

*A Mini Project Synopsis on*  
**Sentiment Analysis**

**T.E. - I.T Engineering**

**Submitted By**

**Pratham Dhanesha                      19104021**

**Shreya Desai                              16104018**

**Pavan Chopra                            19104069**

**Under The Guidance Of**

**Prof. Manasi Choche**



**DEPARTMENT OF INFORMATION TECHNOLOGY**

**A.P.SHAH INSTITUTE OF TECHNOLOGY**

**G.B. Road, Kasarvadavali, Thane (W), Mumbai-400615**

**UNIVERSITY OF MUMBAI**

**Academic year : 2021-22**

## CERTIFICATE

This to certify that the Mini Project report on Sentiment Analysis has been submitted by PRATHAM DHANESHA (19104021), SHREYA DESAI (16104018), PAVAN CHOPRA (19104069) who are a Bonafede students of A. P. Shah Institute of Technology, Thane, Mumbai, as a partial fulfilment of the requirement for the degree in **Information Technology**, during the academic year **2021-2022** in the satisfactory manner as per the curriculum laid down by University of Mumbai.

**Prof. Manasi Choche**

Guide

Prof. Kiran Deshpande

Head Department of Information Technology

Dr.Uttam D. Kolekar

Principal

External Examiner(s)

- 1.
- 2.

Place:A.P.Shah Institute of Technology, Thane

Date:

## ACKNOWLEDGEMENT

This project would not have come to fruition without the invaluable help of our guide **Prof. Manasi Choche**. We express gratitude towards our **HOD, Prof. Kiran Deshpande**, and the Department of Information Technology for providing us with the opportunity as well as the support required to pursue this project. We would also like to thank our teachers who gave us their valuable suggestions and ideas when we needed them. We would also like to thank our peers for their helpful suggestions.

## TABLE OF CONTENTS

1. Introduction.....	5
1.1.Purpose.....	6
1.2.Objectives.....	7
1.3 Scope.....	8
2. Problem Definition.....	9
3. Proposed System.....	10
3.1 PreProcessing of Data.....	11
3.2. Project Architecture.....	12
3.3. Features and Functionality.....	13
4. Project Outcomes.....	14
5. Software Requirements .....	15
6. References.....	16
7. Project Scheduling.....	17
8.Conclusion.....	18

# Chapter No: 01

## **INTRODUCTION**

Nowadays, the age of Internet has changed the way people express their views, opinions. It is now mainly done through blog posts, online forums, product review websites, social media, etc. Nowadays, millions of people are using social network sites like Facebook, Twitter, Google Plus, etc. to express their emotions, opinion and share views about their daily lives. Through the online communities, we get an interactive media where consumers inform and influence others through forums. Social media is generating a large volume of sentiment rich data in the form of tweets, status updates, blog posts, comments, reviews, etc. Moreover, social media provides an opportunity for businesses by giving a platform to connect with their customers for advertising. Our experiments show that features that have to do with Twitter-specific features (emoticons, hashtags etc.) add value to the classifier but only marginally. Features that combine prior polarity of words with their parts-of-speech tags are most important for both the classification tasks. Sentiment analysis (SA) tells user whether the information about the product is satisfactory or not before they buy it. Marketers and firms use this analysis data to understand about their products or services in such a way that it can be offered as per the users requirements. Textual Information retrieval techniques mainly focus on processing, searching or analyzing the factual data present. Facts have an objective component but, there are some other textual contents which express subjective characteristics. These contents are mainly opinions, sentiments, appraisals, attitudes, and emotions, which form the core of Sentiment Analysis (SA). It offers many challenging opportunities to develop new applications, mainly due to the huge growth of available information on online sources like blogs and social networks. For example, recommendations of items proposed by a recommendation system can be predicted by taking into account considerations such as positive or negative opinions about those items by making use of SA.

## 1.1 Purpose

- A social media sentiment analysis **tells you how people feel about your brand online.**
- Rather than a simple count of mentions or comments, sentiment analysis considers emotions and opinions.
- It involves collecting and analyzing information in the posts people share about your brand on social media.
- Sentiment analysis is used **to determine whether a given text contains negative, positive, or neutral emotions.** It's a form of text analytics that uses natural language processing (NLP) and machine learning.

## 1.2 OBJECTIVES

- Twitter sentiment analysis allows you to keep track of what's being said about your product or service on social media, and can help you detect angry customers or negative mentions before they they escalate.
- Regardless of the name, the goal of sentiment analysis is the same: to know a user or audience opinion on a target object by analyzing a vast amount of text from various sources.
- You can analyze text on different levels of detail, and the detail level depends on your goals.

## 1.3 Scope

- Examining public opinion polls and political polls.
- To predict the outcome of an election, anyone can use sentiment analysis to compile and analyze large amounts of text data, such as news, social media, opinions, and suggestions.
- It takes into account how the general public feels about both candidates. Twitter sentiment analysis allows you to keep track of what's being said about your product or service on social media, and can help you detect angry customers or negative mentions before they they escalate.
- Regardless of the name, the goal of sentiment analysis is the same: to know a user or audience opinion on a target object by analyzing a vast amount of text from various sources.
- You can analyze text on different levels of detail, and the detail level depends on your goals.



## **Chapter No:2**

### **PROBLEM DEFINITION**

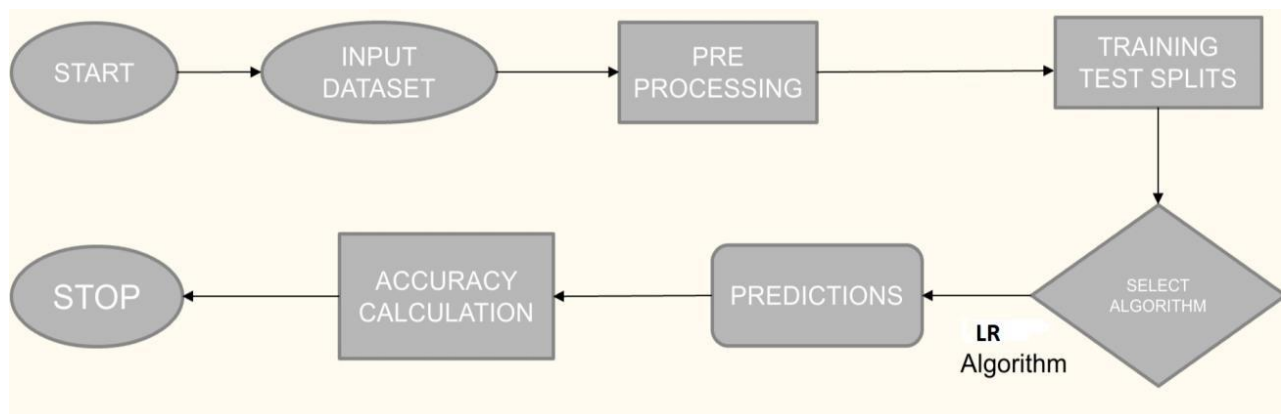
Since the Advent of the Internet, humans have used it as a communication tool in the form of mostly text messages and nowadays video and audio streams and as we increase our dependence on technology it becomes increasingly important to better gauge human sentiments expressed with the help of technology. However, in this textual communication data, we lose the access to sentiments or the emotions conveyed behind a sentence, as we often use our hands and facial expressions to express our intent behind the statement. From this textual data, we can gain insights into the individual. Insights which can be used for multiple different uses such as content recommendation based on current mood, market segmentation analysis and psychological analysis in humans.

In this project, we have attempted to classify human sentiment into two categories namely positive and negative. Which helps us better understand human thinking and gives us an insight which can be used in a variety of ways as stated above. The main aim of this project is to predict stock prices using Logistic Regression(LR)

## Chapter No: 3

### PROPOSED SYSTEM

Sentiments from the dataset given is classified with the help of machine learning and natural language processing (NLP) algorithms, we use the datasets from Kaggle which was crawled from the internet and labeled positive/negative. The data provided comes with emoticons (emoji), usernames and hashtags which are required to be processed (so as to be readable) and converted into a standard form. We also need to extract useful features from the text such unigrams and bigrams which is a form of representation of the “tweet”. We use various machine learning algorithms based on NLP (Natural Language Processing) to conduct sentiment analysis using the extracted features. Finally, we report our experimental results and findings at the end.

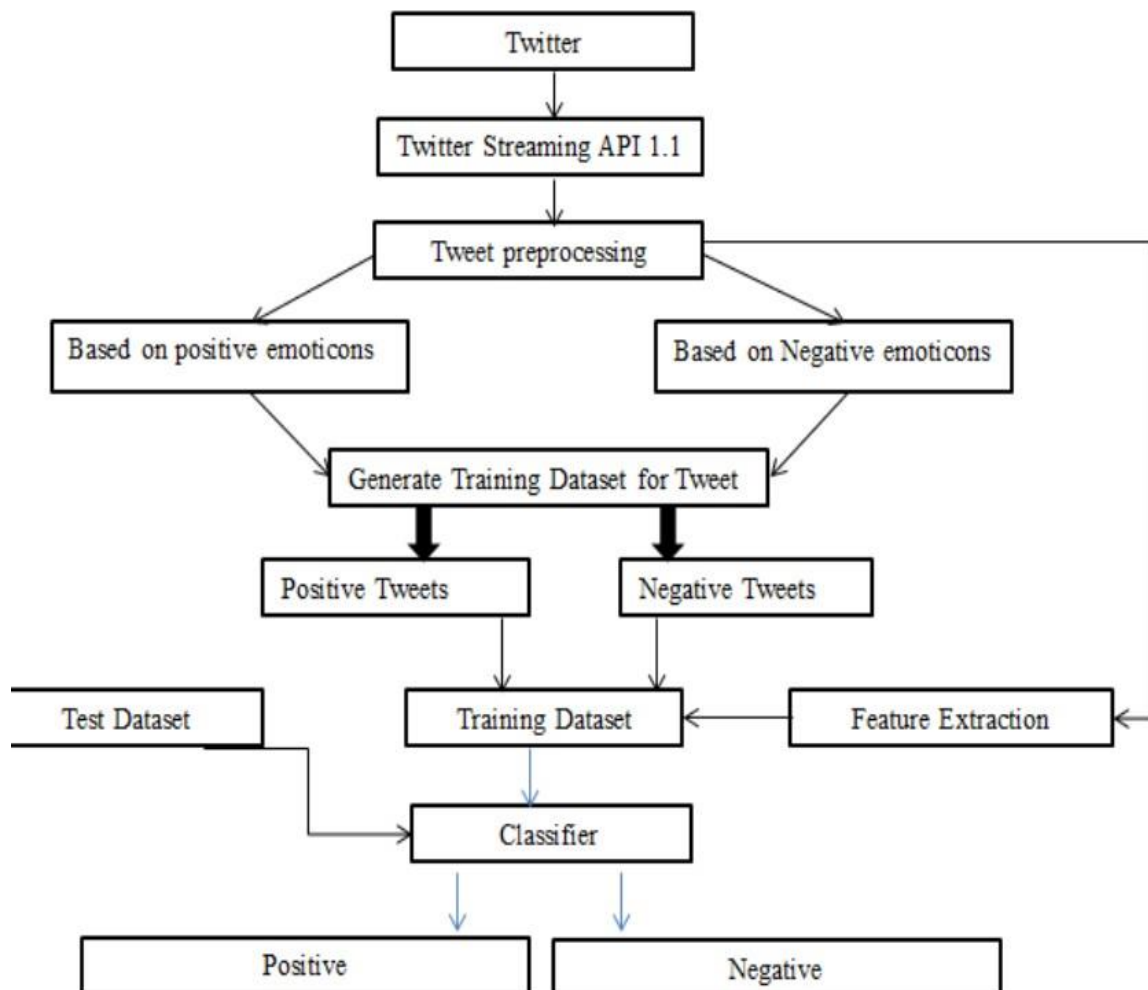


## 3.1 Pre-processing of data:

A tweet contains a lot of opinions about the data which are expressed in different ways by different users. The twitter dataset used in this survey work is already labeled into two classes viz. negative and positive polarity and thus the sentiment analysis of the data becomes easy to observe the effect of various features. The raw data having polarity is highly susceptible to inconsistency and redundancy. Preprocessing of tweet include following points,

- Remove all URLs (e.g. [www.xyz.com](http://www.xyz.com)), hash tags (e.g. #topic), targets (@username)
- Correct the spellings; sequence of repeated characters is to be handled
- Replace all the emoticons with their sentiment.
- Remove all punctuations ,symbols, numbers
- Remove Stop Words
- Expand Acronyms(we can use a acronym dictionary)
- Remove Non-English Tweet

## 3.2 Problem Architecture



### **3.3 Features and Functionality**

Sentiment analysis or opinion mining is field of research that can have significant impact on today's business. Increasing number of consumers' reviews created the need of its automatic analysis. This issue is gaining popularity for both – researchers and entrepreneurs, for whom consumers' reviews are important source of business information. There are three main areas of opinion mining: opinions classification , feature based opinion mining, comparative sentences analysis. This paper is focused on feature based sentiment analysis in which not the sentiment of the whole opinion is analyzed but how particular features of opinion's subject are seen. For the purpose of research ontology-based approach and pattern based approach to opinion mining are used. First of them allows to present products features in a form of hierarchy while second automatizes the process of features with its sentiments extraction. To every feature the sentiment is assigned.

## **Chapter No: 4**

### **Project Outcomes**

1. Twitter allows businesses to engage personally with consumers. However, there's so much data on Twitter that it can be hard for brands to prioritize which tweets or mentions to respond to first.
2. That's why sentiment analysis has become a key instrument in social media marketing strategies.
3. Sentiment analysis is a tool that automatically monitors emotions in conversations on social media platforms.

## **Chapter No: 5**

### **Software Requirements**

- Backend- Jupyter/colab
- Front end- Streamlit
- Python 3.10.4 in Google Colab is used for data pre-processing, model training and prediction.
- Operating System: windows 7 and above

## **Chapter No: 6**

### **PROJECT SCHEDULING**

No.	Group Member	Time Duration	Work to be done
	Pratham Dhanesha	1 <sup>st</sup> week of January	Worked on backend
		2 <sup>nd</sup> week of January	Trained and tested dataset using algorithm
	Shreya Desai	3 <sup>rd</sup> week of January	Worked on GUI
	Pavan Chopra	By the end of march month	Connect backend and front end



## 7. REFERENCES

- M. S. Neethu and R. Rajasree, "Sentiment analysis in twitter using machine learning techniques," 2013 Fourth International Conference on Computing, Communications and Networking Technologies (ICCCNT), 2013, pp. 1-5, doi: 10.1109/ICCCNT.2013.6726818.
- G. Gautam and D. Yadav, "Sentiment analysis of twitter data using machine learning approaches and semantic analysis," 2014 Seventh International Conference on Contemporary Computing (IC3), 2014, pp. 437-442, doi: 10.1109/IC3.2014.6897213.
- M. Kanakaraj and R. M. R. Guddeti, "Performance analysis of Ensemble methods on Twitter sentiment analysis using NLP techniques," Proceedings of the 2015 IEEE 9th International Conference on Semantic Computing (IEEE ICSC 2015), 2015, pp. 169-170, doi: 10.1109/ICOSC.2015.7050801

## **8. CONCLUSION**

In this paper, we provide a survey and comparative study of existing techniques for opinion mining including machine learning and lexicon-based approaches, together with cross domain and cross-lingual methods and some evaluation metrics. Research results show that machine learning methods, such as SVM and naive Bayes have the highest accuracy and can be regarded as the baseline learning methods, while lexicon-based methods are very effective in some cases, which require few effort in human-labeled document .We also studied the effects of various features on classifier. We can conclude that more the cleaner data, more accurate results can be obtained. Use of bigram model provides better sentiment

