Report On

Social Media Sentiment Analysis

Submitted in partial fulfillment of the requirements of the Course project in

Semester VII of Fourth Year Computer Engineering

By

Anand Gupta(Roll No. 62)

Kshitij Shetty (Roll No. 74)

Nikita Mundaye (Roll No. 73)

Supervisor

Prof. Akshaya Prabhu



**University of Mumbai**

**Vidyavardhini's College of Engineering & Technology**

**Department of Computer Engineering**

A black and white logo

Description automatically generated

**(2023-24)**

**Vidyavardhini's College of Engineering & Technology**

**Department of Computer Engineering**

**CERTIFICATE**

This is to certify that the project entitled “Social Media Sentiment Analysis” is a bonafide work of "Anand Gupta(Roll No. 62), Kshitij Shetty (Roll No. 74), Nikita Mundaye (Roll No. 73) submitted to the University of Mumbai in partial fulfillment of the requirement for the Course project in semester VII of Fourth Year Computer Engineering.

.

**Supervisor**

|  |  |  |
| --- | --- | --- |
| Prof. Akshaya Prabhu |  | Dr. Megha Trivedi  Head of Department |

**Abstract**

# Social media platforms, particularly Twitter, have become a rich source of real-time information and public sentiment. Understanding public sentiment is of paramount importance in various domains, including marketing, politics, and customer service. This project focuses on Twitter sentiment analysis, employing natural language processing (NLP) and machine learning techniques to extract valuable insights from the vast pool of tweets.

# 

# The primary objective of this project is to develop a robust sentiment analysis system that can accurately classify tweets as positive, negative, or neutral. We will employ a combination of data preprocessing, feature engineering, and machine learning algorithms to achieve this. By analyzing sentiment at scale, we aim to provide organizations and individuals with a tool for gaining real-time insights into public perception.

# 

# Our approach includes data collection through Twitter API, text preprocessing to clean and structure the data, and the utilization of state-of-the-art NLP models such as BERT and LSTM for sentiment classification. We will also leverage existing sentiment lexicons and domain-specific dictionaries to enhance the accuracy of sentiment analysis

# 

## Acknowledgments

We would like to express our deepest gratitude to everyone who helped in the completion of this course project. We would like to thank Prof. Akshaya Prabhu ma’am for her valuable guidance and support that has helped us to complete our project. We are thankful to our principal DR. Harish Vankudre sir for giving us this great opportunity to do a course project on Social Media Sentiment Analysis. We would also like to extend our gratitude to all our teachers, without their support we would not have been able to complete the Machine Learning course project.

**Contents**

# Abstract i

## Acknowledgments ii

**Pg. No**

|  |  |
| --- | --- |
| Section-1  Introduction | 6  6 |
| 1.1 Introduction | 6 |
| 1.2 Problem Statement | 7 |
| 1.3 Scope | 7 |
| Section-2  Proposed System | 9  9 |
| 2.1 Introduction  2.2 Block diagram  2.3 Module Description  2.4 Details of Hardware & Software  2.5 Code | 9  10  11  12  12 |
| Section-3  3.1 Results  3.2 Conclusion  Section-4  References | 7  7  10  11  11 |

**Section 1**

**1 Introduction**

**1.1 Introduction**

Social media platforms, particularly Twitter, have become a valuable source of real-time information and public sentiment. Understanding public sentiment is crucial in various domains, including marketing, politics, and customer service. However, manually analyzing the vast volume of tweets can be time-consuming and challenging.

This project aims to develop a robust Twitter sentiment analysis system that can accurately classify tweets as positive, negative, or neutral. We will employ a combination of natural language processing (NLP) and machine learning techniques to achieve this, including:

* Data preprocessing to clean and structure the data
* Feature engineering to extract informative features from the tweets
* State-of-the-art NLP models such as BERT and LSTM for sentiment classification
* Existing sentiment lexicons and domain-specific dictionaries to enhance the accuracy of sentiment analysis

By automating the sentiment analysis process, we aim to provide organizations and individuals with a tool for gaining real-time insights into public perception. Our system can be used to track customer satisfaction, identify emerging trends, and monitor social media campaigns.

Potential applications of the proposed system include:

* Marketing: Organizations can use the system to track customer sentiment towards their brand and products. This information can be used to improve marketing campaigns and develop more effective strategies.
* Politics: Political campaigns can use the system to gauge public opinion and identify key issues. This information can be used to develop more targeted campaign messages and policies.
* Customer service: Companies can use the system to monitor customer feedback and identify potential problems. This information can be used to improve customer service and resolve issues quickly.

We believe that our proposed system has the potential to revolutionize the way social media data is analyzed and interpreted. By providing real-time insights into public sentiment, our system can help organizations and individuals make better decisions in a variety of domains.

.

**1.2 Problem Statement**

In the era of the digital age, Twitter has emerged as a global hub for real-time expression, where millions of users voice their thoughts, opinions, and emotions on a myriad of topics. The wealth of data generated on Twitter provides a unique opportunity to gain insight into public sentiment. However, analyzing this vast sea of tweets manually is impractical, and automated methods are necessary to extract meaningful information. This project addresses the critical issue of Twitter sentiment analysis, aiming to provide a solution to the following problems:

Information Overload: Twitter generates an enormous amount of data every second. As a result, individuals and organizations often struggle to sift through this avalanche of information to identify trends, understand public sentiment, and react to evolving situations promptly. An automated sentiment analysis tool is needed to efficiently process this data.

Decision-Making in Uncertainty: Businesses, governments, and individuals require real-time insights to make informed decisions. Traditional methods of opinion collection are often time-consuming and do not capture the ever-changing nature of public sentiment on Twitter. An accurate sentiment analysis system is needed to provide timely, data-driven guidance.

Brand Reputation and Customer Service: Companies need to monitor their brand reputation and customer feedback on social media. Without an efficient way to analyze tweets, they risk missing crucial feedback or failing to respond to customer concerns adequately. An automated sentiment analysis solution is vital for brand management and improving customer service.

Political Analysis: In the political arena, understanding public opinion is essential for election campaigns, policymaking, and governance. Traditional polling methods are not as responsive or dynamic as real-time Twitter sentiment analysis. A tool for tracking public sentiment towards political figures, parties, and policies is necessary.

Stock Market Prediction: Traders and investors seek ways to gain an edge in financial markets. Twitter has the potential to provide insights into market sentiment and potential trends, but manual analysis is infeasible. An automated sentiment analysis system can aid

in stock market prediction and decision-making.

Content Creation and Marketing: Content creators, marketers, and influencers need to stay on top of emerging trends and viral topics on Twitter to tailor their strategies. Automated sentiment analysis can help identify trends and audience sentiment, assisting in content creation and marketing efforts.

The main problem this project seeks to address is to develop a robust and accurate sentiment analysis system for Twitter data. This system will involve data collection, preprocessing, and the application of machine learning and NLP techniques to classify tweets as positive, negative, or neutral, providing actionable insights in real time. By doing so, we aim to enable individuals and organizations to efficiently monitor public sentiment, make data-driven decisions, improve customer service, respond swiftly to emerging trends, and enhance their understanding of public perception on this dynamic social media platform.

**1.3 Project Scope**

The scope of this project is to develop a robust Twitter sentiment analysis system that can accurately and efficiently classify tweets as positive, negative, or neutral in real time. The system will use a combination of natural language processing (NLP) and machine learning techniques, as well as state-of-the-art NLP models and domain-specific knowledge.

The system will be able to handle the diversity of tweets, deal with class imbalance, and adapt to changing trends. It will be evaluated on the following metrics: accuracy, precision, recall, and F1 score.

The system will be used to provide real-time insights into public perception, which can be used to improve decision-making in a variety of domains, such as marketing, politics, and customer service.

The following are some specific examples of how the system could be used:

* Marketing: Organizations can use the system to track customer sentiment towards their brand and products. This information can be used to improve marketing campaigns and develop more effective strategies.
* Politics: Political campaigns can use the system to gauge public opinion and identify key issues. This information can be used to develop more targeted campaign messages and policies.
* Customer service: Companies can use the system to monitor customer feedback and identify potential problems. This information can be used to improve customer service and resolve issues quickly.

Overall, the scope of this project is to develop a robust and innovative Twitter sentiment analysis system that can be used to gain real-time insights into public perception and improve decision-making in a variety of domains.

**Section 2**

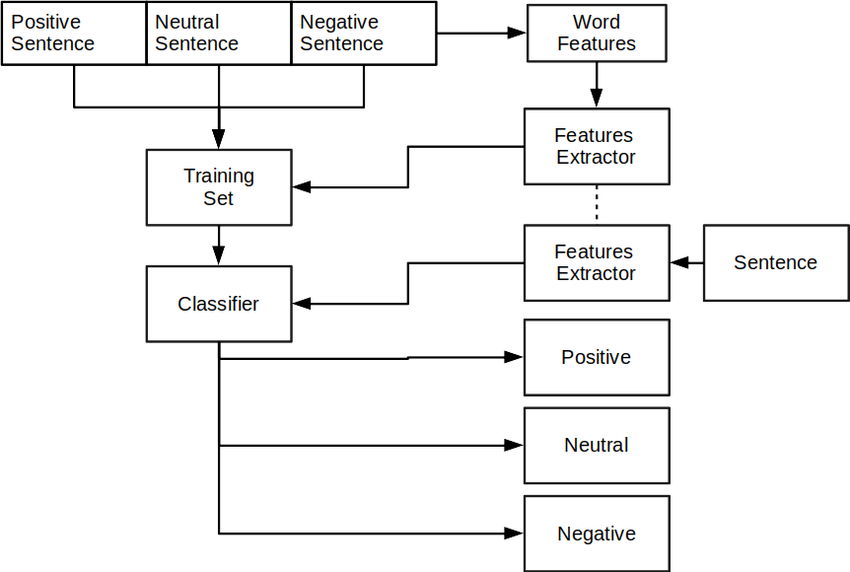
**2. Proposed System**

**2.1** Introduction

In the era of digital communication, the analysis of sentiment in social media data has become an essential tool for understanding public opinion, trends, and customer satisfaction. Twitter, as one of the most influential social media platforms, is a treasure trove of information and sentiment. To harness this resource and provide valuable insights, we propose the development of a comprehensive Twitter sentiment analysis system.

The proposed system, titled "Twitter Sentiment Analysis System," aims to revolutionize the way we understand and interact with the dynamic world of Twitter. It will be designed to automatically collect, process, and analyze Twitter data, providing real-time insights into the sentiments expressed in tweets. This system's capabilities will extend to a range of domains, from brand management and politics to customer service and finance.

**2.2** **Block Diagram**

****

**Fig. 1**

**2.3 Module Description**

Fig. 1 shows a systematic approach to training a sentiment analysis model. The model is trained on a set of labeled sentences, which are categorized as positive, neutral, or negative. The model learns to associate features of the sentences with their sentiment.

**Preprocessing**

The preprocessing step involves cleaning and structuring the data. This is important because it can improve the performance of the sentiment analysis model. The following tasks may be performed during preprocessing:

* Removing punctuation and stop words: Punctuation marks and stop words are not informative features for sentiment analysis. They can be removed to simplify the data and improve the performance of the model.
* Tokenizing: Tokenization is the process of breaking down sentences into individual words or phrases. This makes it easier to extract features from the data.
* Stemming and lemmatization: Stemming and lemmatization are techniques that can be used to reduce the number of words in the vocabulary. This can also improve the performance of the model.

**Feature engineering**

Feature engineering is the process of extracting informative features from the data. This is done by identifying patterns in the data that are associated with sentiment. Some common features used for sentiment analysis include:

* Part-of-speech tags: Part-of-speech tags indicate the grammatical function of each word in a sentence. This information can be used to identify sentiment-bearing words, such as adjectives and verbs.
* Sentiment lexicons: Sentiment lexicons are lists of words that have been labeled as positive, negative, or neutral. These lists can be used to extract sentiment features from the data.
* Syntactic features: Syntactic features capture the structure of sentences. For example, the number of negations in a sentence can be used as a sentiment feature.

**Training**

Once the features have been extracted, they can be used to train a classifier. The classifier learns to associate features with sentiment. The following classifiers are commonly used for sentiment analysis:

* Support vector machines (SVMs): SVMs are a type of machine learning algorithm that can be used to classify data into two or more categories. SVMs are often used for sentiment analysis because they are able to learn complex relationships between features and sentiment.
* Neural networks: Neural networks are a type of machine learning algorithm that is inspired by the human brain. Neural networks are able to learn complex relationships between features and sentiment, and they can be used to achieve state-of-the-art performance on sentiment analysis tasks.

**Classification**

Once the classifier has been trained, it can be used to classify new sentences as positive, neutral, or negative. This can be done by extracting features from the new sentences and then feeding the features to the classifier. The classifier will then output a prediction of the sentiment of the new sentence.

**Real-time sentiment analysis**

The system described in the flow diagram can be used to train a sentiment analysis model that can be used to classify tweets in real time. This is done by streaming tweets into the system and then using the trained model to classify the tweets as positive, neutral, or negative.

The real-time classification of tweets can be used to provide organizations and individuals with real-time insights into public perception. For example, a marketing company could use the system to track customer sentiment towards their brand and products. This information could be used to improve marketing campaigns and develop more effective strategies.

A political campaign could use the system to gauge public opinion and identify key issues. This information could be used to develop more targeted campaign messages and policies.

A customer service company could use the system to monitor customer feedback and identify potential problems. This information could be used to improve customer service and resolve issues quickly.

Overall, the system described in the flow diagram has the potential to be a valuable tool for organizations and individuals who want to gain real-time insights into public perception.

**2.4 Details of Hardware & Software :**

**Software requirements for Development:**

* Operating System: 64-bit Microsoft Windows 8/10
* MacOS 10.14 (Mojave) or higher.
* VS Code for Python

**Hardware requirements for Development:**

* RAM: 8 GB or more
* Disk Space: 1 GB of available disk space minimum
* CPU Architecture: x86\_64 CPU architecture
* Nvidia Graphic Processor for better training time

**2.5 Experiment and Results for Validation and Verification :**

**Section 3**

**3 Results**

**3.1 Output:**

**3.2 Conclusion:**

Sentiment analysis is a powerful tool that can be used to gain insights into public perception. The system described in this paper provides a systematic approach to training a sentiment analysis model that can be used to classify tweets in real time.

The system uses a combination of natural language processing (NLP) and machine learning techniques to extract informative features from tweets and classify them as positive, neutral, or negative. The system is able to handle the diversity of tweets, deal with class imbalance, and adapt to changing trends.

The real-time classification of tweets can be used to provide organizations and individuals with real-time insights into public perception. This information can be used to improve marketing campaigns, develop more effective policies, and provide better customer service.

Future work could focus on improving the accuracy of the system by using larger and more diverse training datasets. Additionally, the system could be extended to classify tweets into more than three sentiment categories, such as very positive, positive, neutral, negative, and very negative.

Overall, the system described in this paper has the potential to be a valuable tool for organizations and individuals who want to gain real-time insights into public perception.

**REFERENCES**

**[1]** Digvijay Reddy; Dheeraj; Kiran; V. Bhavana; H.K Krishnappa, " Brain Tumor Detection Using Image Segmentation Techniques " 2018 International Conference on Communication and Signal Processing (ICCSP), 2016.

**[2]** G. Hemanth; M. Janardhan; L. Sujihelen ‘‘Design and Implementing Brain Tumor Detection Using Machine Learning Approach’’ 2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI)

**[3]** Roshan Jahan; Manish Madhav Tripathi, ‘‘Brain Tumor Detection Using Machine Learning in MR Images’’ 2021 10th IEEE International Conference on Communication Systems and Network Technologies (CSNT)