# Defense_Shahin_Ashraful_ (6).pdf

*By* Sahin


| | | | |
|---|---|---|---|
| WORD COUNT | 9832 | TIME SUBMITTED | 26-DEC-2022 03:17PM |
| | | PAPER ID | 94380355 |

A Thesis for the Degree of Bachelor of Science


# Machine Learning Based Sentiment Analysis on Russia-Ukraine War Tweets

**By**

Ashraful Haque Tani
ID: B-170305007
Md.Shahin Alam
ID: B-170305044
Session: 2017-2018

**Supervised By**

Dr. Md. Zulfikar Mahmud
Associate Professor

**Department of Computer Science and Engineering**
**Jagannath University**
**Dhaka, Bangladesh**

**January 02, 2023**

# Recommendation of the Board of Examiners

The Thesis Report ***"Machine Learning Based Sentiment Analysis on Russia-Ukraine War Tweets"*** submitted by Ashraful Haque Tani ID: B170305007 and Md. Shahin Alam ID:B170305044 to the Department of Computer Science Engineering, Jagannath University Dhaka, has been accepted as satisfactory for the partial ful- fillment of the requirements for the degree of B.Sc. in Computer Science Engi- neering and approved as to its style and contents

## Examiners

1 ..... Supervisor

2 ..... Examiner

3 ..... Examiner

4 ..... Chairman

# Declaration of Authorship

We, hereby, declare that the work presented in this thesis is the outcome of the investigation performed by us under the supervision of Dr. Md. Zulfikar Mahmud, Associate Professor, Department of Computer Science & Engineering, Jagannath University Dhaka. We also declare that no part of this thesis and thereof has been or is being submitted elsewhere for the award of any degree or Diploma.

| Countersigned | Signature | Signature |
|---|---|---|
| ..... | ..... | ..... |
| ( Dr. Md. Zulfikar Mahmud ) | (Ashraful Haque Tani) | (Md.Shahin Alam ) |
| **Supervisor** | **Candidate** | **Candidate** |

# Dedication

*Dedicated to*

*Our beloved parents*

# Abstract

On February 24, 2022, the conflict between Ukraine and Russia broke out, shocking the entire world and dominating social media. People from all around the world have expressed their opinions regarding the conflict between Russia and Ukraine ever since the war began. One of the most important social media platforms, Twitter, offers useful perception into what people are thinking. In this work, we take the twitter data related to the Russia Ukraine war for sentiment analysis of the people all around the world. we used the Natural Language Toolkit (NTLK) called VADER (Valence Aware Dictionary and Sentiment Reasoner) which is pre-trained sentiment analyzer to score the tweets in the dataset.

This paper will able to understand the psychology and behavior of societies in Russia and Ukraine war by analysis of sentiments like positive, negative or neutral.we used the Natural Language Toolkit (NTLK) called VADER (Valence Aware Dictionary and Sentiment Reasoner) which is pre-trained sentiment analyzer to score the tweets in the dataset.Respectively using different vectorization (BoW Tfidf) and Machine Learning algorithm (Logistic Regression,Naive Bayes,Support Vector Machine, Decision Tree) we can can predict what people think about this issue.



**Key words:** sentiment analysis, Russia Ukraine war, machine learning, opinion mining.

# Acknowledgment

It is our great pleasure to express our heartiest thankfulness to those who gave me the valuable time and supported me in making this dissertation possible. Thanks all of you to make my dream successful. we owe our deepest sense of gratitude to our honorable supervisor, Assoc/Prof **Dr. Zulfikar Mahmud** for his excellent supervision, meaningful suggestions, persistent encouragements and other fruitful help during each stage of our B.Sc. study. His thoughtful comments and guidance helped us to complete our thesis work and present them in productive ways. We have really been lucky in working with a person like him. Needless to say, it would not have been possible to complete this thesis without his guidance and active support. We would like to acknowledge Dept. of Computer Science and Engineering, Jagannath University for giving us opportunity to complete B.Sc. Last but not least, We are grateful to all of our teachers for their unconditional loves, supports, patience, and confidence on us are the biggest rewards as well as the driving forces of our life.

Ashraful Haque Tani

Md. Shahin Alam

# Table of Contents

# List of Figures

# List of Tables

# Chapter 1

# Introduction

## 1.1 Introduction

Social media have evolved into engaging debate forums for many groups, allowing for the expression of individual opinions. Because they allow users to freely express their thoughts and emotions and spread an infinite amount of information in contemporary society, these platforms have become essential to people's daily lives. Today's population spends a significant amount of time on social media websites as a result of the global adoption of the Internet, making social media an indispensable information source[3]. In a similar vein, social media users are increasingly likely to have either beneficial or bad effects on society due to the platforms' growing popularity among people worldwide. Online sources have a significant impact on how the public feels about current affairs like politics, the economy, or social life. Among them, Twitter is the most commonly used social media platform where individuals everyday express their thoughts and feelings on a certain subject [4]. Millions of people use Twitter daily to share billions of different pieces of material.

As a result, Twitter sentiment analysis may be useful in a variety of contexts, such as assessing people's reactions to certain events, products, movies, and songs, among

others. In certain situations, natural language can be beneficial. Sentiment Analysis has already been studied extensively in the context of other similar issues. We provide a novel approach of sentiment analysis using twitter data on the Russia Ukraine war in our suggested system, with the objective of bringing some predictions on polarity kinds for political reforms. In this work, we will exhibit an end-level decision of how many individuals are commenting in favor of war and how many are commenting in opposition to war , followed by categorization of positive, negative and neutral responses. This research will demonstrate how non-polarized tweets may be minimized by deleting stop words, symbols, and repetitive tweets from the dataset prior to analysis.

## 1.2 History

Starting with the annexation of Crimea by Russia in 2014, the Russia-Ukraine crisis turned into a war as a result of Russian attacks on Ukrainian territory on 24 February 2022. The Russian invasion of Ukraine witnessed many violent battles and led to different reactions around the world, including on Twitter. During wartime, many press members pay attention to reporting daily events justly and displaying an impartial attitude towards the war [5]. Such conflicts will result in long-term economic, political, and psychological problems in a given society. People use social media platforms to convey their thoughts and feelings about this political situation all over the world. However, as online users, we are often limited to our social environment when exploring different individuals' opinions on the conflict. Not surprisingly, it is almost impossible for an individual to reach and read all tweets about a certain topic. In this respect, the analysis of tweets about the Ukraine-Russia war on Twitter using NLP techniques will help us gain insight into an impartial view of global tendencies, thus giving a unique data source for press member reports and articles. Therefore, a literature review of various discourses on a certain Twitter discussion topic may facilitate categorizing opinions using machine learning based

approaches.

## 1.3 Motivation

In recent years, scholars have conducted extensive Sentiment Analysis and Opinion Mining studies on a variety of topics and events. Sentiment Analysis can have a greater influence on community service as social media becomes a greater source of data. The Russia-Ukraine war is currently one of the most pressing issues on the international stage. As Russia is the 3rd Oil producer and 2nd natural gas provider in the world, The war has affected global oil and gas prices, as well as wheat, corn, sunflower, and various precious metals. The global impact of the Russia-Ukraine war is also being felt in Bangladesh. Suppose the dispute between Ukraine and Russia continues for a long time and spreads across Europe. In that case, the country's garment industry could be threatened as 64 percent of the country's garment exports, and 58 percent of the total exports are destined for the European market. As a result, we've determined that utilizing NLP and Machine Learning Algorithms, we can categories people's thoughts and opinions on this issue. After that, we may arrive at a satisfactory conclusion that will aid decision-making in the political arena.

## 1.4 Problem Statement

People are experimenting with diverse data in the modern world. There, new facts and ideas are discovered. However, examining scientific publications is difficult because to a number of factors that influence the evaluation of sentiment reviews. To distinguish the sentiment class, they employed several analysis techniques. The most crucial factor in making a decision is the sense and segmentation of the statement. There is a lot of study on sentiment analysis, but there isn't much on the refugee crisis. Existing studies on the refugee crisis, on the other hand, are primarily concerned with sentiment comparisons

between persons from various countries. The earlier work's flaw is that it failed to provide any useful results. As a result, we attempted to provide that study in a more efficient and exact manner by utilizing MLA (Machine Learning Algorithm) to forecast correctness.

## 1.5 Objectives

The objective of my paper is to understand people's emotion and opinion about Russia-Ukraine war and give prediction of polarity (polarity can be positive, negative or neutral, as determined by tweets after extracting various elements from tweets and evaluating them using natural language processing and an existing machine learning algorithm).

- To analyze people's emotion and opinion about Russia-Ukraine conflict.
- To extract the positive, negative or neutral sentiment about Russia-Ukraine Crisis.
- To train the machine by extracting features from our data so that machine can predict what people think about the issue.

## 1.6 Contributions

The following section examines my contributions to the development of this task, which include the following:

- In this work, we used the Natural Language Toolkit (NTLK) , specifically the pre-trained sentiment analyzer called VADER (Valence Aware Dictionary and Sentiment Reasoner) to score the polarity in the tweets of the dataset(polarity may be positive, negative or neutral).
- We used Various vectorization (BoW, Tfidf, W) and classification techniques were compared in terms of performance (Naïve Bayes, Logistic Regression,SVM Decision Tree).

## 1.7 Summary

We have outlined our project objectives, contributions that we made, a brief comment on the issue area, and the motivation behind our effort in this chapter. We've named it "motivation" and provided some historical context for Russia-Ukraine war.

## 1.8 Structure of The Thesis

There are six chapters in this research paper.The book is laid out as follows:

**Chapter 1: Introduction**

In this chapter, we discussed study's background, issue description, motivation, objective, contributions, and how the entire project is organized are all covered.

**Chapter 2: Background Study**

In this chapter, we discussed the background study of this research-related word.

**Chapter 3: Literature Review**

This chapter provides an overview of several existing works that are relevant to our thesis.

**Chapter 4: System Methodology and Implementation**

In the fourth chapter, discusses the system's methodology and execution. The implemented models' methodology is described, along with a description of the dataset.

**Chapter 5: Results and Discussions**

In this chapter we have discussed about the miner selection time .....We have shown some of the differences with existing work and their limits

**Chapter 6: Conclusions and Recommendations**

In chapter 6 we concluded our proposed model and its implementation.This section will be covered in more detail in the future.

The Bibliography includes all the references used in this work.

# Chapter 2

# Background Study

## 2.1 Introduction

The act of evaluating, experimenting, and testing people's opinions after gathering data in text form is known as sentiment analysis and text mining. In the following section we will discuss about it briefly.

## 2.2 Sentiment Analysis

People's emotions, views, thoughts, and ideas are referred to as sentiments. There are two kinds of sentiment: facts and opinion-based information. Fact is simply a statement on a thing or occurrence, whereas opinion is a person's subjective perspective, thinking, intention, or idea about an event or issue. Sentiment analysis is a way of assessing, experimenting, and testing people's opinions or facts using text data that has been gathered and preprocessed. We shall briefly address it in the next part. It's also known as information extraction, emotion collecting, decisions mining, and idea generation on a subject that might be subjective or objective [6].

The term "subjective" refers to a person's own view, whereas "objective" refers to the purpose of providing only the facts of an occurrence. The subjective sentences have three different types:

- Private situations as references. For example: "Rakib was taking with fear."
- Expressing private situations as references to speech. For example: "The editors of the New Age paper attacked the police officer."
- Expressive subjective entities. For example: "Rabbi is a brilliant."

## 2.3 Importance of Sentimental Analysis

Facebook, Twitter, Whatsapp, Google+, Skype, Sina Weibo, Instagram, and other social media platforms have billions of users. Where users express their reactions, opinions, thoughts, ideas, and points of view on various events and issues. As a result, online daily feelings have become the most essential decision-making resource. According to a latest survey done by World Research, online customer reviews account for the same percentage of sales as personal referrals.

According to a study (Spiegel Research Center, 2017), approximately 95% of customers read online reviews before making a purchase. This study also found that posting reviews may raise sales rates by 270 percent. According to a 2016 study ((Fan and Fuel), 94% of customers read online reviews, and 97% of buyers believe reviews affect their purchasing decisions [7].

## 2.4 Reason of Hardness of Sentiment Analysis

One of the most difficult tasks for computers to solve is sentiment analysis. Machines find it difficult, if not impossible, to recognize particular items and traits, but humans find it quite simple. The following are some difficult circumstances for computers to solve:

### 2.4.1 Dealing with Sarcasm

It is hard to comprehend a sentence's opposing meaning. This may sometimes be detected using a particular technique, although it is unreliable to rely on this approach.

### 2.4.2 Dealing with Strength of an Opinion

It might be difficult for a computer to evaluate the strength of a given opinion or review. Opinions differ in terms of their intensity. Some of them are very strong: "This war should be stopped" and some of them are weak: "We think this war can be stopped" [8]. Although a dictionary of weak or strong opinion words may be created specific application, computers are still uncomfortable because when the strength of an opinion is combined with the stance of that view, the document in many circumstances entirely changes the polarity.

## 2.5 Sentiment Analysis (SA) Natural Language processing (NLP)

### 2.5.1 Overview

In the next part, we'll go over several key definitions based on sentiment terminology that we've learned.

#### 2.5.1.1 Natural language processing (NLP)

NLP is a branch of computer science and artificial intelligence concerned with human-computer communication. NLP is linked to the topic of Human Computer Interaction (HCI). Understanding and interacting with natural language is a difficult task for NLP, as it requires computers to derive meaning from provided human or natural language input [9]. A wide range of strategies and methods NLP established by programming are

used for automated production, analysis of inner meaning of phrases, and interaction with humans. NLP inherits so many approaches from AI's knowledge base that it affects fresh disciplines. To comprehend NLP models and procedures, we must first go through some very fundamental definitions and terminology.

#### 2.5.1.2 Token

The incoming text must first be split into linguistic components such as words, punctuation, and numbers or alphanumeric before any serious processing can begin. Tokens are the units that are recognized.

#### 2.5.1.3 Sentence

An ordered series of tokens is referred to as sentence.

#### 2.5.1.4 Tokenization

The practice of collapsing a text into tokens is known as tokenization. Because of the presence of whitespace, tokenization is significantly easier in English, which is also known as a segmented language.

#### 2.5.1.5 Corpus

A corpus is sometimes referred to as a vast number of sentences, papers, blog data, website data, or simply a body of text.

#### 2.5.1.6 Part-of-speech (POS) Tag

A POS tag is just a representation of symbols that allows a word to be classified into one or more categories, such as NN (Noun), VB (Verb), AJ (Adjective), and AT (Adjective) (Article). The Brown Corpus tag set is one of the traditional and widely used tag sets.

#### 2.5.1.7 Parse Tree

A parse tree is a tree that divides a sentence into tree structures. It generates the syntactic framework of a given sentence after giving fundamental terminology and formal grammar.

### 2.5.2 Some Mutual NLP tasks:

#### 2.5.2.1 Computational Morphology

Morphemes, also known as stems, are fundamental building blocks that are created from a vast number of natural language words. The term Computational Morphology is used to identify and study the internal structure of words using a computer system.

#### 2.5.2.2 Parsing

In some kind of a parsing process, a parser creates a parse tree for such a given text. In order to parse, few parsers need the existence of a set of grammatical rules, while more contemporary parsers are capable of deducing parse trees directly from the provided data using complicated statistical models. Most parsers work in a supervised environment and require that the sentence be POS-tagged before being processed. In the field of natural language processing, statistical parsing is a hot topic.

#### 2.5.2.3 Machine Translation (MT)

The purpose of machine translation is to translate a given text from one natural language to another without the use of a human translator. This is one of the most challenging problems in NLP, and it has been approached in a variety of ways throughout the years. Preliminary steps in almost all MT methods include POS tagging and parsing.

#### 2.5.2.4 Subjective Sentence

Subjective sentences are created when a writer or user conveys their own opinions, feelings, or sentiments about any situations, entities, or events. For example: "We like to give shelter for refugee".

#### 2.5.2.5 Objective Sentence

We call it an objective statement when a writer or user provides facts about occurrences, entities, or events. For example: "Rohingya and Syria crisis made millions of people helpless and".

#### 2.5.2.6 Opinion

Opinion is nothing more than a belief or judgment based on knowledge of a certain subject. Sometimes opinions are called as explicit opinion like: "Refugees are facing dangerous situation of their life". But sometimes hidden in the sentiment of a sentence, for example; "Current refugee problem has no solution yet".

#### 2.5.2.7 Opinion words

To express positive or negative sentiment, opinion words are words that are commonly used. For example: Support, pretty, love Positive sentiment Crisis, stop, hate Negative sentiment.

#### 2.5.2.8 Sentiment Orientation

Sentiment orientation is a term to indicate the expressed opinion by opinion words is positive, negative. For example: "The government good decision for staying them with us" is positive.

#### 2.5.2.9 Opinion Sentence

It is a sentence where it has one or more opinion words. For example: "The government policy was amazing as they were facing political harassment and our government showed humanity".

#### 2.5.2.10 Object / Features

For Example: "A state wants to begin Refugee Crisis for showing off their heroism".

**Object**: heroism
Explicit object- feature: show off heroism.

Here **crisis** word is objective word.
In this example: the explicit feature is voice quality, but sometimes object features should be inferred from the sentence. This kind of feature is called: "implicit feature".

For example: "The crisis is going dangerously"
**Object** crisis

Implicit feature: dangerous
**Opinion** word: dangerously

## 2.6 Sentiment analysis tool: VADER

It was introduced by Vader, Hutto et al. [10] in 2014. VADER is a pre-trained sentiment analysis library with a lexicon and some rule-based scores which can be easily applied to the sentiment data obtained from social media platforms. It contains a list of lexical

features and related sentiment intensity scores. A few rules are created based on grammatical rules and syntactic use in a language to identify sentiment intensity in a given text. The VADER lexicon is mostly a list of terms with semantic labels for each word, such as positive or negative. It also assigns features a score range of to indicate their sentiment polarity and intensity. Later, the sum of scores for $[-4, +4]$ each word is calculated and adjusted according to the rules to reach a normalized compound score in a range of $[-1, +1]$. The calculated compound score displays a flawless performance when it is applied to the data obtained from social media. In this respect, the present study benefits from the VADER method to identify sentiment intensity (positive, neutral, and negative) in a tweet dataset about the Ukraine-Russia war. In the present study, the standardized threshold for sentiment polarities in a sentence is as follows:

$$Fp_i = \begin{cases} Possitive, & v_s \geq .05 \\ Negitive, & v_s \leq .05 \\ Neutral, & otherwise \end{cases} \tag{2.1}$$

For example- Words like 'love', 'enjoy', 'happy', 'like' all convey a positive sentiment. Also VADER is intelligent enough to understand the basic context of these words, such as "did not love" as a negative statement. It also understands the emphasis of capitalization and punctuation, such as "ENJOY"

## 2.7 Machine Learning

Machine learning is a domain of artificial intelligence (AI). It gives the systems the capability to spontaneously improve by learning from the past experiences. Basically machine learning's aim to develop computer programs to access the data and use the data for learning themselves. In modern age machine learning is the most used buzzword.

Firstly with observations from a huge number of data the learning process begins. It looks for patterns in data from direct experiences, instructions, or examples. It makes better decisions from those patterns for the future. Allowing the computers or programs to learn spontaneously and also without any kind of human or manual interference is the main goal, after learning the programs also should take accurate decisions and actions according to the learning.

The aim of machine learning is to program computers to use data or past experience to solve a problem and predict accordingly. In the present era there are lots of applications of machine learning around us, including systems that analyze customer previous buying data and predict customer behavior for future, optimize workers behavior so that they can complete a task using minimum resources, extract knowledge from data and many more. The most useful use of machine learning is in medical fields, it can be used for solving a lot of medical treatment related problems which can decrease the death rate and provide healthy and secure life to the human being. It also used in many more fields like statistics, pattern recognition, neural networks, signal processing, control, artificial intelligence, In order to present a unified treatment of machine learning problems and solutions, it discusses many methods from different fields, including statistics, pattern recognition, neural networks, artificial intelligence, signal processing, control, and data mining[11].

## 2.8 Applied Algorithms

The algorithms that i have used in this research are: 1. Logistic Regression, 2. Naïve Bayes, 3. Decision Tree, 4. Support Vector Machine(SVM). The discussion of them are given below.

### 2.8.1 Logistic Regression

Logistic regression is a supervised machine learning technique which categorizes problems. On labeled datasets, assisted machine learning algorithms are taught, and accuracy is evaluated using an answer key. The goal of the model is to find a mapping function that roughly connects the input variables x1, x2, and xn to the output variable f(Xi) = Y. (Y). The procedure is referred to as supervised learning since the model predictions are continuously evaluated and altered in light of the output values. Using a sample Twitter dataset, we train a sentiment classifier built using logistic regression. The supplied dataset comes in the form of tweets, which is not the easiest format for a model to comprehend. Therefore, in order to transform the provided text into a form that the model can easily understand, some data pre-processing and cleaning will be necessary.

### 2.8.2 Naïve Bayes

Bayesian approaches, according to Wikipedia, are those that use and utilise the Bayes Theorem for probabilistic issues like classification and regression. The Gaussian Naive Bayes classifier is a popular Bayesian method that is frequently used for text categorization in many fields. It is a basic classification technique that uses the Naive Bayes theorem to predict a new document's class [12].

$$P(X|polarity) = \frac{P(X) * P(polarity|X)}{P(polarity)} \tag{2.2}$$

$$P(X|subjectivity) = \frac{P(X) * P(subjectivity|X)}{P(subjectivity)} \tag{2.3}$$

### 2.8.3 Decission Tree

Decision tree algorithm is a data mining induction technique that recursively partitions a data set of records using depth-first greedy approach or breadth-first approach until all the data items belong to a particular class. A decision tree structure is made of root,

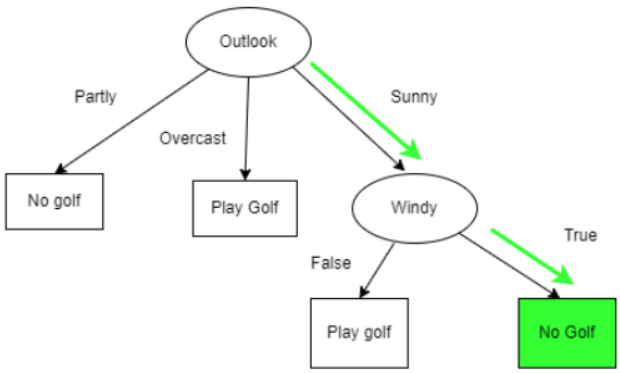


Figure 2.1: An image for understanding Decision Tree[1]

internal and leaf nodes. The tree structure is used in classifying unknown data records. At each internal node of the tree, a decision of best split is made using impurity measures. The tree leaves is made up of the class labels which the data items have been group. Decision tree classification technique is performed in two phases: tree building and tree pruning. Tree building is done in top-down manner. It is during this phase that the tree is recursively partitioned till all the data items belong to the same class label. It is very tasking and computationally intensive as the training data set is traversed repeatedly. Tree pruning is done is a bottom-up fashion. It is used to improve the prediction and classification accuracy of the algorithm by minimizing over-fitting (noise or much detail in the training data set). Over-fitting in decision tree algorithm results in misclassification error. Tree pruning is less tasking compared to the tree growth phase as the training data set is scanned only once. In the proposed system, the decision tree classification provides a better option for the end user to classify the positive and negative tweets. It is done by comparing the maximum frequent items generated by the rules in the training data have been compared with the maximum frequent items of the test data and hence the classification can be made easily.

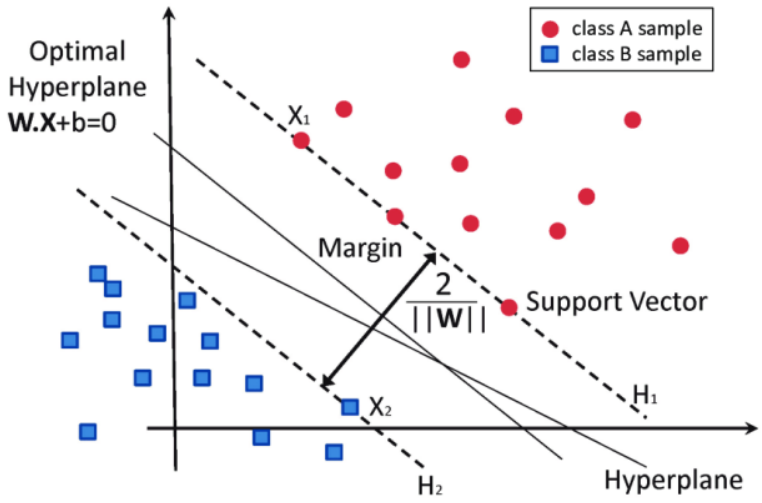


Figure 2.2: An image for understanding SVM [2]

.

### 2.8.4 Support Vector Machine

The supervised(feed-me) machine learning algorithm known as SVM can be applied to classification or regression problems. Regression predicts a continuous value, whereas classification predicts a label or group. By locating the hyper-plane that distinguishes the classes we plotted in n-dimensional space, SVM accomplishes classification.

By converting our data with the aid of mathematical operations referred to as "Kernels," SVM creates that hyperplane. Linear, sigmoid, RBF, non-linear, polynomial, etc., are some types of kernels. When there is no prior knowledge of the data, the general-purpose kernel Kernel — "RBF" is employed as a tuning parameter for non-linear situations. Kernel — " linear" is for issues with linear separability. We will use "linear SVM" since our problem is linear (consisting only of positive and negative values). Steps needed to build a model:

- Perfect Data gathering for training and testing
- Data vectorizing
- • Creating a Linear SVM Model to train and then predict

# Chapter 3

# Literature Review

## 3.1 Introduction

It's high time to examine literatures that are relevant to our study once we've learned more about context, goal, contribution, and aim of our research. Following that, we discussed various sentiment analysis research papers. We've included the paper's title first, followed by a review of their work.

## 3.2 Related Works

In this part, we will offer an overview of current studies in the literature:

There are three categories of sentiment analysis techniques used nowadays are machine learning, dictionary-based, and deep learning-based. In the recent issue Russia-Ukraine war there are few sentiment analysis work based on deep learning method. In this research work [13] author López Ramírez et al. provide a sentiment analysis approach that can effectively distinguish between positive and negative sentiment analysis on Russia-Ukraine conflict. In this paper, authors also used the same dataset that we used. For score the sentiment they also used VADER sentiment analyzer tool. They used RNN model that can accurately classify sentiment value with 93% accuracy. The efficiency of

this model may also be determined using various huge datasets.

Table 3.1: Existing works on Sentiment Analysis

| Paper Name | Author Name | Year | Algorithm | Accuracy (%) | Limitations |
|---|---|---|---|---|---|
| A sentiment analysis of the Ukraine-Russia conflict tweets using Recurrent Neural Networks. [13] | López Ramírez et al. | 2022 | RNN | 93% | Complex in large dataset |
| MF-CNN-BILSTM: A Deep Learning-Based Sentiment Analysis Approach and Topic Modeling of Tweets Related to the Ukraine-Russia Conflict. [14] | Aslan, Serpil et al. | 2022 | CNN, LSTM, Bi-LSTM | 89.5%(CNN), 91.5%(LSTM), 91.5%(Bi-LSTM) | Complex in large dataset |
| Twitter sentiment analysis on coronavirus: Machine learning approach. [15] | Machuca, Cristian R et al. | 2021 | Logistic Regression | 78.57% | Low accuracy |
| US Based COVID-19 tweets sentiment analysis using textblob and supervised machine learning algorithms [16] | Khan, Rashid, et al. | 2021 | SVM, LR | 92%(LR), 94%(SVM) | Low accuracy |
| Product review sentiment analysis by using NLP and machine learning in Bangla language [17] | Shafin, Minhajul Abedin, et al. | 2020 | KNN, SVM, DT, LR | 88.81%(SVM), 83.03%(DT), 88.09%(LR) | Low accuracy |
| Comparison study of sentiment analysis of tweets using various machine learning algorithms [18] | Kanakaraddi, Suvarna G., et al. | 2020 | SVM, NB, DT | 79.90%(SVM), 70.58%(NB), 75.98%(DT) | Low accuracy |

In the paper titled as 'MF-CNN-BILSTM: A Deep Learning-Based Sentiment Analysis Approach and Topic Modeling of Tweets Related to the Ukraine-Russia Conflict' [14] authors used CNN, LSTM and Bi-LSTM model to predict the public emotion on Russia-Ukraine War. The accuracy their work 89.5% for CNN, 91.5% for LSTM and 91.5% for

Bi-LSTM. They also used UKraine-Russia conflict based dataset.

In the research work [15] Machuca, Cristian R et al used machine learning approach to sentiment analysis on twitter sentiment analysis.They used Logistic regression. The accuracy of their work is 78.57% only.

In the research work [16] Khan, Rashid, et al. used TextBlob and machine learning algorithm to sentiment analysis on twitter sentiment analysis.They used Logistic regression and Support Vector Machine. The accuracy rate of SVM is 94% and the accuracy of Logistic Regression is 92%

In the paper titled as 'Product review sentiment analysis by using NLP and machine learning in Bangla language' [17] authors Shafin, Minhajul Abedin, et al. work to predict sentiment analysis on twitter data. They used different types of algorithm. They used KNN, SVM, Random Forest, DT, LR in their work. The accuracy rate of Decission Tree algorithm is 83.03%. The accuracy rate of support vector machine is 88.81%. The accuracy rate of Logistic Regression is 88.09% in their work.[19]

'Comparison study of sentiment analysis of tweets using various machine learning algorithms'[18] provides various machine learning approaches perfomance analysis on twitter data. In this work they used Support Vector Machine, Naive Bayes and Decission Tree. The accuracy rate of these algorithm are respectably 79.90%(SVM), 70.58%(NB), 75.98%(DT).

In the above study we can see that the accuracy of machine learning algoritm is low. The accuracy of deep learning is better but if the dataset are large then the model can not work accurately .

## 3.3 Advantage and Limitation of Existing Work

The strong point of the existing analysis of sentiment which are as follows:

- In many experiment researchers made their own system to work with Text Mining and Sentiment Analysis.
- From their working experience we got to understand that big data can be handled by using different tools.
- Researcher gave hint about RapidMiner that is data science workable environment and right gateway to work with data.
- Existing research can help to find out the best way of marketing, reviewing.
- In some research they worked with 5 classes of polarity.

Though we noticed many advantages in existing research, however we have found some drawback in existing environment and system. Here is some:

- As It is difficult for machine to find out the accurate sentiment from a text every time. Sometime prediction may not true.
- Most of the time existing research is done by working with smaller data set. Smaller dataset can be handled without facing any difficulty and hardships.
- While in performing Sentiment Analysis sometimes they concerned only positivity of the text and then made the rest of data as negative.
- The classification result they got can be improved.

## 3.4 Summary

In this chapter we discussed about different existing paper on sentiment analysis. We have tried to give some idea about some paper. We've also discussed about strong and weak point about existing work.

# Chapter 4

# System Methodology and Implementation

## 4.1 Methodology Overview

Methodology is the process of how we work implemented and modeling refers, can be explained as elaborated description of followed strategy and diagram. Each of part described in details.

## 4.2 Overview of the full System Design

In the following (Fig 4.1) flow chart shows the overall procedure for sentimental analysis. The main goals of this research to analyze people's emotion and opinion about Russia-Ukraine War and proposed a model that gives better accuracy.

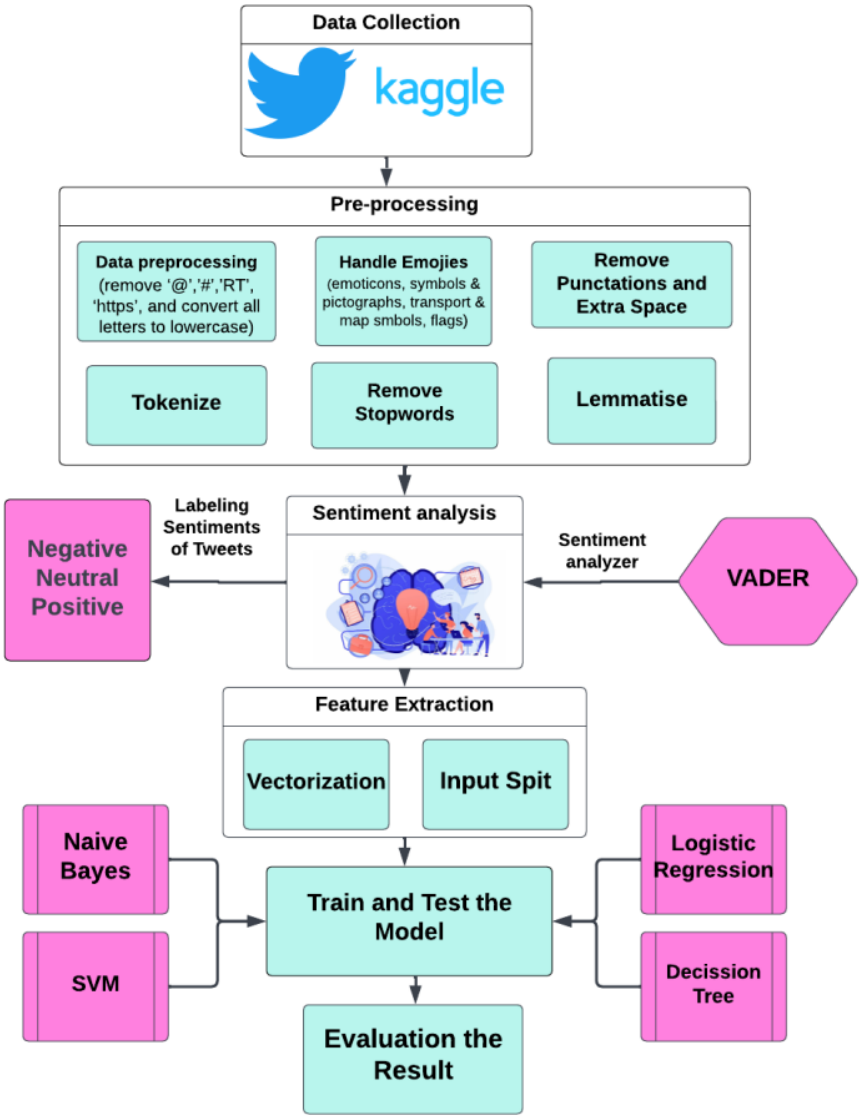


Figure 4.1: Block diagram of Sentiment analysis process steps.

## 4.3 Data Description

We collected the dataset from Kaggle with the name "Ukraine Conflict Twitter Dataset" which Contain around 53 million tweets. In this research we Worked for 3,64,875 tweets.There are 18 columns with data types int64 and object in this dataset.Mainly we uses tweets attribute for our work.

## 4.4 Dataset Partitioning

Because of large dataset we divided the data into two parts: training and testing. Trained tweets aid in the detection of data trends, the minimization of error percentages, and the testing of the data set used to measure model performance. We use 80% data for training purposed and 20% data for testing purpose.

## 4.5 Data Preprocessing

The data which extracted from Twitter or other social media website contains different non-sentimental contents such as duplicate tweets, website links, emotions, mentioned symbol or username, white spaces, retweet tag, hashtag etc. which is removed before processing my tweets so that the sentiment can generate accurately. Preprocessing step include followings:

### 4.5.1 Removal of Additional White Spaces

There may be consists of extra white space in the data and it needs to be removed. By removing white spaces, the analysis to be done more efficiently.

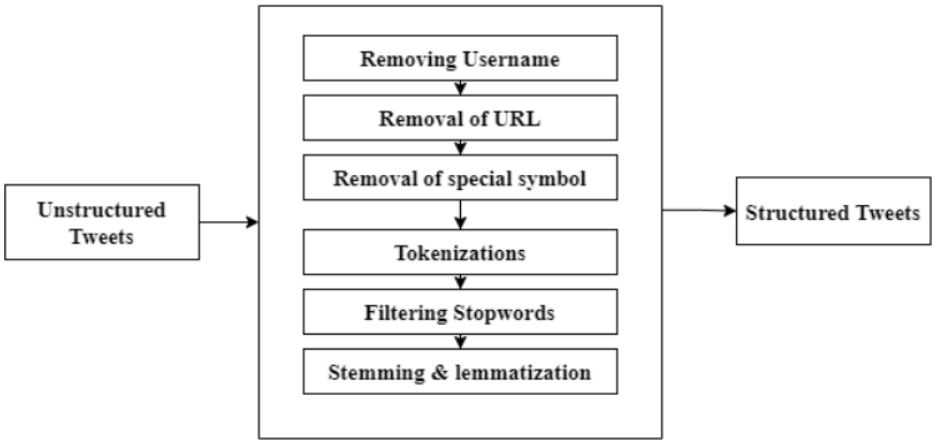


Figure 4.2: Data preprocessing diagram.

### 4.5.2 Duplicate Tweets Removal

Every day peoples are sharing huge number of tweets. When we extract their shared tweets from social media platform, sometime one tweet comes multiple times. This is what we called duplicated tweets and that should be removed for our desired analysis.

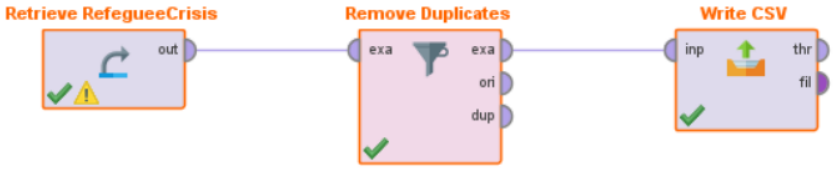


Figure 4.3: Duplicate Tweets Removal diagram.

### 4.5.3 Removal of Website Link

Extracted Twitter data so called tweets consist of different type of information what we called as URL. If Tweeter user posted any video, audio, article link which is unnecessary for use in sentiment analysis. Therefore, that URL should be removed from my tweet. The URL we found here many as YouTube and Facebook's video link.

### 4.5.4 Special Symbol Removing Part

There are different types of symbols what we called as special symbol used by the social media user. We removed punctuation mark (!), full stop (.) mentioning symbol (@), single quote (“”), single quote (‘’) etc. which does not contain sentiment. So, we removed all the special symbols from my tweet dataset by following the RapidMiner provided operator with the help of regular expression.

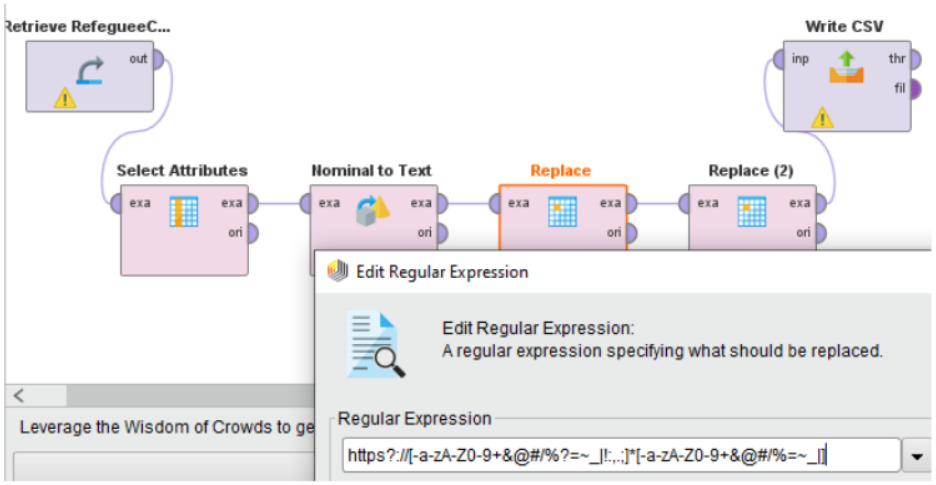


Figure 4.4: Symbol removal diagram.

### 4.5.5 Username Removing Part

One user can use one username and that is should be unique by following the guide line of Twitter. So, anything is posted by a user there is his/her username proceeding by @ which is used as proper nouns. For example, @someones_username. This also removed from my dataset for effective analysis.

### 4.5.6 Removal of Stop Words

"Stop words" means some common words that don't carry useful information of sentence like "the," "a," and "and". Removing stop words means that model won't able to see these words and will be trained on a clean dataset. Text Analysis Platform, TAP recognizes stopwords based on a manually curated, comprehensive list, but since what defines a stopword can vary depending on context. Since shorter documents like tweets contain such little text, for training a model filtering stop word is needed.


Figure 4.5: Removal of stop words.

Following stop words are removed from our dataset: I, me, my, myself, we, our, ours, ourselves, you, your, yours, yourself, yourselves, he, him, his, himself, she, her, hers, herself, it, its, itself, they, them, their, theirs, themselves, what, which, who, whom, this, that, these, those, am, is, are, was, were, be, been, being, have, has, had, having, do, does, did, doing, a, an, the, and, but, if, or, because, as, until, while, of, at, by, for, with,

about, against, between, into, though, during, before, after, above, below, to, from, up, down, in, out, on, off, over, under, again, further, then, once, here, there, when, where, why, how, all, any, both, each, few, more, most, other, some, such, no, nor, not, only, own, same, so, than, too, very, can, will, just, don, should, now.

## 4.6 Methodology of Sentiment Analyzer tool(VADER)

As our dataset is not labeled by Sentiment score, we applied Sentiment analyzer tool called VADER(Valence Aware Dictionary and sEntiment Reasoner) to score the tweets after preprocess the dataset.The flow chart of VADER is given below:

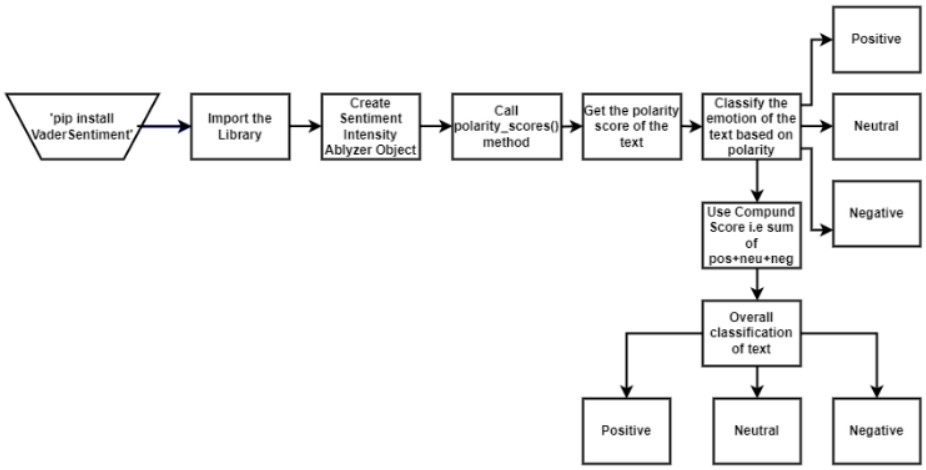


Figure 4.6: Flow Chart of VADER.

## 4.7 Feature Extraction

The process of translating raw data into numerical features that can be examined while keeping the information in the original data set is known as feature extraction. It give

better outcomes than applying machine learning to raw data directly. So the next stage in the procedure is to extract the appropriate features from the processed text. There are many way to extract features: such as CountVectorizer, Term Frequency Inverse Document Frequency (TFIDF), Word2Vec,Bag of Words(BoW).

In Our work, we used two method to extract features.These are BoW and TF-IDF

## 4.8 Model Training

We provide an experimental research that compares various machine learning model for multilable categorization. We use 80% of the dataset for training purposed, while 20% of the dataset for testing purposed.Accuracy, F1-Score, Recall and Precision are four metric parameters used to evaluate the outcomes achieved for each combination.

## 4.9 Machine Learing based algorithms

Different machine learning model are utilized in this research work. Diverse classification algorithms are supported by machine learning algorithms like Logistic Regression, Naive bayes, Decision Tree, SVM etc.

## 4.10 Evaluation Metrics

Evaluation metrics measure the performance of prediction model. In this study, four performance metrics are conducted to measure the performance results of the proposed architecture. We used evaluation measures are accuracy, precision, recall f1-score. To better understand a model performance we must analyze the performance via evaluation metrics.

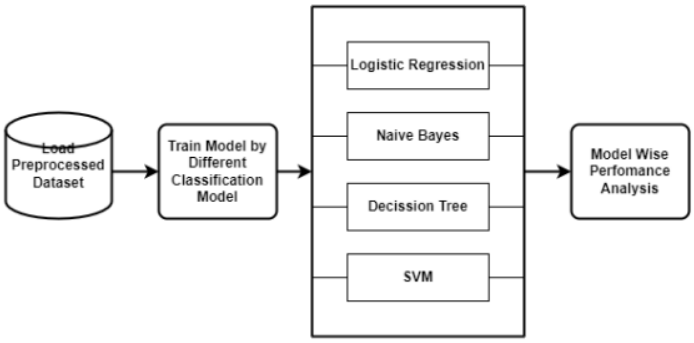


Figure 4.7: Flow Diagram of Different Machine Learning Model.

### 4.10.1 Confusion Matrix

This yields a matrix that describes the overall performance of the model.

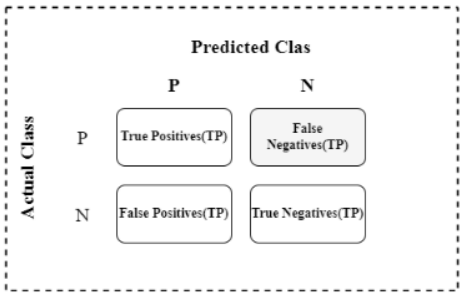


Figure 4.8: Confusion Matrix.

1. **True Positives** occur when the prediction is YES and the actual output is YES.
2. **True Negatives** occur when the prediction is NO and the actual output is NO.
3. **False Positives** occur when the prediction is YES but the actual output is NO.

4. **False Negatives** occur when the prediction is NO and the actual output is YES.

### 4.10.2 Accuracy

The number of correct predictions divided by the total number of input samples is the appropriate prediction ratio.

The matrix's accuracy is computed as following equation 4.1.

$$\frac{(TP + TN)}{(TP + FP + TN + FN)} \tag{4.1}$$

### 4.10.3 Precision

When compared to the total number of values that were correctly predicted, equation 4.2 measure displays the number of True Positives that are in fact positive.

$$\frac{TP}{(TP + FP)} \tag{4.2}$$

### 4.10.4 Recall

The Recall Metric displays the number of True Positives categorized by the model out of the total number of samples that should have been positive.

$$\frac{TP}{(TP + FN)} \tag{4.3}$$

### 4.10.5 F1 Score

The F1 Score is the Harmonic Mean of accuracy and recall. The F1 Score ranges from 0 to 1. It indicates the precision of your classifier, i.e. how many instances it properly classifies, as well as the robustness of your classifier, i.e. how many examples it does not miss.

$$F1 - Score = \frac{(2 * precision * recall)}{(precision + recall)} \tag{4.4}$$

# Chapter 5

# Results and Discussions

## 5.1 Introduction

In this chapter we discuss the outcomes of different model used for sentimental analysis. The outcomes are evaluated in light of the topic of this research, as well as their relevance to the issue statement and methodology.

## 5.2 Data Analysis and Visualization

### 5.2.1 Data Preprocessing

We performed different preprocessing techniques on this dataset. After cleannig the data and labeled the sentiment are shown in figure (Fig 5.1).

| | Unnamed: 0 | clean_tweet | sentiment1 | sentiment |
|---|---|---|---|---|
| 0 | 0 | ukrainian air forc would like address misinfor... | neutral | 1.0 |
| 1 | 1 | chernihiv oblast ukrainian welcom liber russia... | neutral | 1.0 |
| 2 | 2 | america prepar someth wors russianukrainianwar... | positive | 2.0 |
| 3 | 3 | anonym hack amp releas email marathon group ru... | positive | 2.0 |
| 4 | 4 | public mint livefor billionaire_wmnwin public ... | positive | 2.0 |
| ... | ... | ... | ... | ... |
| 184882 | 184882 | western spi agenc weapon intellig attempt unde... | negative | 0.0 |
| 184883 | 184883 | video social media allegedli eastern ukrain sh... | positive | 2.0 |
| 184884 | 184884 | ist es klar dass die zahlungen fr ga mit ber d... | neutral | 1.0 |
| 184885 | 184885 | el decreto firmado por putin faculta gazprom c... | neutral | 1.0 |
| 184886 | 184886 | putin biden chhattisgarh riteshmishraht thealo... | neutral | 1.0 |

Figure 5.1: After cleaning and labelized the data.

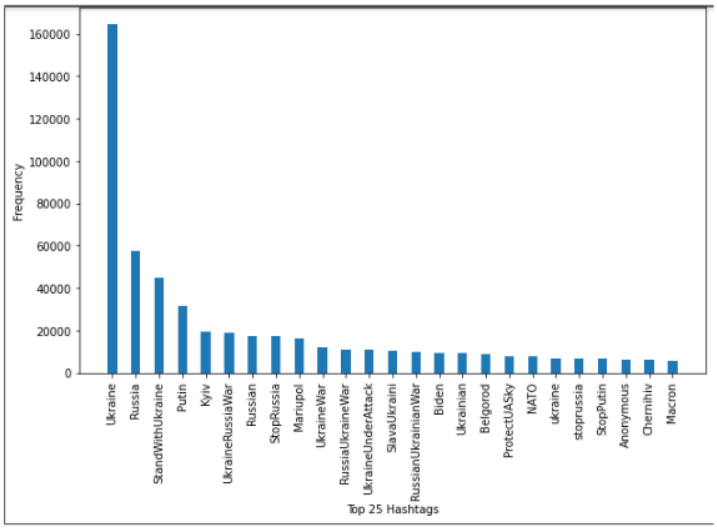


Figure 5.2: Top 25 hashtags

### 5.2.2 Visualized the data

Data are visualized in many way. we can use some visualized technique to show the data.Figure(5.2) represent the top 25 hashtag in dataset.Figure (5.4) show the number of positive, negative and neutral sentiment in the dataset.In this (Fig5.3)shows the word cloud of sample dataset.Figure (5.4 represent the distribution of Tweets based on Sentiment which we got by using VADER sentiment analysis tool.


((a)) WordCloud for all word

((b)) WordCloud for positive word

((c)) WordCloud for negative word

((d)) WordCloud for neutral word

Figure 5.3: Sample Data Word Cloud

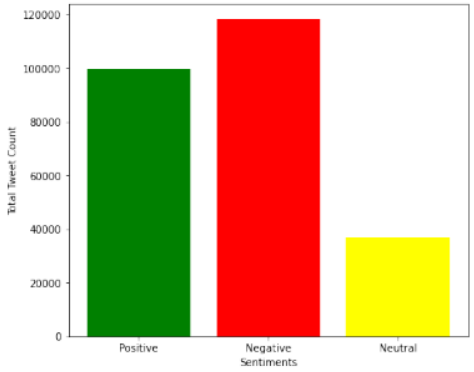


Figure 5.4: Distribution of Tweets based on Sentiment

## 5.3 Splitting dataset in train and test

Because the large dataset we can take 80% as train data and 20% of test data.

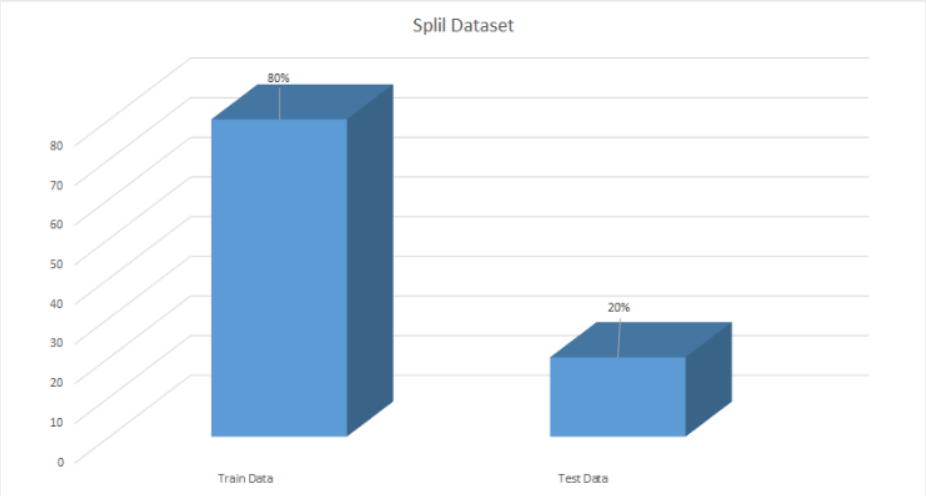


Figure 5.5: Split Dataset.

## 5.4 Evaluation of The Model

In this work, We utilized different machine learning models (Logistic Regression, Naive Bayes,Decission Tree and SVM) with two types of Vectorization.

### 5.4.1 Vectorization with BoW and Classification

In (Fig 5.6), (Fig 5.7), (Fig 5.8) and (Fig 5.9) show the confusion matrics with Bow vectorization. It is a technique for summarizing performance of classification algorithm.

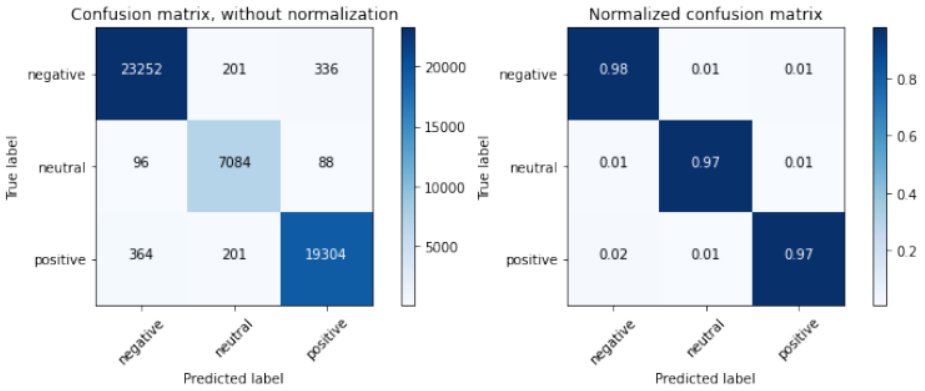


Figure 5.6: Confusion Matrix of Logistic Regression with BoW

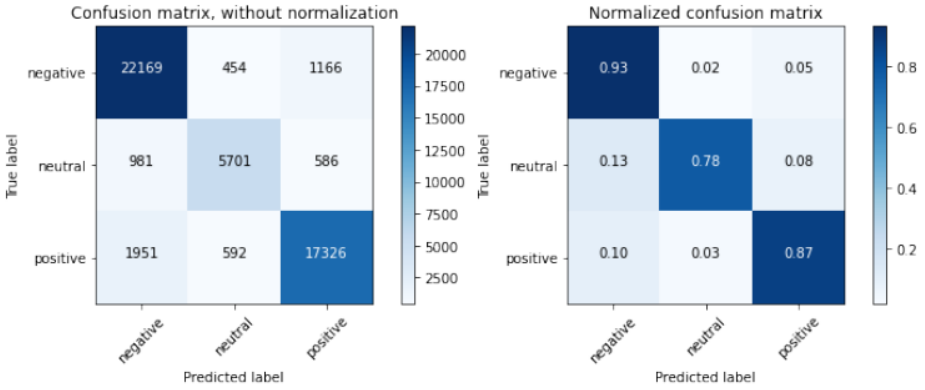


Figure 5.7: Confusion Matrix of Naïve Bayes with Bow

### 5.4.2 Vectorization with TF-IDF and Classification

In (Fig 5.10), (Fig 5.11), (Fig 5.12) and (Fig 5.13) show the confusion matrics with TF-IDF vectorization. It is a technique for summarizing performance of classification algorithm.

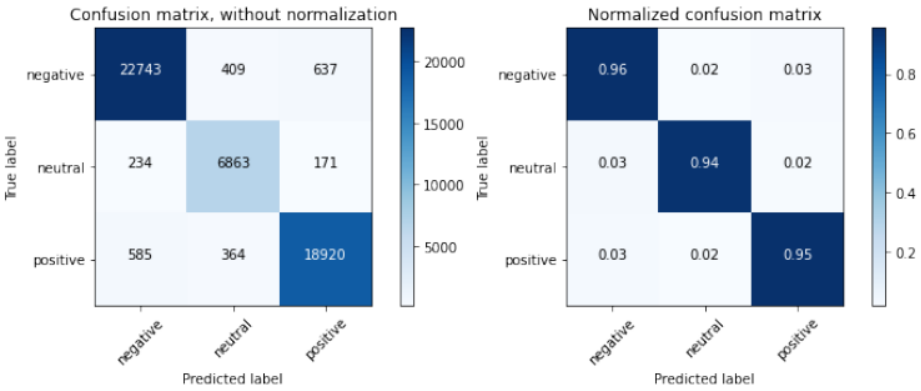


Figure 5.8: Confusion Matrix of Decission Tree with Bow

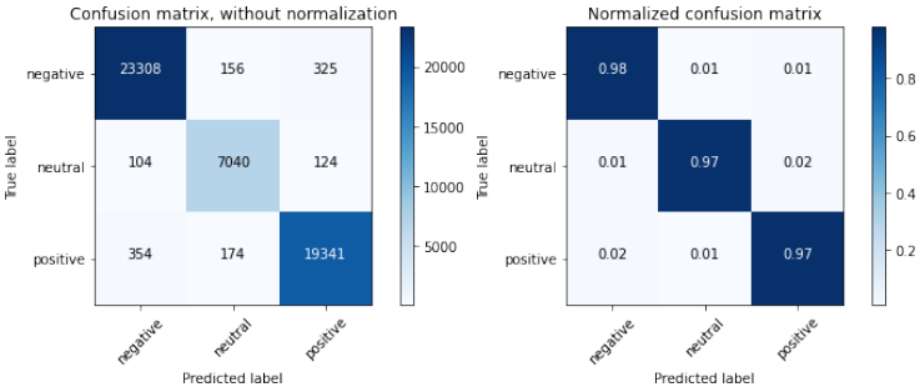


Figure 5.9: Confusion Matrix of SVM with Bow

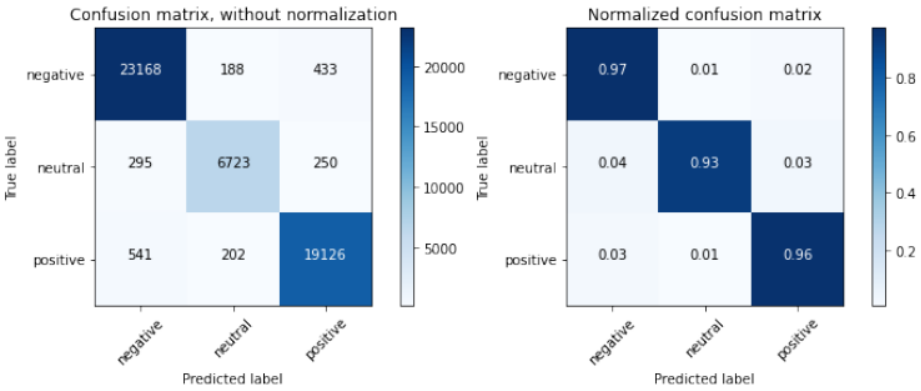


Figure 5.10: Confusion Matrix of Logistic Regression with TF-IDF

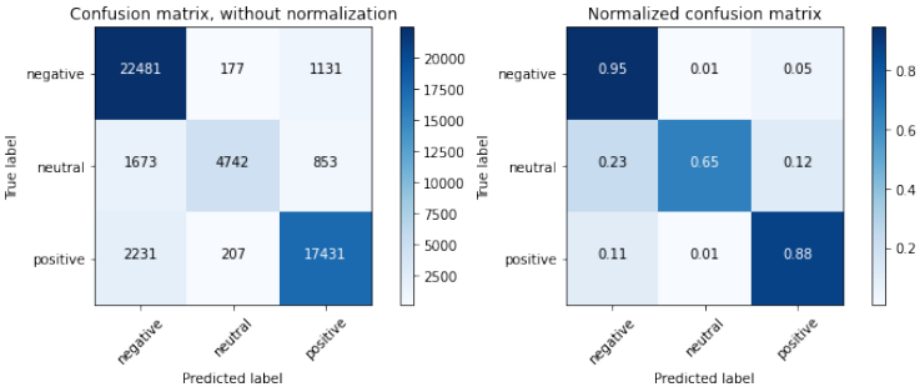


Figure 5.11: Confusion Matrix of Naïve Bayes with TF-IDF

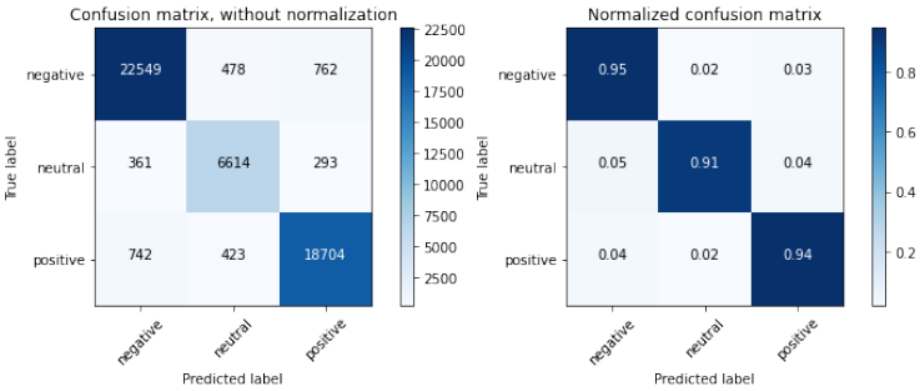


Figure 5.12: Confusion Matrix of Decission Tree with TF-IDF

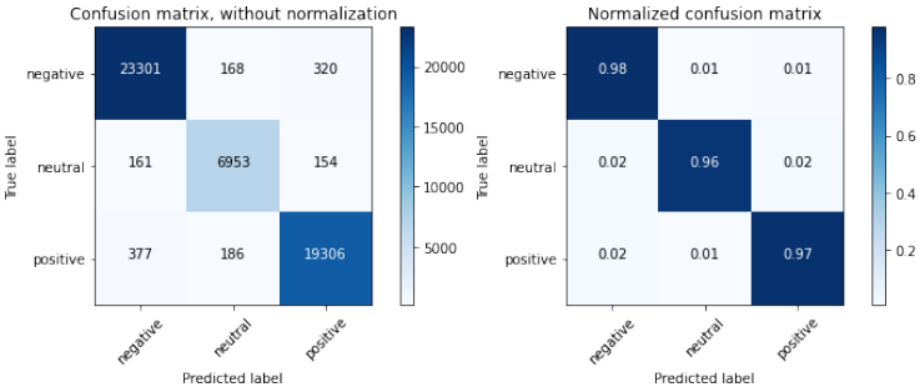


Figure 5.13: Confusion Matrix of SVM with TF-IDF

## 5.5 Model Wise Performance Analysis

In the Table 5.1 and 5.2shows the model wise performance in this research work. From this table 5.1 we shows that we achieved the highest accuracy from Support Vector Machine with BoW vectorization.Its accuracy is 98%. Using BoW vectorization technique we achieved highest precision value of .98 for negtive sentiment status by considering Support vector Machine and .98 for positive sentiment status by considering Support vector machine and precision value for neutral is .96 . The highest recall value 0.98 achieved for positive considering Support vector Machine and .98 for negative sentiment status by considering SVM. The highest F1-score is .98 for negtive sentiment status by considering Support vector Machine and .98 for positive sentiment status by considering Support vector machine and precision value for neutral is .96.

For considering Tfidf technique we achieved highest accuracy 97% from Support Vector Machine.5.2 Using TF-IDF vectorization technique we achieved highest precision value of .98 for negtive sentiment status by considering Support vector Machine and .98 for positive sentiment status by considering Support vector machine and precision value for neutral is .95 . The highest recall value 0.98 achieved for positive considering Support vector Machine and .98 for negative sentiment status by considering SVM. The highest F1-score is .98 for negtive sentiment status by considering Support vector Machine and .98 for positive sentiment status by considering Support vector machine and F1-score for neutral is .95.

| Model | Accuracy | Class | Precision | Recall | F1-score | Support |
|---|---|---|---|---|---|---|
| LR | 0.97 | Negative(0) | **0.98** | **0.98** | **0.98** | 23789 |
| | | Neutral(1) | 0.95 | 0.97 | 0.96 | 7268 |
| | | Positive(2) | **0.98** | 0.97 | **0.98** | 19869 |
| | | Macro avg | 0.97 | 0.97 | 0.97 | 50926 |
| | | Weighted avg | 0.97 | 0.97 | 0.97 | 50926 |
| NB | 0.89 | Negative(0) | 0.88 | **0.93** | **0.91** | 23789 |
| | | Neutral(1) | 0.84 | 0.78 | 0.81 | 7268 |
| | | Positive(2) | **0.91** | 0.87 | 0.89 | 19869 |
| | | Macro avg | 0.88 | 0.86 | 0.87 | 50926 |
| | | Weighted avg | 0.89 | 0.89 | 0.89 | 50926 |
| DT | 0.95 | Negative(0) | **0.98** | **0.96** | **0.96** | 23789 |
| | | Neutral(1) | 0.95 | 0.94 | 0.92 | 7268 |
| | | Positive(2) | **0.98** | 0.95 | **0.96** | 19869 |
| | | Macro avg | 0.97 | 0.95 | 0.95 | 50926 |
| | | Weighted avg | 0.97 | 0.95 | 0.95 | 50926 |
| SVM | 0.98 | Negative(0) | **0.98** | **0.98** | **0.98** | 23789 |
| | | Neutral(1) | 0.96 | 0.97 | 0.96 | 7268 |
| | | Positive(2) | **0.98** | 0.97 | **0.98** | 19869 |
| | | Macro avg | 0.97 | 0.97 | 0.97 | 50926 |
| | | Weighted avg | **0.98** | **0.98** | **0.98** | 50926 |

Table 5.1: Report of Machine Learning model with BoW vectorization

| Model | Accuracy | Class | Precision | Recall | F1-score | Support |
|---|---|---|---|---|---|---|
| LR | 0.96 | Negative(0) | **0.97** | **0.97** | **0.97** | 23789 |
| | | Neutral(1) | 0.95 | 0.93 | 0.93 | 7268 |
| | | Positive(2) | **0.97** | 0.96 | 0.96 | 19869 |
| | | Macro avg | 0.96 | 0.95 | 0.96 | 50926 |
| | | Weighted avg | 0.96 | 0.96 | 0.96 | 50926 |
| NB | 0.88 | Negative(0) | **0.85** | **0.95** | **0.90** | 23789 |
| | | Neutral(1) | 0.93 | 0.65 | 0.77 | 7268 |
| | | Positive(2) | 0.90 | 0.88 | 0.89 | 19869 |
| | | Macro avg | 0.89 | 0.82 | 0.85 | 50926 |
| | | Weighted avg | 0.88 | 0.88 | 0.87 | 50926 |
| DT | 0.94 | Negative(0) | **0.95** | **0.95** | **0.95** | 23789 |
| | | Neutral(1) | 0.88 | 0.65 | 0.89 | 7268 |
| | | Positive(2) | **0.95** | 0.88 | 0.94 | 19869 |
| | | Macro avg | 0.93 | 0.82 | 0.93 | 50926 |
| | | Weighted avg | 0.94 | 0.88 | 0.94 | 50926 |
| SVM | 0.97 | Negative(0) | **0.98** | **0.98** | **0.98** | 29610 |
| | | Neutral(1) | 0.95 | 0.96 | 0.95 | 9149 |
| | | Positive(2) | **0.98** | 0.97 | 0.97 | 24898 |
| | | Macro avg | 0.97 | 0.97 | 0.97 | 63657 |
| | | Weighted avg | 0.97 | 0.97 | 0.97 | 63657 |

Table 5.2: Report of Machine Learning model with TF-IDF vectorization

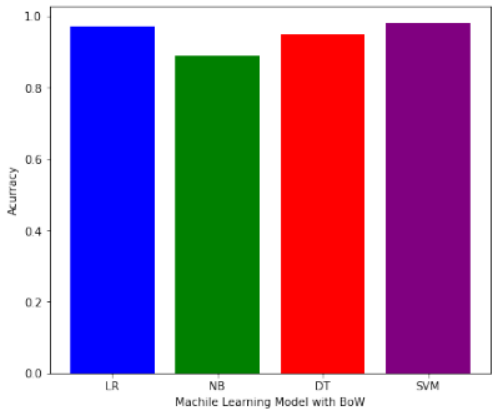


Figure 5.14: Performance comparison of techniques in terms of accuracy with BoW Vectorization

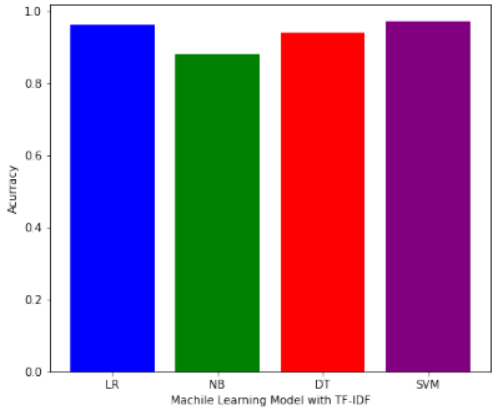


Figure 5.15: Performance comparison of techniques in terms of accuracy with TF-IDF Vectorization

## 5.6 Comparison of Existing work

In this table (5.3) represents the performance of exiting work and our research work. In this section we can see that different paper worked with different model and their outcomes where the best accuracy is 94%. From our work we preffer Support Vector Mschine for sentiment analysis with BoW vectorization as we get the 98% accuracy .

| **Related Work** | | **This Work** | |
|---|---|---|---|
| Model | Accuracy(Best) | Model | Accuracy |
| Support Vector Machine | 94% | Support Vector Machine | 98% |
| Logistic Regression | 88.09% | Logistic Regression | 97% |
| Decision Tree | 83.03% | Decision Tree | 95% |
| Naïve Bayes | 70.58% | Naïve Bayes | 89% |
| RNN | 93% | | |
| CNN,LSTM | 91.5% | | |

Table 5.3: Performance comparison of techniques in terms of accuracy

# Chapter 6

# Conclusions and Recommendations

## 6.1 Conclusions

The present study was a detailed analysis on machine learning techniques for sentiment analysis on Russia-Ukraine war issue. As an outcome, we can observe the changes in the results chapter, where the Machine learning based Support Vector Machine provided the most accurate and best result in support and against other approaches.

This work analyze public opinion on war and visualization of debated topics and sentiment polarities will enable governments to recognize concerns and take necessary precautions according to changing conditions. The sentiment analysis helps us understand that, as predicted, a high percentage of the tweets related to the Russia-Ukraine war were negative. Even though the current circumstances are horrible, people around the world, one way or another, tried to contribute by spreading some positiveness in their tweets.

## 6.2 Limitation and Future Work

In this work we see that it's tough for a machine to always extract the correct sentiment from a sentence.

We are willing to provide such a user interface for sentimental analysis so that individuals may quickly assess sentimental analysis in a variety of contexts.

We want to create a Bangla natural language processing framework where we can create a list of Bangla stop words and portions of a tagging model for solely studying Bangla text.

# Bibliography


[1] "Intuitive guide to understanding decision trees," https://towardsdatascience.com/light-on-math-machine-learning-intuitive-guide-to-understanding-decision-trees-adb2165ccab7, accessed: 2020-12-12.

[2] E. García-Gonzalo, Z. Fernández-Muñiz, P. J. Garcia Nieto, A. Sánchez, and M. Menéndez, "Hard-rock stability analysis for span design in entry-type excavations with learning classifiers," *Materials*, vol. 9, p. 531, 06 2016.

[3] A. A. Müngen, İ. Aygün, and K. Mehmet, "Finding the relationship between news and social media users' emotions in the covid-19 process," *Sakarya University Journal of Computer and Information Sciences*, vol. 3, no. 3, pp. 250–263, 2020.

[4] A. Pak and P. Paroubek, "Twitter as a corpus for sentiment analysis and opinion mining," in *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*, 2010.

[5] S. M. Graber, "War of perception: a habermasian discourse analysis of human shield newspaper reporting during the 2014 gaza war," *Critical Studies in Media Communication*, vol. 34, no. 3, pp. 293–307, 2017.

[6] R. P. Haseena, "Opinion mining and sentiment analysis challenges and applications," 2014.

[7] A. V. Kristen McCabe, "Predictions to drive success," 2022. [Online]. Available: https://learn.g2.com/business-forecasting

[8] D. M. E.-D. M. Hussein, "Analyzing scientific papers based on sentiment analysis," *Information System Department Faculty of Computers and Information Cairo University, Egypt*, 2016.

[9] S. Bird, E. Klein, and E. Loper, *Natural language processing with Python: analyzing text with the natural language toolkit.* " O'Reilly Media, Inc.", 2009.

[10] C. Hutto and E. Gilbert, "Vader: A parsimonious rule-based model for sentiment analysis of social media text," in *Proceedings of the international AAAI conference on web and social media*, vol. 8, no. 1, 2014, pp. 216–225.

[11] E. Alpaydin, *Introduction to Machine Learning*, 2nd ed. The MIT Press, 2010.

[12] Z. Qin, "Naive bayes classification given probability estimation trees," in *2006 5th International Conference on Machine Learning and Applications (ICMLA'06)*. IEEE, 2006, pp. 34–42.

[13] I. López Ramírez and J. Méndez Vargas, "A sentiment analysis of the ukraine-russia conflict tweets using recurrent neural networks," 06 2022.

[14] S. Aslan, "Mf-cnn-bilstm: A deep learning-based sentiment analysis approach and topic modeling of tweets related to the ukraine-russia conflict," *Available at SSRN 4218398*.

[15] C. R. Machuca, C. Gallardo, and R. M. Toasa, "Twitter sentiment analysis on coronavirus: Machine learning approach," in *Journal of Physics: Conference Series*, vol. 1828, no. 1. IOP Publishing, 2021, p. 012104.

[16] R. Khan, F. Rustam, K. Kanwal, A. Mehmood, and G. S. Choi, "Us based covid-19 tweets sentiment analysis using textblob and supervised machine learning algo-

rithms," in *2021 international conference on artificial intelligence (ICAI)*. IEEE, 2021, pp. 1–8.

[17] M. A. Shafin, M. M. Hasan, M. R. Alam, M. A. Mithu, A. U. Nur, and M. O. Faruk, "Product review sentiment analysis by using nlp and machine learning in bangla language," in *2020 23rd International Conference on Computer and Information Technology (ICCIT)*. IEEE, 2020, pp. 1–5.

[18] S. G. Kanakaraddi, A. K. Chikaraddi, K. C. Gull, and P. Hiremath, "Comparison study of sentiment analysis of tweets using various machine learning algorithms," in *2020 International Conference on Inventive Computation Technologies (ICICT)*. IEEE, 2020, pp. 287–292.

[19] M. B. Garcia and A. Cunanan-Yabut, "Public sentiment and emotion analyses of twitter data on the 2022 russian invasion of ukraine," in *2022 9th International Conference on Information Technology, Computer, and Electrical Engineering (ICITACEE)*. IEEE, 2022, pp. 242–247.

# Defense_Shahin_Ashraful_ (6).pdf

ORIGINALITY REPORT

26%
SIMILARITY INDEX

PRIMARY SOURCES

| | | |
|---|---|---|
| 1 | www.researchgate.net<br>Internet | 303 words — 3% |
| 2 | jcreview.com<br>Internet | 232 words — 2% |
| 3 | researchonline.federation.edu.au<br>Internet | 124 words — 1% |
| 4 | vdoc.pub<br>Internet | 105 words — 1% |
| 5 | papers.ssrn.com<br>Internet | 81 words — 1% |
| 6 | businessinspection.com.bd<br>Internet | 79 words — 1% |
| 7 | www.journals.pen2print.org<br>Internet | 69 words — 1% |
| 8 | Babacar Gaye, Dezheng Zhang, Aziguli Wulamu. "Sentiment classification for employees reviews using regression vector- stochastic gradient descent classifier (RV-SGDC)", PeerJ Computer Science, 2021<br>Crossref | 68 words — 1% |
| 9 | dspace.daffodilvarsity.edu.bd:8080<br>Internet | |

67 words — 1%

| | | |
|---|---|---|
| 10 | dspace.ewubd.edu:8080<br>Internet | 57 words — 1% |
| 11 | Omar Ali, Alexander Gegov, Ella Haig, Rinat Khusainov. "Conventional and Structure Based Sentiment Analysis: A Survey", 2020 International Joint Conference on Neural Networks (IJCNN), 2020<br>Crossref | 48 words — < 1% |
| 12 | www.cse.cuhk.edu.hk<br>Internet | 46 words — < 1% |
| 13 | ijettcs.org<br>Internet | 45 words — < 1% |
| 14 | Md. Hamidur Rahman, Md. Saiful Islam, Md. Mine Uddin Jowel, Md. Mehedi Hasan, Ms. Subhenur Latif. "Classification of Book Review Sentiment in Bangla Language Using NLP, Machine Learning and LSTM", 2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT), 2021<br>Crossref | 44 words — < 1% |
| 15 | Artur Sokolovsky, Thomas Gross, Jaume Bacardit. "Is it feasible to detect FLOSS version release events from textual messages? A case study on Stack Overflow", PLOS ONE, 2021<br>Crossref | 43 words — < 1% |
| 16 | repository.ihu.edu.gr<br>Internet | 43 words — < 1% |
| 17 | Ludeman, John. "Runway Location for Autonomous Aircraft Using Neural Networks and | 35 words — < 1% |

Computer Vision", University of Nebraska at Omaha, 2022
ProQuest

| | | |
|---|---|---|
| 18 | kth.diva-portal.org<br>Internet | 35 words — < 1% |
| 19 | saucis.sakarya.edu.tr<br>Internet | 35 words — < 1% |
| 20 | thesai.org<br>Internet | 35 words — < 1% |
| 21 | www.ijraset.com<br>Internet | 34 words — < 1% |
| 22 | Wael F. Al-Sarraj, Heba M. Lubbad. "Bias Detection of Palestinian/Israeli Conflict in Western Media: A Sentiment Analysis Experimental Study", 2018 International Conference on Promising Electronic Technologies (ICPET), 2018<br>Crossref | 32 words — < 1% |
| 23 | ijs.uobaghdad.edu.iq<br>Internet | 32 words — < 1% |
| 24 | lib.buet.ac.bd:8080<br>Internet | 32 words — < 1% |
| 25 | G R Usha, L. Dharmanna. "Sentiment Analysis on Business Data using Machine Learning", 2021 Second International Conference on Smart Technologies in Computing, Electrical and Electronics (ICSTCEE), 2021<br>Crossref | 31 words — < 1% |
| 26 | webthesis.biblio.polito.it<br>Internet | 29 words — < 1% |

| | | |
|---|---|---|
| 27 | www.dialog-21.ru<br>Internet | 29 words — < 1% |
| 28 | aryum Bibi, Wajid Arshad Abbasi, Wajid Aziz, Sundus Khalil, Mueen Uddin, Celestine Iwendi, Thippa Reddy Gadekallu. "A Novel Unsupervised Ensemble Framework using Concept-based Linguistic Methods and Machine Learning for Twitter Sentiment Analysis", Pattern Recognition Letters, 2022<br>Crossref | 28 words — < 1% |
| 29 | eprints.nottingham.ac.uk<br>Internet | 27 words — < 1% |
| 30 | Manuel B. Garcia, Armi Cunanan-Yabut. "Public Sentiment and Emotion Analyses of Twitter Data on the 2022 Russian Invasion of Ukraine", 2022 9th International Conference on Information Technology, Computer, and Electrical Engineering (ICITACEE), 2022<br>Crossref | 25 words — < 1% |
| 31 | medium.com<br>Internet | 25 words — < 1% |
| 32 | www.scpe.org<br>Internet | 24 words — < 1% |
| 33 | Jasper van der Waa, Tjeerd Schoonderwoerd, Jurriaan van Diggelen, Mark Neerincx. "Interpretable confidence measures for decision support systems", International Journal of Human-Computer Studies, 2020<br>Crossref | 22 words — < 1% |
| 34 | www.icicel.org<br>Internet | 22 words — < 1% |

| | | |
|---|---|---|
| 35 | ns2.thinkmind.org<br>Internet | 21 words — < 1% |
| 36 | www.mdpi.com<br>Internet | 21 words — < 1% |
| 37 | cicero.ensea.fr<br>Internet | 20 words — < 1% |
| 38 | Ernesto Lee, Furqan Rustam, Imran Ashraf, Patrick Bernard Washington, Manideep Narra, Rahman Shafique. "Inquest of Current Situation in Afghanistan Under Taliban Rule Using Sentiment Analysis and Volume Analysis", IEEE Access, 2022<br>Crossref | 19 words — < 1% |
| 39 | Uddagiri Sirisha, Bolem Sai Chandana. "Aspect based Sentiment & Emotion Analysis with ROBERTa, LSTM", International Journal of Advanced Computer Science and Applications, 2022<br>Crossref | 19 words — < 1% |
| 40 | library.samdu.uz<br>Internet | 18 words — < 1% |
| 41 | www.jmir.org<br>Internet | 18 words — < 1% |
| 42 | digitalcommons.njit.edu<br>Internet | 17 words — < 1% |
| 43 | link.springer.com<br>Internet | 17 words — < 1% |
| 44 | "Machine Learning for Healthcare Applications", Wiley, 2021<br>Crossref | 16 words — < 1% |

| | | |
|---|---|---|
| 45 | core.ac.uk<br>Internet | 15 words — < 1% |
| 46 | "Congress on Intelligent Systems", Springer Science and Business Media LLC, 2022<br>Crossref | 14 words — < 1% |
| 47 | "Proceedings of Third International Conference on Intelligent Computing, Information and Control Systems", Springer Science and Business Media LLC, 2022<br>Crossref | 14 words — < 1% |
| 48 | Basri Ciftci, Mehmet Serkan Apaydin. "A Deep Learning Approach to Sentiment Analysis in Turkish", 2018 International Conference on Artificial Intelligence and Data Processing (IDAP), 2018<br>Crossref | 14 words — < 1% |
| 49 | Kazi Nabiul Alam, Md Shakib Khan, Abdur Rab Dhruba, Mohammad Monirujjaman Khan, Jehad F. Al-Amri, Mehedi Masud, Majdi Rawashdeh. "Deep Learning-Based Sentiment Analysis of COVID-19 Vaccination Responses from Twitter Data", Computational and Mathematical Methods in Medicine, 2021<br>Crossref | 14 words — < 1% |
| 50 | jnu.ac.bd<br>Internet | 14 words — < 1% |
| 51 | open.uct.ac.za<br>Internet | 14 words — < 1% |
| 52 | www.frontiersin.org<br>Internet | 14 words — < 1% |
| 53 | www.semanticscholar.org | |

Internet

14 words — < 1%

54 Christopher P. Furner, Tom E. Yoon, Robert Zinko, Samuel H. Goh. "The Influence of Reviewer and Consumer Congruence in Online Word-of-Mouth Transactions", Journal of Electronic Commerce in Organizations, 2021
Crossref

13 words — < 1%

55 Nitin Madnani. "Getting started on natural language processing with Python", Crossroads, 09/01/2007
Crossref

13 words — < 1%

56 Wang, Zhenqian. "Attack Resilient Pulse Based Synchronization.", Clemson University, 2020
ProQuest

13 words — < 1%

57 dspace.plymouth.ac.uk
Internet

13 words — < 1%

58 www.cse.iitb.ac.in
Internet

13 words — < 1%

59 www.iitp.ac.in
Internet

13 words — < 1%

60 Ahmad, Tariq. "Classification of Tweets Using Multiple Thresholds with Self-Correction and Weighted Conditional Probabilities", The University of Manchester (United Kingdom), 2021
ProQuest

12 words — < 1%

61 Perin, F.. "Linguistic style checking with program checking tools", Computer Languages, Systems & Structures, 201204
Crossref

12 words — < 1%

| | | |
|---|---|---|
| 62 | repositori.uji.es<br>Internet | 12 words — < 1% |
| 63 | web.engr.uky.edu<br>Internet | 12 words — < 1% |
| 64 | "Advanced Prognostic Predictive Modelling in Healthcare Data Analytics", Springer Science and Business Media LLC, 2021<br>Crossref | 11 words — < 1% |
| 65 | Prachi Juyal. "Multi-modal Sentiment Analysis of Audio and Visual Context of the Data using Machine Learning", 2022 3rd International Conference on Smart Electronics and Communication (ICOSEC), 2022<br>Crossref | 11 words — < 1% |
| 66 | Suvarna G Kanakaraddi, Ashok K Chikaraddi, Karuna C. Gull, P S Hiremath. "Comparison Study of Sentiment Analysis of Tweets using Various Machine Learning Algorithms", 2020 International Conference on Inventive Computation Technologies (ICICT), 2020<br>Crossref | 11 words — < 1% |
| 67 | github.com<br>Internet | 11 words — < 1% |
| 68 | scholar.ufs.ac.za<br>Internet | 11 words — < 1% |
| 69 | vsip.info<br>Internet | 11 words — < 1% |
| 70 | docplayer.net<br>Internet | 10 words — < 1% |
| 71 | kops.uni-konstanz.de | |

| | | |
|---|---|---|
| | Internet | 10 words — < 1% |
| 72 | aclanthology.org<br>Internet | 9 words — < 1% |
| 73 | ousar.lib.okayama-u.ac.jp<br>Internet | 9 words — < 1% |
| 74 | www.hindawi.com<br>Internet | 9 words — < 1% |
| 75 | Communications in Computer and Information Science, 2015.<br>Crossref | 8 words — < 1% |
| 76 | Xue-Jing Liu. "Sentiments and Emotions for Vaccination in 2021: An International Comparison Study", Cold Spring Harbor Laboratory, 2022<br>Crossref Posted Content | 8 words — < 1% |
| 77 | go.gale.com<br>Internet | 8 words — < 1% |
| 78 | scholar.sun.ac.za<br>Internet | 8 words — < 1% |
| 79 | www.coursehero.com<br>Internet | 8 words — < 1% |
| 80 | www.gecekitapligi.com<br>Internet | 8 words — < 1% |
| 81 | Youness Arjoune, Naima Kaabouch. "A Comprehensive Survey on Spectrum Sensing in Cognitive Radio Networks: Recent Advances, New Challenges, and Future Research Directions", Sensors, 2019<br>Crossref | 7 words — < 1% |

82 "Machine Learning Algorithms and Applications", Wiley, 2021 — 6 words — < 1%
Crossref

83 Bagga, Sunyam. "Detecting Abuse on the Internet: It's Subtle.", McGill University (Canada), 2020 — 6 words — < 1%
ProQuest

84 Miftahul Qorib, Timothy Oladunni, Max Denis, Esther Ososanya, Paul Cotae. "Covid-19 Vaccine Hesitancy: Text Mining, Sentiment Analysis and Machine Learning on COVID-19 Vaccination Twitter Dataset", Expert Systems with Applications, 2022 — 6 words — < 1%
Crossref

| | | | |
|---|---|---|---|
| EXCLUDE QUOTES | OFF | EXCLUDE SOURCES | OFF |
| EXCLUDE BIBLIOGRAPHY | OFF | EXCLUDE MATCHES | OFF |