**Mini Project Report on**



**TWITTER SENTIMENT ANALYSIS**



**Submitted in partial fulfillment of the requirement for the award of the degree of**

**BACHELOR OF TECHNOLOGY**

**IN**

**COMPUTER SCIENCE & ENGINEERING**

**Submitted by:**

**Student Name: Aashish Kuswaha** **University Roll No.2016554**

***Under the Mentorship of***

**Dr. Vishan Kumar Gupta**

**Assistant Professor**



**Department of Computer Science and Engineering**

**Graphic Era (Deemed to be University)**

**Dehradun, Uttarakhand**

**January 2023**



**CANDIDATE’S DECLARATION**

I hereby certify that the work which is being presented in the project report entitled **“Twitter Sentiment analysis”** in partial fulfillment of the requirements for the award of the Degree of Bachelor of Technology in Computer Science and Engineeringof the Graphic Era (Deemed to be University), Dehradun shall be carried out by the under the mentorship of Dr. Vishan kumar Gupta**, Assistant professor** , Department of Computer Science and Engineering, Graphic Era (Deemed to be University), Dehradun.

Name : Aashish Kushwaha MENTOR: Dr Vishan Kumar Gupta

University Roll no: 2016554 Designation: Assistant Professor

Signature: \_\_\_\_\_\_\_\_\_ Sign: \_\_\_\_\_\_\_\_\_\_\_\_\_

**Table of Contents**

|  |  |  |
| --- | --- | --- |
| **Chapter No.** | **Description** | **Page No.** |
| Chapter 1 | Introduction | **1-2** |
| Chapter 2 | Literature Survey | **3-5** |
| Chapter 3 | Methodology | **6-8** |
| Chapter 4 | Result and Discussion | **9-10** |
| Chapter 5 | Conclusion and Future Work | **11** |
|  | References | **12** |

**Chapter 1**

**Introduction**

Generating statistical information regarding emotions, sentiments out of analysis of user’s opinions from tweets, which can be used as an inference to understand how users feel thereby improving users experiences regarding.

* 1. **Introduction**

Sentiment analysis is the task of finding the opinions and affinity of people towards specific topics of interest. Be it a product or a movie, opinions of people matter, and it affects the decision-making process of people. The first thing a person does when he or she wants to buy a product on-line, is to see the kind of reviews and opinions that people have written. Social media such as Facebook, blogs, twitter have become a place where people post their opinions on certain topics. The sentiment of the tweets of a particular subject has multiple usage, including stock market analysis of a company, movie reviews, in psychology to analyze the mood of people that has a variety of applications, and so on.

Sentiments of tweets can be categorized into many categories like positive, negative and neutral. The data, being labeled by humans, has a lot of noise, and its hard to achieve good accuracy.

**1.2 Motivation**

Being extremely interested in everything having a relation with the Machine Learning, the independent project was a great occasion to give me the time to learn and confirm my interest for this field. The fact that we can make estimations, predictions and give the ability for machines to learn by themselves is both powerful and limitless in term of application possibilities. We can use Machine Learning in Finance, Medicine, almost everywhere. That’s why I decided to conduct my project around the Machine Learning.

**1.3 IDEA**

This project was motivated by my desire to investigate the sentiment analysis field of machine learning since it allows to approach natural language processing which is a very hot topic actually. Following my previous experience where it was about face recognition and Machine learning model. I applied the same idea with tweets and try to figure out which is positive or negative.

**Chapter 2**

**Literature Survey**

**2.1 NATURAL LANGUAGE PROCESSING**

Natural Language Processing (NLP) research at Google focuses on algorithms that apply at scale, across languages, and across domains. Our systems are used in numerous ways across Google, impacting user experience in search, mobile, apps, ads, translate and more.

Our work spans the range of traditional NLP tasks, with general-purpose syntax and semantic algorithms underpinning more specialized systems. We are particularly interested in algorithms that scale well and can be run efficiently in a highly distributed environment.

Our syntactic systems predict part-of-speech tags for each word in a given sentence, as well as morphological features such as gender and number. They also label relationships between words, such as subject, object, modification, and others. We focus on efficient algorithms that leverage large amounts of unlabelled data, and recently have incorporated neural net technology.

On the semantic side, we identify entities in free text, label them with types (such as person, location, or organization), cluster mentions of those entities within and across documents (coreference resolution), and resolve the entities to the Knowledge Graph.

Recent work has focused on incorporating multiple sources of knowledge and information to aid with analysis of text, as well as applying frame semantics at the noun phrase, sentence, and document level.

**2.2 A Study of Sentiment Analysis: Concepts, Techniques, and Challenges**

Sentiment analysis (SA) is a process of extensive exploration of data stored on the Web to identify and categorize the views expressed in a part of the text. The intended outcome of this process is to assess the author attitude toward a particular topic, movie, product, etc. The result is positive, negative, or neutral. This study illustrated different techniques in SA approach for extracting and analytics sentiments associated with the polarity of positive, negative, or neutral on the topic selected. Social networks SA can be a useful source of information and data. SA acquires important in many areas of business, politics, and thought. So, this study contains a comprehensive overview of the most important studies in this field from the past to the recent studies till 2017. The main aim of this study is to provide full concept about SA techniques and its classification and methods used it. Also, we give a brief overview of big data techniques and its relation and use in SA field. Because the recent period has witnessed a remarkable development in the use of Big Data (Hadoop) in the process collection of data and reviews from social networks for analysis.

Sentiment analysis (SA) is a process of extensive exploration of data

stored on the Web to identify and categorize the views expressed in a part of the

text. The intended outcome of this process is to assess the author attitude toward a

particular topic, movie, product, etc. The result is positive, negative, or neutral. These

study illustrated different techniques in SA approach for extracting and analytics

sentiments associated with the polarity of positive, negative, or neutral on the topic

selected. Social networks SA can be a useful source of information and data. SA

acquires important in many areas of business, politics, and thought. So, this study

contains a comprehensive overview of the most important studies in this ﬁeld from

the past to the recent studies till 2017. The main aim of this study is to provide full

concept about SA techniques and its classiﬁcation and methods used it. Also, we

give a brief overview of big data techniques and its relation and use in SA ﬁeld.

Because the recent period has witnessed a remarkable development in the use of Big

Data (Hadoop) in the process collection of data and reviews from social networks

for analysis.

Sentiment analysis (SA) is a process of extensive exploration of data

stored on the Web to identify and categorize the views expressed in a part of the

text. The intended outcome of this process is to assess the author attitude toward a

particular topic, movie, product, etc. The result is positive, negative, or neutral. These

study illustrated different techniques in SA approach for extracting and analytics

sentiments associated with the polarity of positive, negative, or neutral on the topic

selected. Social networks SA can be a useful source of information and data. SA

acquires important in many areas of business, politics, and thought. So, this study

contains a comprehensive overview of the most important studies in this ﬁeld from

the past to the recent studies till 2017. The main aim of this study is to provide full

concept about SA techniques and its classiﬁcation and methods used it. Also, we

give a brief overview of big data techniques and its relation and use in SA ﬁeld.

Because the recent period has witnessed a remarkable development in the use of Big

Data (Hadoop) in the process collection of data and reviews from social networks

for analysis.

Sentiment analysis (SA) is a process of extensive exploration of data

stored on the Web to identify and categorize the views expressed in a part of the

text. The intended outcome of this process is to assess the author attitude toward a

particular topic, movie, product, etc. The result is positive, negative, or neutral. These

study illustrated different techniques in SA approach for extracting and analytics

sentiments associated with the polarity of positive, negative, or neutral on the topic

selected. Social networks SA can be a useful source of information and data. SA

acquires important in many areas of business, politics, and thought. So, this study

contains a comprehensive overview of the most important studies in this ﬁeld from

the past to the recent studies till 2017. The main aim of this study is to provide full

concept about SA techniques and its classiﬁcation and methods used it. Also, we

give a brief overview of big data techniques and its relation and use in SA ﬁeld.

Because the recent period has witnessed a remarkable development in the use of Big

Data (Hadoop) in the process collection of data and reviews from social networks

for analysis.

Sentiment analysis (SA) is a process of extensive exploration of data

stored on the Web to identify and categorize the views expressed in a part of the

text. The intended outcome of this process is to assess the author attitude toward a

particular topic, movie, product, etc. The result is positive, negative, or neutral. These

study illustrated different techniques in SA approach for extracting and analytics

sentiments associated with the polarity of positive, negative, or neutral on the topic

selected. Social networks SA can be a useful source of information and data. SA

acquires important in many areas of business, politics, and thought. So, this study

contains a comprehensive overview of the most important studies in this ﬁeld from

the past to the recent studies till 2017. The main aim of this study is to provide full

concept about SA techniques and its classiﬁcation and methods used it. Also, we

give a brief overview of big data techniques and its relation and use in SA ﬁeld.

Because the recent period has witnessed a remarkable development in the use of Big

Data (Hadoop) in the process collection of data and reviews from social networks

for analysis.

Sentiment analysis (SA) is a process of extensive exploration of data

stored on the Web to identify and categorize the views expressed in a part of the

text. The intended outcome of this process is to assess the author attitude toward a

particular topic, movie, product, etc. The result is positive, negative, or neutral. These

study illustrated different techniques in SA approach for extracting and analytics

sentiments associated with the polarity of positive, negative, or neutral on the topic

selected. Social networks SA can be a useful source of information and data. SA

acquires important in many areas of business, politics, and thought. So, this study

contains a comprehensive overview of the most important studies in this ﬁeld from

the past to the recent studies till 2017. The main aim of this study is to provide full

concept about SA techniques and its classiﬁcation and methods used it. Also, we

give a brief overview of big data techniques and its relation and use in SA ﬁeld.

Because the recent period has witnessed a remarkable development in the use of Big

Data (Hadoop) in the process collection of data and reviews from social networks

for analysis.

Diagram, schematic

Description automatically generated

**Fig. 2.1** Sentiment analysis process steps

**2.3 Sentiment Evaluation**

Sentiment analysis is crucial for understanding the thoughts that are being expressed on social media and other websites at an increasingly rapid rate. The vast volume of information that has recently exploded in communication sites, air traffic, and alternative marketplaces cannot be regulated or examined using the conventional methods, thus scientists and researchers have devised high-efficiency systems to handle this data. To make the best choice, the SA must analyze the input and understand its polarity. Data processing in SA consists of five steps: data collection, text preparation, sentiment detection, sentiment categorization, and output display.

## **2.4 Data Collection**

The data collection is the first step in sentiment analysis. The collection of data from sources like user groups, Twitter, Facebook, blogs and commercial website such as amazon.com and alibaba.com, etc. This data cannot be analyzed using traditional methods like scanning, text analysis, and language processing, which is used for extraction and classification. Wei Xu and Tapan Kumar proposed a certain method for a task of paraphrasing and gathering the tweet data called Twitter Streaming API.

## **Text Preparation**

Text preparation examines the data before analyzing it. Some reviews and conversations in the communication sites contain offensive and inappropriate words, so they are examined and preparation to be the result more reliable analysis. This process selects the contents that are not related to the analysis and then removes it. Objective of the process is the removal of spam and inappropriate reviews before sent to automated analysis. In this case, we can use part of speech (POS) technique which are used for text preparation before analysis.

## **Sentiment Detection**

Sentiment detection is the process of finding the sentiment newline expressed in a review by using machine learning technique or NLP technique; these are also called opinion mining (OM) new line and sentiment analysis. Sentiment detection consists of the examination of phrases and sentences extracted from reviews and ideas. All the sentences containing self-expressions like beliefs, opinions, and abuse are retained. Many research studies in this field included different methods of detection, like Lakshmish Kaushik ; one of the recent studies propose a system for automatic sentiment detection in natural audio streams by using POS technique.

Diagram

Description automatically generated

**Fig 2.2** Sentiment classification techniques

Sentiment analysis (SA) is a process of extensive exploration of data

stored on the Web to identify and categorize the views expressed in a part of the

text. The intended outcome of this process is to assess the author attitude toward a

particular topic, movie, product, etc. The result is positive, negative, or neutral. These

study illustrated different techniques in SA approach for extracting and analytics

sentiments associated with the polarity of positive, negative, or neutral on the topic

selected. Social networks SA can be a useful source of information and data. SA

acquires important in many areas of business, politics, and thought. So, this study

contains a comprehensive overview of the most important studies in this ﬁeld from

the past to the recent studies till 2017. The main aim of this study is to provide full

concept about SA techniques and its classiﬁcation and methods used it. Also, we

give a brief overview of big data techniques and its relation and use in SA ﬁeld.

Because the recent period has witnessed a remarkable development in the use of Big

Data (Hadoop) in the process collection of data and reviews from social networks

for an

**Chapter 3**

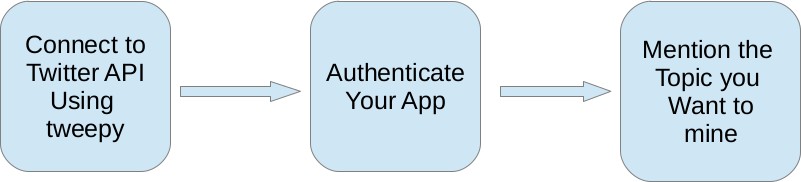
**Methodology**

**3.1 Methodology**

The sentiment analysis of Twitter data is an emerging field that needs much more attention. We use Tweepy an API to stream live tweets from Twitter. User based on his interest chooses a keyword and tweets containing that keyword are collected and stored into a list .Then we make it a labelled dataset using pandas data frame. Next we perform pre-processing to clean, remove unwanted text, characters out of the tweets. Using textblob and setting the sentiment fields accordingly. Thus our train data set without pre-processing is ready Then we train our classifier by fitting the train data to the classifier, there after prediction of results over unseen test data set is made which there after provides us with the accuracy with which the classifier had predicted the outcomes. There after we present our results in a pictorial manner which is the best way to showcase results because of its easiness to understand information out of it.

**3.2 Extraction of data**

Tweets based on a keyword of user’s choice of interest have been collected using a twitter API known as Tweepy and stored into a list. This data set collected for sentiment analysis have tweets based on a keyword e.g.,tatasafari .Tweets mimicking various emotions as a dataset downloaded from twitter is used for sentiment analysis. In order to extract the opinion, first of all data is selected and extracted from twitter in the form of tweets. After selecting the data set of the tweets, these tweets were cleaned from emoticons, unnecessary punctuation marks and a pandas dataframe was created to store this data in a specific transformed structure. In this structure, all the transformed tweets are in lowercase alphabets and are divided into different columns. The details about the steps adopted for the transformation of information are described in next subsections.



**Fig 3.1** Extraction of data

**3.3 Tokenization:**

Tokenization is the process of converting text into tokens before transforming it into vectors. It is also easier to filter out unnecessary tokens. For example, a document into paragraphs or sentences into words. In this case we are tokenizing the reviews into words

**3.4 Classification of tweets:**

Now to classify tweets into positive, negative and neutral we used TextBlob.sentiment.polarity and TextBlob.sentiment.subjectivity function which gives us the sentiment score .On the pandas dataframe we applied getPolarity() and getSubjectivity() which used polarity and subjectivity operation of TextBlob library . The polarity function is used for generating the sentiment scores for each tweet and subjectivity is used to measure the personal opinion of the tweet users. The Sentiment Score of each tweet can be positive or negative or neutral on the basis of opinions of public.

**3.4.1 Sentiment score:**

The Senti-Strength Scale is used for measuring the polarity (Sentiment) score for positive ,negative and neutral tweets in which +1 means Positive , -1 for Negative and 0 for Neutral . The sentiment score is a more specific numerical illustration of the Opinions of people.

Chart, bar chart

Description automatically generated

**Chapter 4**

**Result and Discussion**

**4.1Tweets Collected**

Tweets are collected with the help of Twitter API. When we ran the script shown in Figure 4.1 a .csv file is generated

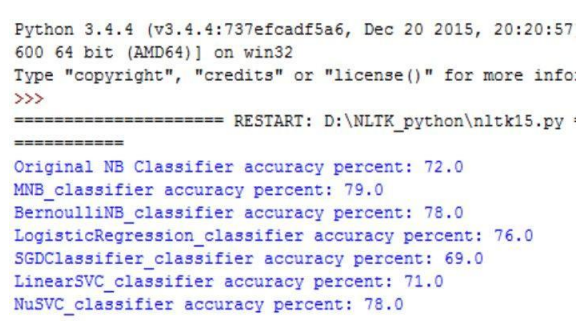
**Graphical user interface, text, application, Word

Description automatically generated**

**Fig 4.1** sample tweets collected

**4.2 Classifier Accuracy for Training Data**

Once we ran the script shown in Figure 4.5, we get the accuracy of each classifier for movie reviews training data

****

**Fig 4.2** Classifiers accuracy for training data

As, we can show from figure 4.2 almost all classifiers are giving accuracy of average 75% and above. Thus, our build classifier is fully trained and ready for sentiment analysis of twitter data.

**Chapter 5**

**Conclusion and Future Work**

We furnished results for Sentiment and Emotional Analysis on twitter data. With the help of TextBlob library which is based on Naive-bayes Classifier. We firmly conclude that implementing sentiment analysis using the TextBlob library will help in deeper understanding of textual data which can essentially serve a potential platform for businesses

Sentiment analysis is used to identifying people’s opinion, attitude and emotional states. The views of the people can be positive or negative. Commonly, parts of speech are used as feature to extract the sentiment of the text. An adjective plays a crucial role in identifying sentiment from parts of speech. Sometimes words having adjective and adverb are used together then it is difficult to identify sentiment and opinion.

To do the sentiment analysis of tweets, the proposed system first extracts the twitter posts from twitter by user. The system can also compute the frequency of each term in tweet. Using machine learning supervised approach help to obtain the results.

Twitter is large source of data, which make it more attractive for performing sentiment analysis. We perform analysis on around 15,000 tweets total for each party, so that we analyze the results, understand the patterns and give a review on people opinion. We saw different party have different sentiment results according to their progress and working procedure. We also saw how any social event, speech or rally cause a fluctuation in sentiment of people. We also get to know which policies are getting more support from people which are started by any of these parties. It was shown that BJP is more successful political part in present time based on people opinion. It is not necessary that our classifier can only be used for political parties. It is general classifier. It can be used for any purpose based on tweets we collect with the help of keyword. It can be used for finance, marketing, reviewing and many more.

**References**

[1] Shiv Dhar, Suyog Pednekar, Kishan Borad- Methods for Sentiment Analysis Computer Engineering, VIVA Institute of Technology, University of Mumbai, India.

[2] <https://youtu.be/ujId4ipkBio>

[3] <https://www.geeksforgeeks.org/twitter-sentiment-analysis-using-python/>

[4]Machine Learning Tom M. Mitchell