

# Twitter Sentiment Detection using Machine Learning

Sahaja P<sup>1</sup>, Dr M Nagalakshmi<sup>2</sup>

<sup>1</sup>Marri Laxman Reddy Institute of Technology and Management, Dundigal

<sup>2</sup>Associate Professor, Dept. of Computer Science and Engineering, Marri Laxman Reddy Institute of Technology and Management College, Telangana, India

\*\*\*

**Abstract** - With the rise of social networking epoch and its growth, the internet has become a promising platform for online learning, exchanging ideas, and sharing opinions. Social media contain an enormous amount of sentiment data within the sort of tweets, blogs, and updates on the status, posts, etc. In this paper, the foremost popular micro-blogging platform Twitter is employed. Twitter sentiment detection is an application of sentiment analysis on data from Twitter (tweets), to extract user's opinions and sentiments. The most goal is to explore how text analysis techniques are often used to probe a variety of the data during a series of posts that concentrate on different trends of tweets languages, tweets volumes on twitter. Experimental evaluations show that the proposed machine learning classifiers are efficient and perform better in terms of accuracy. The proposed algorithm is implemented in python.

**Key Words:** Sentiment Detection, Python, Twitter, Sentiments, Opinions, Tweets

## 1. INTRODUCTION

Sentiment Detection may be a process of the deriving sentiment of a specific statement or sentence. It's a way that derives opinion from the tweets and formulates a sentiment and on the idea of which, sentiment classification is performed. Sentiments are subjective to the subject of interest. It's required to formulate what features will decide for the sentiment it embodies. The sentiment we ask maybe a class of entities that the person performing sentiment analysis wants to seek out within the tweet. Sentiment analysis approaches are often broadly categorized into two classes – lexicon-based and machine learning-based. Lexicon based approach is unsupervised because it proposes to perform analysis using lexicons and a scoring method to gauge opinions. Whereas the machine learning approach involves the use of feature extraction and training the model using a feature set and a few datasets. the essential steps for performing sentiment analysis include data collection, pre-processing of data, feature extraction, selecting baseline features, sentiment detection, and performing classification either using simple computation alternatively machine learning approaches. Twitter has emerged as a serious micro-blogging website, having over 100 million users generating over 500 million tweets a day.

With such an outsized audience, Twitter has consistently attracted users to convey their opinions and perspective

about any issue, brand, company, or the other topic of interest. So as to extract sentiment from tweets, sentiment analysis is employed.

The results from this will be utilized in many areas like analyzing and monitoring changes of sentiment with an occasion, sentiments regarding a specific brand or release of a specific product, analyzing public view of state policies, etc.

### 1.1 Existing System

Surveys and Questionnaires: A group of questions with a choice of answers, devised for a product or a service helps tons to realize knowledge on effectiveness.

Interviews: Watching Interviews with tech experts helps in being updated with the newest technology.

Feedback: Many products and services invite consumer feedback. Companies make a record of both negative and regeneration and use these records to return up with a far better product and repair.

However, the prevailing traditional methods are time-consuming and as they're performed manually, their efficiency decreases with a rise within the frequency. Another important problem is that a customer might not convey information properly during a formal setting in comparison to an off-the-cuff medium like social media.

### 1.2 Problem Statement

Previous studies came up with technical solutions to extract user sentiment on government health monitoring system, influenza, Netflix, product. Despite the extensive literature, none of the solutions have how forum relationships affect network dynamics. within the proposed system, sentiment analysis is completed using tongue Processing, which defines a relation between user-posted tweets and opinion on the merchandise, and additionally, suggestions of a way better product are often provided to the users. Thereby the disadvantage of identifying a relation between user post and therefore the opinion related to the post is overcome.

## 2. LITERATURE SURVEY

### Opinion Mining on Social Media Data [8]:

Po-Wei Liang et.al used Twitter API to collect data from twitter. Tweets which contain opinions were filtered out. Unigram Naive Bayes model was developed for polarity identification. They also worked for elimination of unwanted features by using the Mutual Information and Chi square feature extraction method. Finally, the approach for predicting the tweets as positive or negative did not give better accuracy by this method.

### A Sentiment Analysis Prototype System for Social Network Data [3]:

This paper discusses a prototype system to enable automated opinion-based analysis of social network data for evaluation of the public social conscience of user provided topics and events.

### Sentiment Analysis for Social Media [1]:

Sentiment Analysis is a problem of text based analysis, but there are some challenges that make it difficult as compared to traditional text based analysis This clearly states that there is need of an attempt to work towards these problems and it has opened up several opportunities for future research for handling negations, hidden sentiments identification, slangs , polysemy. However, the growing scale of data demands automatic data analysis techniques. In this paper, a detailed survey on different techniques used in Sentiment Analysis is carried out to understand the level of work.

### Sentiment Analysis for Using Twitter Dataset [6]:

The research of sentiment analysis of Twitter data can be performed in different aspects. This paper shows sentiment analysis types and techniques used to perform extraction of sentiment from tweets. In this survey paper, we have taken comparative study of different techniques and approaches of sentiment analysis having twitter as a data.

## 3. PROPOSED SYSTEM

Within the proposed system, the primary step involves the gathering of tweets from twitter and making it a knowledge set, the second step is preprocessing of the related tweets. within the third step, sentiment analysis is performed using the tongue Processing (NLP) algorithm, which is predicated on numerical statistics. Assigned sentiment value using NLP, is employed as a weighting think about sentiment analysis. within the fourth step, similar data is identified and analyzed, then the ultimate results, which are suggestions of other changes available for the merchandise, are provided.

## 3.1. SYSTEM ARCHITECTURE

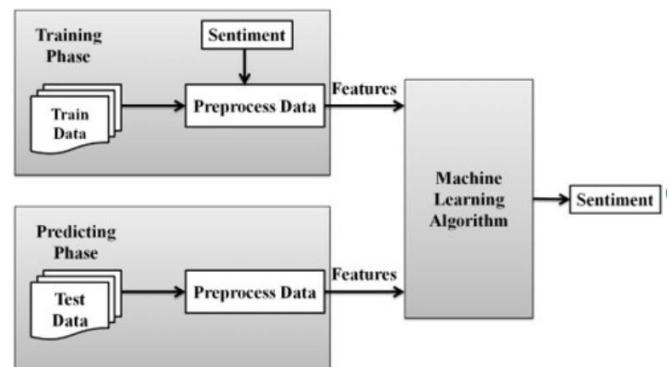


Fig -1: System Architecture

Train data or learning set — is that the information used to train an algorithm. The training data includes both input data and thus the corresponding expected output. Test data includes only input file, not the corresponding expected output. The testing data is used to assess how well your algorithm was trained, and to estimate model properties. Train data and Test data are preprocessed with meta-features on which a machine learning algorithm is applied.

## 3.2. WORKFLOW OF SYSTEM

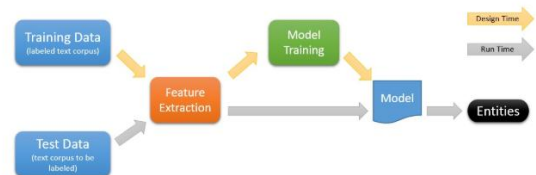


Fig -2: Workflow of System

A Named Entity Recognition (NER) can discover entity elements from data and determines the category the element belongs to. The system reads the sentence and highlights the important entity elements within the text. NER could be given separate sensitive entities counting on the project. this suggests that NER systems designed for one project might not be reused for an additional task. Once the user selects the test, train files, he/she should perform exploratory data analysis to removes stopwords and define the common words within the dataset. Named Entity Recognition algorithm must be used on the dataset and later predict the accuracy and sentiment with the assistance of the algorithm.

## 4. ADVANTAGES

The use of this information is often applied to form wiser decisions associated with the utilization of resources, to form improvements in organizations, providing better products/services, and ultimately to enhance the citizen lifestyle and human relations so as to realize a far better society. An example of this application is that the impact of

tracking people's feelings on products, services, and events, which permit enterprise managers to possess knowledge and parameters for decision-making. Another example is council administrators that would have the chance for improving the services offered to citizens and for addressing challenges of development and sustainability more efficiently supported what people feel.

## 5. CONCLUSION

Twitter sentiment detection comes under the category of text and opinion mining. It focuses on analyzing the sentiments of the tweets and feeding the data to a machine learning model in order to train it and then check its accuracy, so that we can use this model for future use according to the results. It comprises of steps like data collection, text pre-processing, sentiment detection, sentiment classification, training and testing the model. This topic has evolved during the last decade with models reaching the efficiency of almost 85%-90%. But it still lacks the dimension of diversity in the data. Along with this it has a lot of application issues with the slang used and the short forms of words. Many analyzers don't perform well when the number of classes are increased. Also, it's still not tested that how accurate the model will be for topics other than the one in consideration. Hence sentiment analysis has a very bright scope of development in future.

## REFERENCES

- [1] Ke Tao, Claudia Hauff, Geert-Jan Houben, txing AG,TU Delft, "Facilitating Twitter Data Analytics: platform, Language and Functionality".
- [2]A. Pak and P. Paroubek,"Twitter as a Corpus for Sentiment Analysis and Opinion Mining," In Proceedings of the Seventh Conference on International Language Resources and Evaluation, 2010, pp.1320-1326 .
- [3]R. Parikh and M. Movassate, "Sentiment Analysis of User-Generated Twitter Updates using Various Classification Techniques",CS224N Final Report, 2009.
- [4]Go. R.Bhayani, L.Huang, "Twitter Sentiment Classification Using Distant Supervision," Stanford University, Technical Paper,2009.

## BIOGRAPHIES



### **Sahaja P**

Studying at Marri Laxman Reddy Institute of Technology and Management, Dept. of Computer Science and Engineering