	elonmusk
Kharkiv	KyivIndependent
ukraine	EmmanuelMacron

Fig 1.2: Data pre-processing

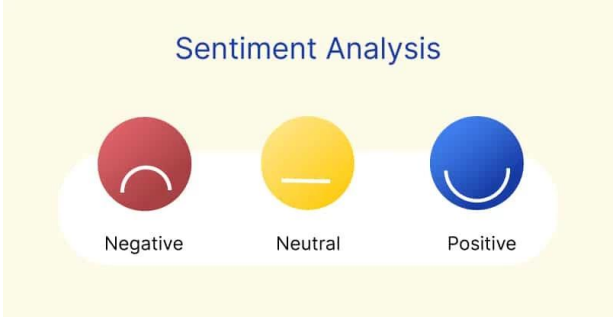


Fig 1.3: Sentiment analysis

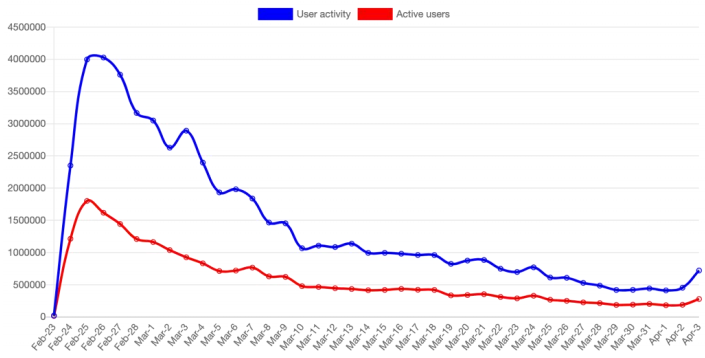


Fig 1.4: Data visualization

Chapter 2

Basic Concepts

The tools used in this project were
i) Python as a programming language
ii) Pre-existing datasets from kaggle for sentiment analysis
iii) Matplotlib library for data visualization
iv) TextBlob as a sentiment analysis library

The techniques used were:
i) Data collection
ii) Data pre-processing
iii) Sentiment analysis
iv) Data analysis and visualization

2.1 Data collection

Tweets are retrived from pre-existing datasets using kaggle site.

2.2 Data pre-processing

Tweets are cleaned by removing irrelevant details and unnecessary elements such as hashtags, URLs, special characters, symbols, etc.

2.3 Sentiment analysis

The chosen library, in this case, TextBlob, analyzes the cleaned text of every tweet from the dataset and categorizes it as positive, negative or neutral.

2.4 Data analysis and visualization

The sentiment distribution is analyzed and calculated and then displayed/visualized in forms of pie charts and graphs.

Chapter 3

Problem Statement / Requirement Specifications

The problem statement is the ongoing conflict of Ukraine and Russia which has been continuing for decades and its discussions on a social media platform like X (Twitter). This project aims to do a sentiment analysis of the tweets expressed by the public regarding the war to understand the situation better along with the stance of the people.

3.1 Project Planning

The steps taken during project planning are:
i) Day 1-3:
[Research and planning]
The topic was chosen first. Then the relevant research questions were looked for. The keywords were defined. Data collection method was chosen i.e datasets and tweets from X (Twitter). Sentiment analysis library was chosen along with various methods of visualization.

ii) Day 4-9:
[Development]
The necessary libraries were looked for and downloaded after python environment was set up. Code for data collection was written. The data pre-processing was implemented. Sentiment analysis library was integrated with this data and code was developed and written for visualization.

iii) Day 9-12:
[Testing and refinement]
Testing of data collection and pre-processing is done. The sentiment analysis accuracy is checked with a sample dataset. The visualizations are refined after repeated testing and checking. The functionalities are documented.

iv) Day 13-15:
[Analysis and reporting]
The project is run with the full dataset procured. The sentiment analysis is done and the trends are documented through visualization of pre-processed data. A report is written based on all of the steps taken while finishing the project.

3.2 Project Analysis

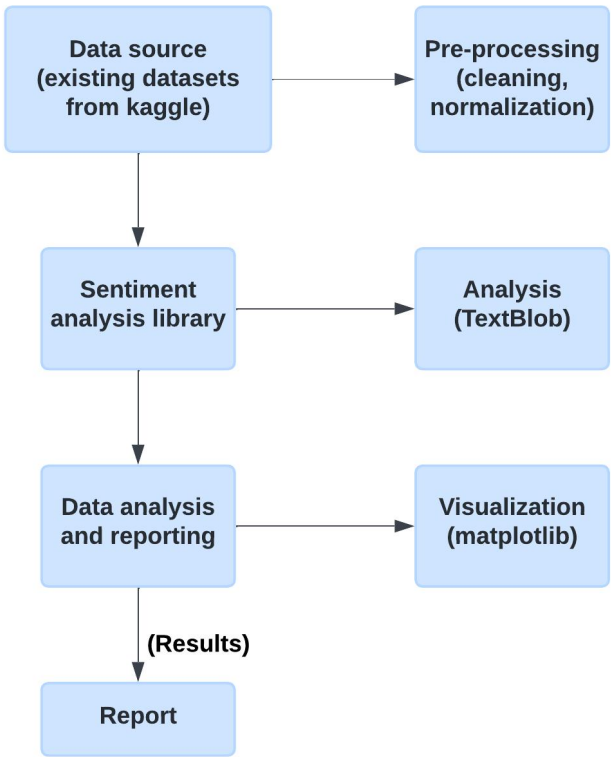
After the preprocessing was finished, the datasets were run through the sentiment analysis part, the data was visualized and the portion of tweets that were positive were 31.54%, for negative it was 54.56%, and for neutral it was 13.90%.

3.3 System Design

3.3.1 Design Constraints

i)Data acquisition is a hassle since relying on pre-existing datasets limits the scope of the analysis.
ii)Sentiment analysis tools are not perfect and already struggle with sarcasm, irony and rhetoric questions. The library had to evaluated properly before being chosen for the output of datasets.
iii)Most analyzers will specify tweets as positive, negative and neutral. But this doesnot do justice to the plethora of emotions existing in the public's minds while they tweeted those tweets.
iv)Removing unnecessary hashtags, URLs, etc. is a big cleaning job which takes additional code for the library.
v)Data visualization is difficult with large datasets. The computational power is also not too good for them. Code has to be optimized for efficiency before and while testing.

3.3.2 Block Diagram



Chapter 4

Implementation

4.1 Methodology OR Proposal

i) The data is retrived from Kaggle pre-existing datasets.
ii) It is pre-processed and sent through the sentiment analysis, TextBlob, for segregation of opinions i.e. positive, negative or neutral.
iii) The processed data after the analyzer is visualized through matplotlib library of python through pie chart and graph.

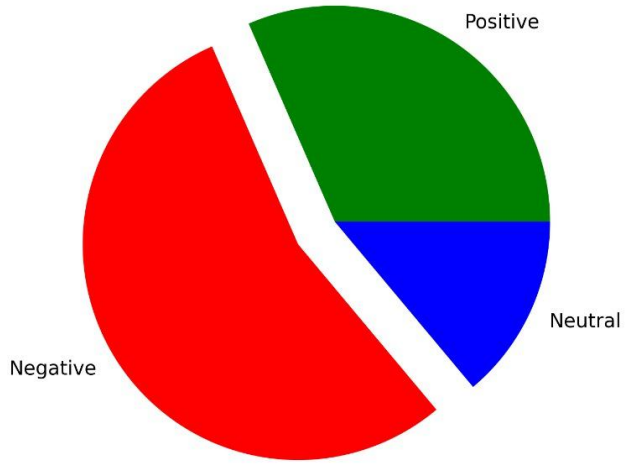
4.2 Testing OR Verification Plan

Testing of dataset:

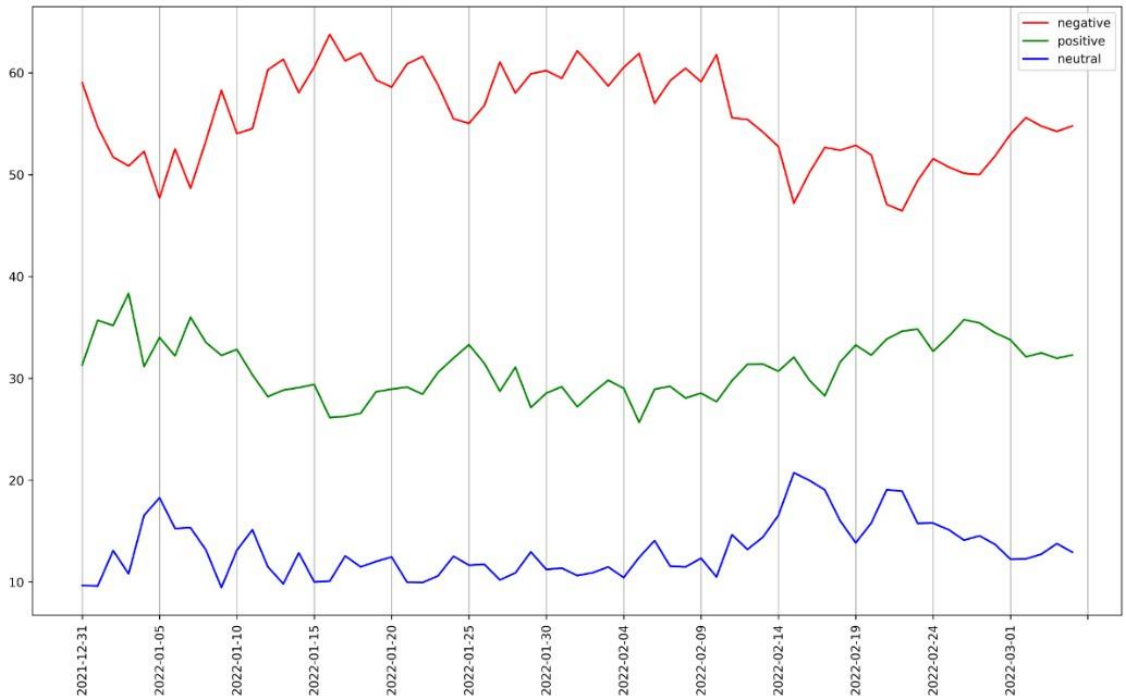
```
In [42]:
1 total_pos = len(tweets.loc[tweets['sentiment'] == "positive"])
2 total_neg = len(tweets.loc[tweets['sentiment'] == "negative"])
3 total_neu = len(tweets.loc[tweets['sentiment'] == "neutral"])
4 total_tweets = len(tweets)
5 print("Total Positive Tweets % : {:.2f}".format((total_pos/total_tweets)*100))
6 print("Total Negative Tweets % : {:.2f}".format((total_neg/total_tweets)*100))
7 print("Total Neutral Tweets % : {:.2f}".format((total_neu/total_tweets)*100))

Total Positive Tweets % : 31.54
Total Negative Tweets % : 54.56
Total Neutral Tweets % : 13.90
```

4.3 Result Analysis OR Screenshots



```
Out[58]: <matplotlib.legend.Legend at 0x15c5fb48750>
```



4.4 Quality Assurance

The quality of the work is checked by our professor, Deependra Singh , under whose guidance we have made this project.

Chapter 5

Standards Adopted

5.1 Design Standards

i) Code readability and maintainability were checked. Proper usage of comments was done and the variables were named differently and not hastily so as to give a sense of clarity when another person worked on it. Code indentation was done so as to keep it neat and tidy.
ii) Data validation was done to ensure the collected data was free of errors.
iii) Code was optimized regularly for the large dataset while being tested to improve its performance.

5.2 Coding Standards

Coding standards are collections of coding rules, guidelines, and best practices. Few of the coding standards used in the project are:
1. Proper naming of variables and functions.
2. Naming convention was used.
3. Appropriate amount of whitespace was given.
4. Proper indentation was done.
5. Comments were written to explain the purpose of every function, loop, variable, etc.

5.3 Testing Standards

i) Unit testing was done to check if every function performed properly as it should.
ii) A test of how various modules behaved with each other was done.
iii) Manual testing was done by the group members and the tweets were classified into categories and later compared to the categorization done by the sentiment analyzer to check if it was working properly.
iv) Code reviews were done regularly for code optimization such as identifying errors or potential errors, inefficiencies, areas of improvement, etc.

Chapter 6

Conclusion and Future Scope

6.1 Conclusion

The project was successful in exploration of the horizons of python and helped in the sentiment analysis of tweets involving the Russia vs. Ukraine war. It analyzed the tweets and segregated them into positive, negative and neutral perspectives. This helps in revealing the public's stance on the whole situation as well as providing insights. The data visualizations helped in providing a clear and concise way for communicating these insights to a wider public audience. This projects demonstrates the power of social media combined with the potential of a programming language in showing us critical decision making, communication, strategies and historical understanding of events.

6.2 Future Scope

i) Using more refined sentiment analysis libraries, we can go beyong the segregation of just positive, negative and neutral and go deeper into the emotions felt and expressed in such tweets.
ii) X (Twitter) provides the geographical location of tweets through which we could analyze if sentiment varies more as a group from region to region or country to country.
iii) A comparative analysis could be done to compare the sentiments shown on X (Twitter) with other platforms such as Instagram comments and hashtags, various newspapers, etc.
iv) Topic modeling could be used to identify the topics and themes around the conflict of the war.

By building more on this project, we can create a better version/model and gain a deeper insight while analyzing the situation better regarding the complex landscape involving a lot of emotions.

References

[1] https://www.turing.com/kb/natural-language-processing-function-in-ai

[2] https://study.com/academy/lesson/project-verification-methods-techniques.html

[3] https://www.bairesdev.com/blog/best-python-sentiment-analysis-libraries/

[4] https://www.kaggle.com/datasets/bwandowando/ukraine-russian-crisis-twitter-dataset-1-2-m-rows

[5] https://www.kaggle.com/code/avinandandutta/twitter-sentiment-analysis-russia-ukraine-conflict

[6] Online Flowchart Maker | Lucidchart

[7] https://github.com/alexdrk14/RussoUkrainianWar_Dataset

RUSSIA VS. UKRAINE: TWITTER SENTIMENT ANALYSIS

ANUSHKA MISHRA
21051717

Abstract: This project aims to understand public sentiment regarding the Russia and Ukraine war by analyzing the tweets with the help of a dataset through a sentiment analyzer. The objective was to use Python sentiment analysis tools in classifying the tweets as positive, negative or neutral. The classification is visualized to provide insights.

Individual contribution and findings: The planning of the project including the research for various sentiment analysis libraries was done by me. I suggested to go with Python as it gives various tools to process the datasets. I contributed in writing some part of the code and helped in overseeing the trial and error phase along with code optimization.

Individual contribution to project report preparation: I wrote the Introduction in Chapter 1, the project planning steps in Chapter 3, and made the Block diagram in Chapter 3.

Individual contribution for project presentation and demonstration: I helped by attaching the screenshots of the project outputs and editing out the grammatical errors.

Full Signature of Supervisor:

…….

Full signature of the student:

……..

RUSSIA VS. UKRAINE: TWITTER SENTIMENT ANALYSIS

BHUVAN PAGILLA
21051748

Abstract: This project aims to understand public sentiment regarding the Russia and Ukraine war by analyzing the tweets with the help of a dataset through a sentiment analyzer. The objective was to use Python sentiment analysis tools in classifying the tweets as positive, negative or neutral. The classification is visualized to provide insights.

Individual contribution and findings: I helped in writing part of the code for sentiment analysis and collected datasets for the project. I helped in code testing and data visualization.

Individual contribution to project report preparation: I wrote Chapter 2 and Chapter 4.

Individual contribution for project presentation and demonstration: I helped in cleaning the code for readability and maintainability.

Full Signature of Supervisor:

…….

Full signature of the student:

……..

RUSSIA VS. UKRAINE: TWITTER SENTIMENT ANALYSIS

UJJWAL KASHYAP
21051780

Abstract: This project aims to understand public sentiment regarding the Russia and Ukraine war by analyzing the tweets with the help of a dataset through a sentiment analyzer. The objective was to use Python sentiment analysis tools in classifying the tweets as positive, negative or neutral. The classification is visualized to provide insights.

Individual contribution and findings: I wrote most part of the code and was the primary in testing all the functions individually and checking the sentiment analysis manually and using sample data multiple times to get results and make it better.

Individual contribution to project report preparation: I wrote the Chapter 5 in the report and helped collect references for it.

Individual contribution for project presentation and demonstration: I helped in data visualization/outputs.

Full Signature of Supervisor:

…….

Full signature of the student:

……..

RUSSIA VS. UKRAINE: TWITTER SENTIMENT ANALYSIS

AADARSH VERMA
21051701

Abstract: This project aims to understand public sentiment regarding the Russia-Ukraine war by analyzing the tweets with the help of a dataset through a sentiment analyzer. The objective was to use Python sentiment analysis tools in classifying the tweets as positive, negative or neutral. The classification is visualized to provide insights.

Individual contribution and findings: I helped in sourcing out the ways in which the steps could progress. I helped in overseeing the code editing process as well as the optimization process while testing.

Individual contribution to project report preparation: I helped in writing the references and the Chapter 6.

Individual contribution for project presentation and demonstration: I helped in editing the code and making the report cleaner.

Full Signature of Supervisor: Full signature of the student:

……. ……..

TURNITIN PLAGIARISM REPORT

(This report is mandatory for all the projects and plagiarism must be below 25%)