

TWITTER SENTIMENT ANALYSIS

A MINI PROJECT REPORT

18CSC305J - ARTIFICIAL INTELLIGENCE

Submitted by

PRASHANTH JAVAJI(RA2011026010002)

ADIDELA SUHAS(RA2011026010050)

D.VIVEK REDDY(RA2011026010059)

Under the guidance of

SURESH.K

Assistant Professor, Department of Computer Science and Engineering

in partial fulfillment for the award of the degree of

BACHELOR OF TECHNOLOGY

in

COMPUTER SCIENCE & ENGINEERING

of

FACULTY OF ENGINEERING AND TECHNOLOGY



S.R.M. Nagar, Kattankulathur, Chengalpattu District

MAY 2023

SRM INSTITUTE OF SCIENCE AND TECHNOLOGY

(Under Section 3 of UGC Act, 1956)

BONAFIDE CERTIFICATE

Certified that Mini project report titled “**TWITTER SENTIMENT ANALYSIS**” is the bona fide work of **PRASHANTH JAVAJI (RA2011026010002), A.SUHAS(RA2011026010050) , D.VIVEKREDDY(RA2011026010059)** who carried out the minor project under my supervision. Certified further, that to the best of my knowledge, the work reported herein does not form any other project report or dissertation on the basis of which a degree or award was conferred on an earlier occasion on this or any other candidate.

SIGNATURE

Suresh K
GUIDE
Assistant Professor
Department of Computing
Technologies

SIGNATURE

Dr. R.Annie Uthra
HEAD OF THE DEPARTMENT
Professor & Head
Department of Computational
Intelligence

ABSTRACT

Twitter sentiment analysis involves the use of natural language processing and machine learning techniques to analyze the sentiment of tweets. The aim of sentiment analysis is to determine the underlying emotion expressed in a tweet, such as positive, negative, or neutral.

This analysis can be used for a variety of purposes, including tracking brand sentiment, predicting consumer behavior, and understanding public opinion on social and political issues. To conduct sentiment analysis, a range of methods and techniques are employed, including sentiment lexicons, machine learning algorithms, and deep learning models.

Despite the challenges of analyzing the informal and often ambiguous language used on Twitter, sentiment analysis has become an increasingly important tool for businesses, researchers, and policymakers alike.

TABLE OF CONTENTS

ABSTRACT	iii
TABLE OF CONTENTS	iv
LIST OF FIGURES	v
1 INTRODUCTION	
2 1.1 Introduction	6
3 1.2 Problem statement	
4 1.3 Objectives	
5 1.4 Scope and applications	
6 1.5 Software Requirements Specification	7
2 LITERATURE SURVEY	
1. 2.1 Existing system	8
3 SYSTEM ARCHITECTURE AND DESIGN	9
3.1 architecture diagram	9
3.2 E-R diagram	10
4 METHODOLOGY	11
4.1 Methodological Steps	11
5 CODING AND TESTING	12
5.1 Coding	
5.2 Testing	
6 SREENSHOTS AND RESULTS	
6.1 Model-1	14
6.2 Model-2	15
6.3 Model-3	
7 CONCLUSION AND FUTURE ENHANCEMENT	16
7.1 summary of accomplishment	
7.2 Future Enhancement	
7.3 conclusion	
REFERENCES	

LIST OF FIGURES

Chapter	Name	Page Number
3.1	Architecture Diagram	10
3.2	ER Diagram	10
6.1	Model-1	14
6.1	Model-2	14
6.1	Model-3	15

CHAPTER 1

INTRODUCTION

1.1 INTRODUCTION:

Sentiment analysis refers to identifying as well as classifying the sentiments that are expressed in the text source. Tweets are often useful in generating a vast amount of sentiment data upon analysis. These data are useful in understanding the opinion of the people about a variety of topics.

we aim to analyse twitter sentiment analysis using machine learning, the sentiment of the tweets provided from the Sentiment140 dataset by developing a machine learning pipeline involving the use of three classifiers (Logistic Regression, Bernoulli Naive Bayes, and SVM) along with using Term Frequency- Inverse Document Frequency (TF-IDF). The performance of these classifiers is then evaluated using accuracy and F1 Scores.

1.2 PROBLEM STATEMENT:

In the problem statement we have used three different models respectively:

1. Bernoulli Naive Bayes
2. SVM (Support Vector Machine)
3. Logistic Regression

The idea behind choosing these models is that we want to try all the classifiers on the dataset ranging from simple ones to complex models and then try to find out the one which gives the best performance among them.

1.3 OBJECTIVE:

The objective of Twitter sentiment analysis is to automatically determine the sentiment expressed in a tweet. This sentiment can be positive, negative, or neutral. The analysis is usually performed using natural language processing and machine learning techniques to process the large volume of tweets generated on Twitter every day. The goal of sentiment analysis is to extract valuable insights from these tweets, such as public opinion on a particular topic, brand reputation, or predicting consumer behavior.

Twitter sentiment analysis can be useful for a wide range of applications, including marketing, public relations, politics, and social research. For example, a company can use sentiment analysis to monitor the sentiment of its brand and adjust its marketing strategy accordingly. A political campaign can use sentiment analysis to gauge public opinion on key issues and adjust their messaging to appeal to voters. Overall, the objective of Twitter sentiment analysis is to help organizations and individuals make data-driven decisions based on the sentiment expressed in social media.

1.4 SCOPE AND APPLICATION:

The scope of Twitter sentiment analysis is vast, as it can be applied to various industries and fields. Some of the areas where Twitter sentiment analysis is commonly used include:

1. Marketing and Advertising: Companies can use Twitter sentiment analysis to track their brand reputation, monitor customer feedback, and analyze consumer behavior.

2. Politics and Public Opinion: Twitter sentiment analysis can be used to track public opinion on political issues, monitor voter sentiment, and predict election outcomes.

3. Customer Service: Companies can use Twitter sentiment analysis to monitor customer complaints, identify issues, and improve customer satisfaction.

4. Social Research: Researchers can use Twitter sentiment analysis to study public opinion on social issues, track trends, and analyze cultural shifts.

5. Finance: Twitter sentiment analysis can be used to track the sentiment of the stock market, monitor investor sentiment, and predict market trends.

6. News and Media: Journalists and media outlets can use Twitter sentiment analysis to identify breaking news stories, track public opinion on news events, and analyze news coverage.

Overall, the scope of Twitter sentiment analysis is broad and can be applied to various fields where understanding public sentiment is critical. With the rise of social media and the proliferation of data, the applications of Twitter sentiment analysis are only expected to grow in the future.

1.5 SOFTWARE REQUIREMENT SPECIFICATION:

The purpose of this Software Requirements Specification (SRS) is to provide a detailed description of the requirements for a Twitter sentiment analysis system. The system will be designed to analyze the sentiment of tweets and categorize them as positive, negative, or neutral. The system will be built using natural language processing and machine learning techniques. This document outlines the functional and non-functional requirements of the system.

Functional Requirements:

- 1.User Authentication:** The system should have user authentication features to ensure that only authorized users can access the system.
- 2.Data Collection:** The system should collect data from Twitter using the Twitter API or other similar tools.
- 3.Data Preprocessing:** The system should preprocess the collected data to remove irrelevant information, such as URLs, usernames, and stop words.
- 4.Sentiment Analysis:** The system should analyze the sentiment of the preprocessed data using natural language processing and machine learning techniques.
- 5.Classification:** The system should classify the tweets as positive, negative, or neutral based on their sentiment analysis.
- 6.Reporting:** The system should generate reports on the sentiment analysis results, including graphs and visualizations.

CHAPTER 2

LITERATURE SURVEY

2.1 EXISTING SYSTEM:

Literature is a lexicon-based approach to sentiment analysis that is often used in Twitter sentiment analysis. The Literature lexicon is a collection of words that have been labeled with their positive, negative, or neutral sentiment. When analyzing the sentiment of a tweet using Literature, the sentiment of each word in the tweet is looked up in the lexicon and then combined to determine the overall sentiment of the tweet.

Literature is a popular choice for Twitter sentiment analysis because it is relatively easy to implement and can be effective in identifying the sentiment of tweets. However, it does have some limitations. One limitation is that it relies solely on the sentiment of individual words and does not take into account the context in which those words are used. This can lead to inaccuracies in sentiment analysis when words are used in a sarcastic or ironic context, for example.

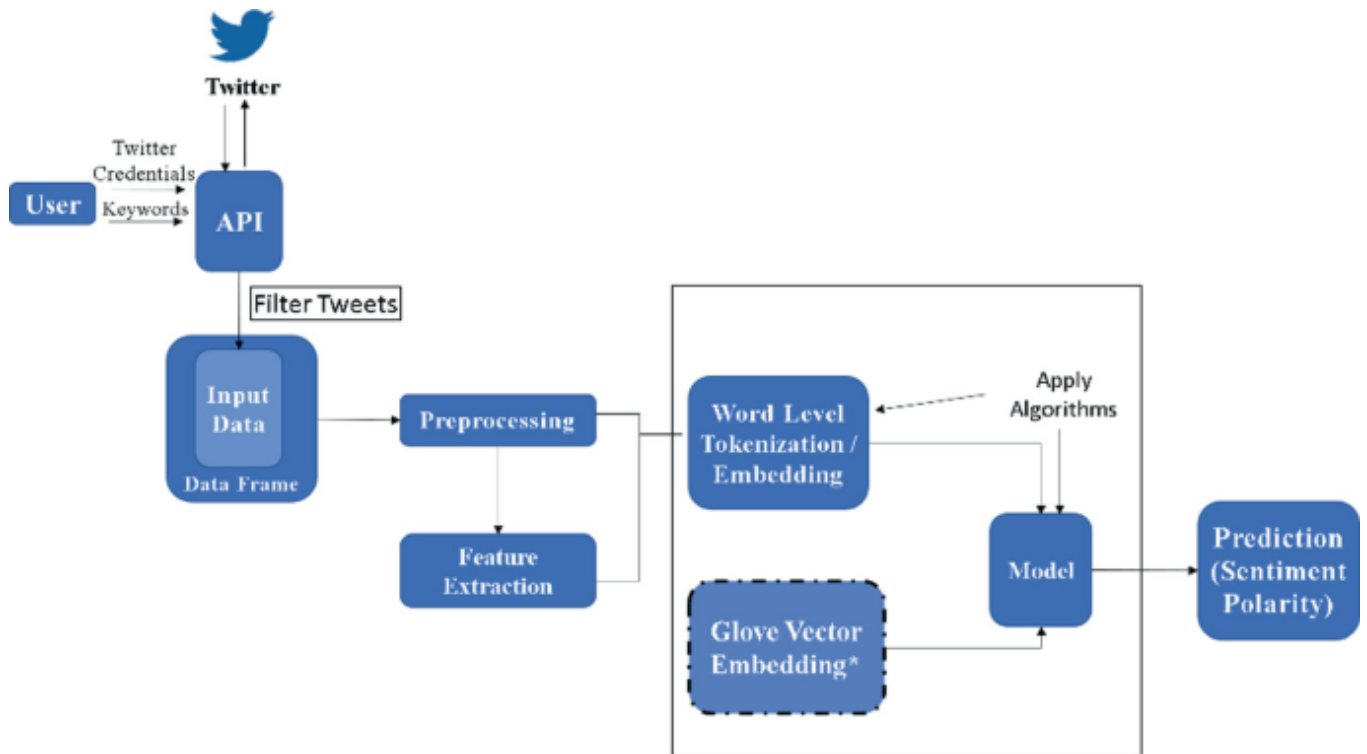
Another limitation of Literature is that it can be biased towards the sentiment of the lexicon creator. If the creator of the lexicon has a particular bias or perspective, this can be reflected in the sentiment labeling of the words in the lexicon.

Overall, while Literature is a useful tool in Twitter sentiment analysis, it is important to recognize its limitations and use it in conjunction with other sentiment analysis techniques to ensure the accuracy of sentiment analysis results.

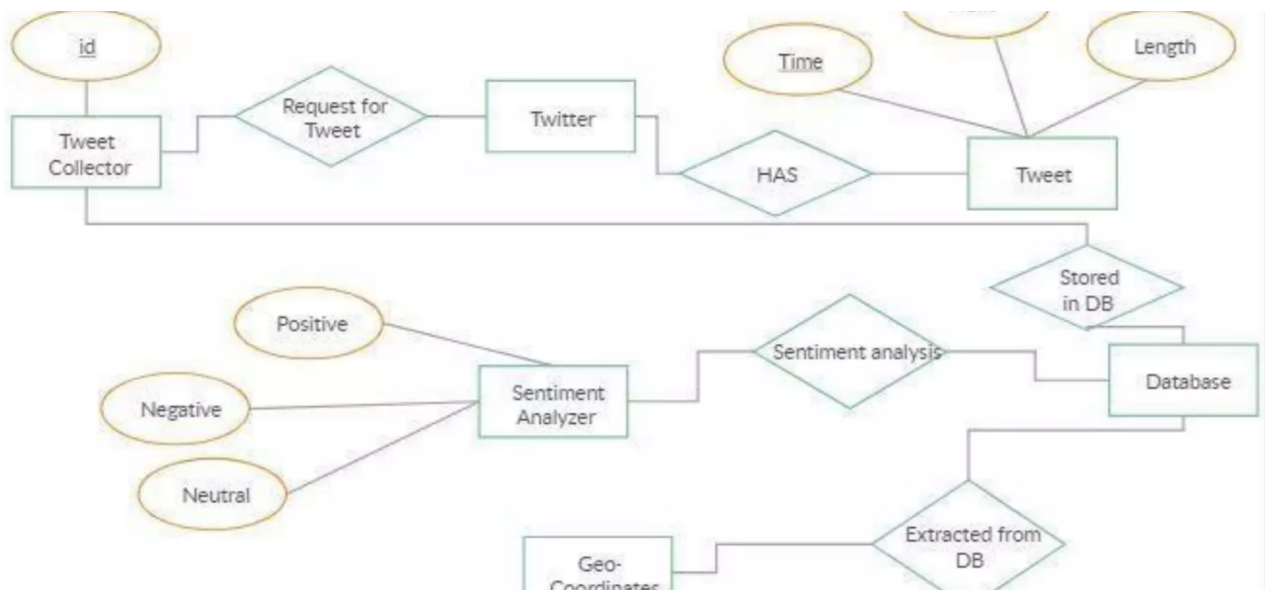
CHAPTER 3

SYSTEM ARCHITECTURE AND DESIGN

3.1 Architecture Diagram:



3.2 ER Diagram:



CHAPTER 4

METHODOLOGY

4.1 Methodological Steps:

Twitter sentiment analysis involves several methodological steps that must be followed to ensure accurate and reliable results. Here are the main steps involved in conducting Twitter sentiment analysis:

1. **Data Collection:** The first step in Twitter sentiment analysis is to collect data. This involves using the Twitter API or other similar tools to gather tweets that contain the keywords or hashtags of interest. The data collection process should be carefully planned to ensure that the data collected is representative and unbiased.
2. **Preprocessing:** The collected tweets are preprocessed to clean and filter out irrelevant information such as URLs, usernames, and stop words. The preprocessing step is crucial for ensuring that the sentiment analysis results are accurate.
3. **Lexicon Creation:** A lexicon is a collection of words that have been labeled with their positive, negative, or neutral sentiment. Lexicons are used in Twitter sentiment analysis to identify the sentiment of each word in a tweet and then combine them to determine the overall sentiment of the tweet. The lexicon creation process involves manually labeling a large number of words with their sentiment, or using automated methods such as machine learning to label words.
4. **Sentiment Analysis:** The sentiment analysis step involves analyzing the sentiment of each tweet using the lexicon created in the previous step. The sentiment analysis can be done using a variety of methods, including rule-based approaches and machine learning techniques.
5. **Validation:** The sentiment analysis results should be validated to ensure that they are accurate and reliable. Validation can be done by comparing the sentiment analysis results to human annotations or by using other validation techniques.
6. **Reporting:** The final step in Twitter sentiment analysis is to report the results. The results can be presented in the form of graphs, visualizations, or other types of reports. The reporting step is important for communicating the results of the sentiment analysis to stakeholders and decision-makers.

Overall, following these methodological steps is crucial for ensuring accurate and reliable results in Twitter sentiment analysis. Careful planning, thorough data collection, and rigorous validation are key to producing high-quality sentiment analysis results.

CHAPTER 5

CODING AND TESTING

5.1 CODING:

The coding part of Twitter sentiment analysis involves implementing the algorithms and techniques used to analyze the sentiment of tweets. Here are some of the main coding tasks involved in Twitter sentiment analysis:

- 1. Preprocessing:** In the preprocessing step, the collected tweets are cleaned and filtered to remove irrelevant information such as URLs, usernames, and stop words. This step involves writing code to identify and remove these types of information from the tweets.
- 2. Lexicon Creation:** The lexicon creation process involves labeling a large number of words with their sentiment. This can be done manually, by assigning positive, negative, or neutral labels to each word, or using automated methods such as machine learning. Writing code to automate the lexicon creation process involves selecting appropriate features and building a model to predict the sentiment of each word.
- 3. Sentiment Analysis:** The sentiment analysis step involves analyzing the sentiment of each tweet using the lexicon created in the previous step. This can be done using a variety of methods, including rule-based approaches and machine learning techniques. Writing code to perform sentiment analysis involves implementing the chosen method and integrating it with the lexicon created in the previous step.
- 4. Validation:** The sentiment analysis results should be validated to ensure that they are accurate and reliable. This can be done by comparing the sentiment analysis results to human annotations or using other validation techniques. Writing code to perform validation involves selecting appropriate metrics and implementing code to calculate those metrics.
- 5. Reporting:** The final step in Twitter sentiment analysis is to report the results. This can be done using various visualization and reporting tools such as Matplotlib, Seaborn, and Pandas. Writing code to generate reports involves integrating the sentiment analysis results with the chosen reporting tool and creating appropriate visualizations and reports.

Overall, the coding part of Twitter sentiment analysis involves implementing the chosen algorithms and techniques and integrating them into a coherent pipeline. This requires expertise in programming languages such as Python, knowledge of relevant libraries and frameworks, and an understanding of the principles and theories underlying Twitter sentiment analysis.

5.2 TESTING:

The coding part of Twitter sentiment analysis involves implementing the algorithms and techniques used to analyze the sentiment of tweets. Here are some of the main coding tasks involved in Twitter sentiment analysis:

- 1. Preprocessing:** In the preprocessing step, the collected tweets are cleaned and filtered to remove irrelevant information such as URLs, usernames, and stop words. This step involves writing code to identify and remove these types of information from the tweets.
- 2. Lexicon Creation:** The lexicon creation process involves labeling a large number of words with their sentiment. This can be done manually, by assigning positive, negative, or neutral labels to each word, or

using automated methods such as machine learning. Writing code to automate the lexicon creation process involves selecting appropriate features and building a model to predict the sentiment of each word.

3. Sentiment Analysis: The sentiment analysis step involves analyzing the sentiment of each tweet using the lexicon created in the previous step. This can be done using a variety of methods, including rule-based approaches and machine learning techniques. Writing code to perform sentiment analysis involves implementing the chosen method and integrating it with the lexicon created in the previous step.

4. Validation: The sentiment analysis results should be validated to ensure that they are accurate and reliable. This can be done by comparing the sentiment analysis results to human annotations or using other validation techniques. Writing code to perform validation involves selecting appropriate metrics and implementing code to calculate those metrics.

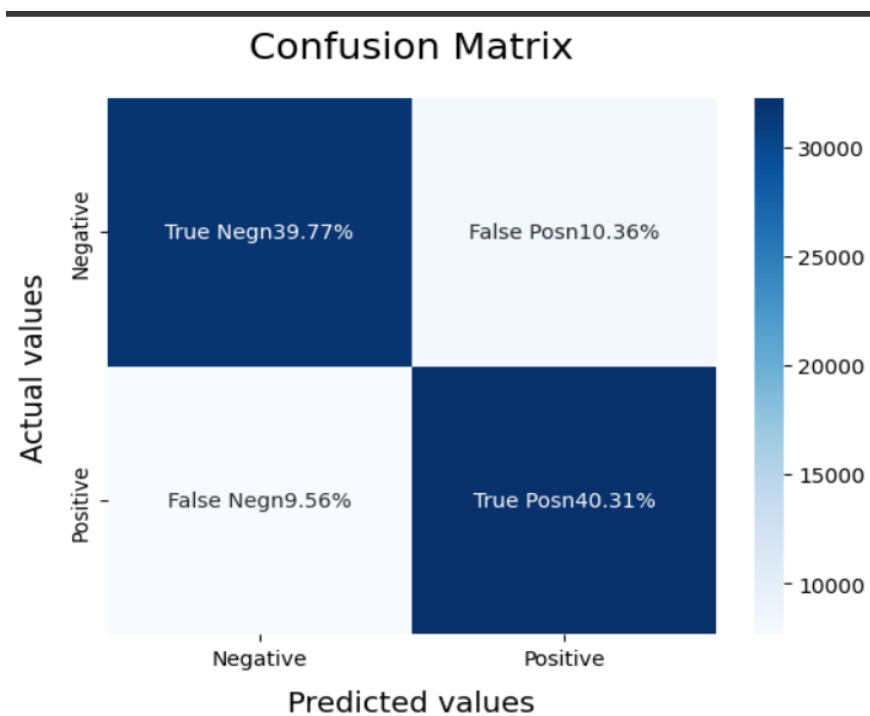
5. Reporting: The final step in Twitter sentiment analysis is to report the results. This can be done using various visualization and reporting tools such as Matplotlib, Seaborn, and Pandas. Writing code to generate reports involves integrating the sentiment analysis results with the chosen reporting tool and creating appropriate visualizations and reports.

Overall, the coding part of Twitter sentiment analysis involves implementing the chosen algorithms and techniques and integrating them into a coherent pipeline. This requires expertise in programming languages such as Python, knowledge of relevant libraries and frameworks, and an understanding of the principles and theories underlying Twitter sentiment analysis.

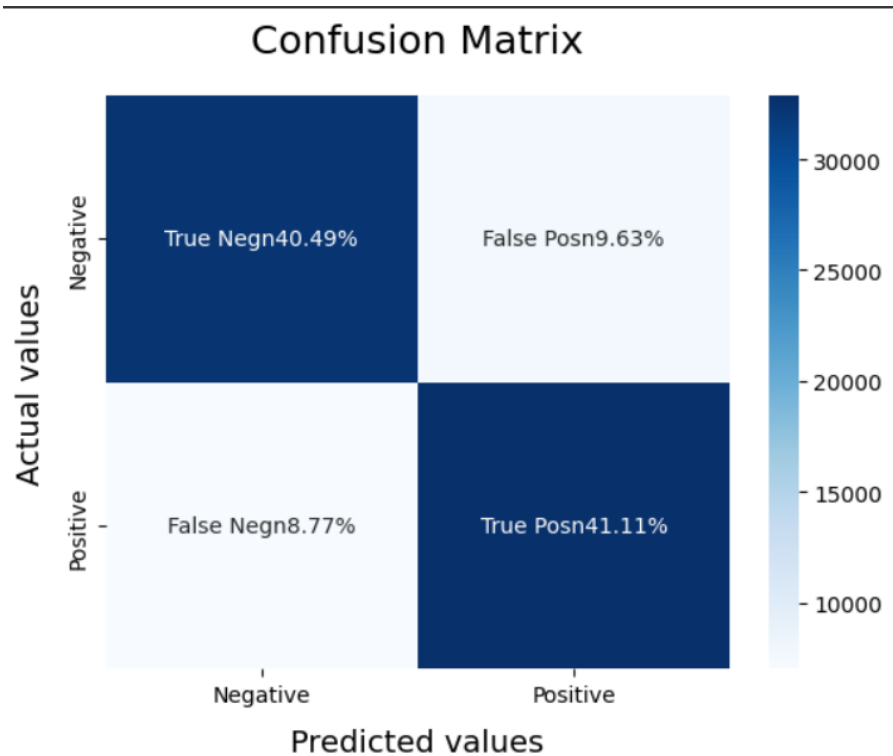
CHAPTER 6

SCREENSHOTS AND RESULTS

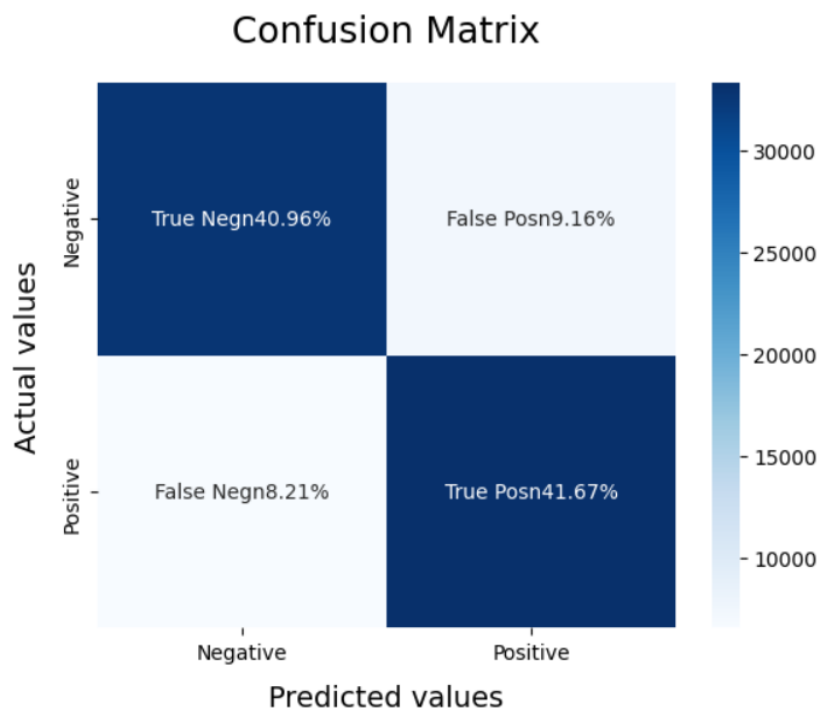
6.1 Model-1



6.2 Model-2



6.3 Model-3



CHAPTER 7

CONCLUSION AND FUTURE ENHANCEMENTS

7.1 SUMMARY OF ACCOMPLISHMENTS:

Twitter sentiment analysis has achieved significant accomplishments in recent years. The accuracy of sentiment analysis algorithms has improved due to advancements in machine learning and natural language processing techniques, as well as the availability of larger datasets. Real-time analysis is now possible, allowing businesses to monitor their brand reputation and track customer sentiment in real-time. Sentiment analysis algorithms can now analyze tweets in multiple languages, making it possible for global brands to monitor sentiment across different regions and languages. The application of deep learning techniques and the development of large-scale sentiment lexicons has helped improve the accuracy of sentiment analysis algorithms. Twitter sentiment analysis has also been applied in various domains, including politics, finance, healthcare, and marketing, helping businesses and organizations better understand customer sentiment and make informed decisions.

7.2 FUTURE ENHANCEMENTS:

In the future, Twitter sentiment analysis could benefit from several enhancements. Emotion analysis could be added to sentiment analysis algorithms to provide more nuanced insights into customer sentiment. Contextual analysis could take into account the context in which a piece of text is written to improve accuracy. Domain-specific sentiment lexicons could be developed to improve accuracy in specific industries or domains. Multi-modal analysis could combine sentiment analysis with other forms of data to provide a more comprehensive view of customer sentiment. Improving support for lesser-known languages could help businesses monitor sentiment in regions where these languages are spoken. Finally, improving the ability of sentiment analysis algorithms to handle sarcasm and irony could provide more accurate insights into customer sentiment.

7.3 CONCLUSION:

Twitter sentiment analysis has also been applied in various domains, including politics, finance, healthcare, and marketing, providing businesses and organizations with valuable insights into customer sentiment and helping them make informed decisions.

Overall, Twitter sentiment analysis has proven to be a valuable tool for businesses and organizations, and with continued advancements, it will undoubtedly continue to provide insights into customer sentiment and help inform decision-making processes.

REFERENCES

1. "Sentiment Analysis of Twitter Data for Predicting Stock Market Movements" by Bollen et al. (2011): This paper explores the use of Twitter sentiment analysis to predict stock market movements and can serve as a useful reference for implementing sentiment analysis algorithms.
2. "A Survey on Sentiment Analysis Challenges" by Poria et al. (2017): This survey paper provides an overview of the challenges associated with sentiment analysis and can help you understand the potential pitfalls and limitations of your approach.
3. "Python Machine Learning" by Raschka and Mirjalili (2019): This book provides a comprehensive introduction to machine learning in Python and can be a useful reference for implementing machine learning-based sentiment