



---

## **Twitter Sentimental Analysis : Predicting the sentiment by analysing the tweets and representing the output in the form of pie chart**

***Arya Patil<sup>a</sup>, Bhakti Koshti<sup>a</sup>, Shweta Vibhute<sup>a</sup>, Preeti Powar<sup>a</sup>, Deepal Shinde<sup>b</sup>***

<sup>a</sup>*Students, Sanjay Ghodawat Polytechnic, Atigre, Kolhapur, Maharashtra, India*

<sup>b</sup>*Teacher, Sanjay Ghodawat Polytechnic, Atigre, Kolhapur, Maharashtra, India*

---

### **ABSTRACT**

This study describes a system for real-time analysis of popular opinion toward presidential candidates stated on Twitter during the 2012 U.S. election. Twitter has become a popular platform for people to voice their thoughts on political parties and politicians. Emerging events or news are frequently followed by a spike in Twitter volume, providing a unique chance to assess the relationship between expressed public mood and political outcomes. Furthermore, sentiment analysis can be used to investigate how these occurrences influence public opinion. Unlike traditional content analysis, which can take days or weeks to complete, the system shown here analyzes sentiment in all of the election-related Twitter traffic and delivers results quickly and continually. It provides a fresh and contemporary perspective on the dynamics of the election process and public opinion to the general public, the media, lawmakers, and academics.

Sentiment analysis is the process of recognizing and categorizing the views and attitudes represented in a source text. Tweets, status updates, blog entries, and other forms of social media generate a large volume of sentiment-rich data. Sentiment analysis of this user-generated data is quite beneficial for determining the general consensus. The two methodologies for analyzing feelings from text are the knowledge base approach and machine learning approach. It is possible to determine the effect of domain information on sentiment classification by doing sentiment analysis in a specific domain. We show a novel feature vector for categorizing tweets as positive or negative, as well as extracting people's opinions on items and firms. Nowadays, social media is getting more attention. Public and private opinions on a wide range of topics are constantly expressed and distributed via a variety of social media platforms.

---

### **1. Introduction**

Twitter sentiment analysis helps you to keep track of what people are saying about your product or service on social media, and it can help you spot unhappy consumers or unfavorable remarks before they become major issues.

Sentiment analysis is a procedure that uses Natural Language Processing to extract attitudes, opinions, perspectives, and emotions from text, audio, tweets, and database sources (NLP). Sentiment analysis is the process of categorizing textual opinions into categories such as "positive," "negative," or "neutral." Sentiment analysis is a strong marketing tool that allows product managers to better understand client sentiments and use that information in their marketing initiatives. When it comes to product and brand awareness, customer loyalty, customer happiness, advertising and marketing success, and product acceptability, it is a critical aspect.

On Twitter, users can send tweets, which are short communications of up to 140 characters. The number of people using Twitter is steadily increasing. Over 100 million active users globally send over 250 million tweets per day, according to the business (Twitter, 2012). As of May 2011, 13 percent of online American adults were actively using it, up from 8% a year earlier (Pew Research Center, 2011). More than two-thirds of members of Congress have set up a Twitter account, and many are utilizing it to communicate with their constituents (Lassen & Brown, 2010; TweetCongress, 2012). We've collected nearly 36 million tweets about the 2012 presidential candidates since October 12, 2012, a quarter million every day on average. We gathered more than half a million relevant tweets in only a few hours during one of the most important political events of the year, the Iowa primary debate on

December 15, 2011. This type of 'big data' much exceeds the capacity of standard content analysis methods, necessitating the development of new computing methods. The majority of previous research has concentrated on post-facto analysis of tweets, with conclusions arriving days or even months after the data was collected. In a separate paper, we will discuss our approaches to these difficulties. Here, we'll show the entire system as well as the visualization dashboards we've created. In section 2, we begin with a review of related work; in section 3, we describe the architecture and components of our system (input, preprocessing, sentiment model, result aggregation, and visualization); and in sections 4 and 5, we assess our early experience with this system and discuss next steps.

---

## 2. Literature Review

Industry 4.0 manufacturing systems are generally described in 2017 as a smart factory with IoT and CPSs at its core. The quality control is regarded as self-predicting and self-aware, and the products are mass personalized. The resource management is self-configured and self-optimized, with the creation of CPS and IoT as development priority. Big data analysis, the Internet of Things, and Cyber-Physical Systems have all been mentioned as key technologies.

Cloud computing, simulation, augmented reality, additive manufacturing, horizontal and vertical system integration, autonomous robotics, cyber security, and the Internet of Services have all recently been added to the three fundamental technologies. The vast range of technologies is predicted to boost the efficiency, agility, and flexibility of manufacturing processes, resulting in cost savings [6, 8]. Manufacturing companies use cloud computing as • Infrastructure as a Service (IaaS): computer resources are housed in the cloud infrastructure. • Platform as a Service (PaaS): Applications that run on cloud infrastructure • Software as a Service (SaaS): Applications that live and function on cloud infrastructure

A. Opinion mining is a vast topic of natural language processing, text mining, and computational linguistics that entails the computer study of sentiments, views, and emotions represented in text. Although a sentiment is a term used to describe a viewpoint or attitude that is based on emotion rather than logic.

B. Twitter: Twitter is a prominent real-time microblogging website that allows users to exchange brief messages called tweets that are limited to 140 characters. Users create tweets to convey their thoughts on a variety of topics that affect their daily lives.

C. Sentiment Analysis: Sentiment can be found in comments or tweets, and it can be used to provide useful indicators for a variety of applications. As a result, and as said, a sentiment can be classified into two groups: negative and positive terms.

---

## 3. Advantages and Disadvantages

### Advantages

- Sentiment Analysis is a text mining technique that is commonly used. As a result, Twitter Sentiment Analysis entails analyzing text sentiments or tweets utilizing cutting-edge text mining approaches in the negative, positive, and neutral modes.
- Twitter Sentiment Analysis is far more than a college assignment or a certification program. Many decent tutorials on Twitter sentiment are available to teach students how to create a Twitter sentiment analysis report and how to use it with Python.
- Some well-known tools for Twitter sentiment analysis are Revealed Context, Enginuity, Steamcrab, SocialMention, and MeaningCloud. Python and R are widely used for sentiment analysis on Twitter.

### Disadvantages

- The key issues with present methodologies are their inability to perform well across domains, their lack of accuracy and performance in sentiment analysis due to insufficient labeled data, and their inability to deal with complicated sentences requiring more than sentiment terms and easy analyses.

---

## 4. Objectives

- Graphical representation of sentiment in the form of a pie-chart
- It attempts to assess people's sentiments, attitudes, opinions, feelings, and so on.
- Create an algorithm for automatically classifying text as good, negative, or neutral.

---

## 5. Methodology

Sentiment analysis of Twitter data is a new topic that requires a lot more research. Figure 1 depicts the procedures involved in conducting sentiment analysis on Twitter data. To begin, the obtained Twitter data is pre-processed in order to do data cleaning. Second, using any of the feature selection methods, the important features are extracted from the clean text.

#### Module 1: Data Sources

A considerable role is played by the data source used to conduct the sentiment analysis. The data sources for social media platforms are divided into three categories: blogs, microblogging sites, and review sites [13-16]. Due of its limited content strength and publically available data, a micro-blogging site such as Twitter has garnered more popularity than others in all categories. It's clear from the following Twitter growth rate statistics that Twitter should be used as a data source for sentiment research.

#### Module 2: Twitter Data Collection Method

The following are three methods for collecting Tweets for study [11]: UCI, Friendster, Kdnuggets, and SNAP are examples of data repositories. • APIs: Twitter offers two different types of APIs: search API and stream API. • Automated tools that are further classified into premium tools like Radian6 [18], Sysmos, Simplify360, Lithium, and non-premium tools like Keyhole, Topsy, Tagboard, and SocialMention to collect Twitter data based on hashtags and stream API to stream real-time data from Twitter.

#### Module 3: Data Preprocessing

Data mining on Twitter is a difficult task. The information gathered is unprocessed. It is necessary to pre-process or clean the raw data before using a classifier. Uniform casing, removal of hashtags and other Twitter notations (@, RT), emoticons, URLs, stop words, decompression of slang terms, and compression of prolonged phrases are all part of the pre-processing operation.

#### Module 4: Feature Extraction

Various discrete properties exist in the pre-processed dataset. Using feature extraction methods, we extract several characteristics such as adjectives, verbs, and nouns, which are then classified as positive or negative in order to determine the polarity of the entire phrase. The methods for Feature Extraction that are most commonly utilized are listed below.

---

## 6. System Configuration

### 1. Hardware Requirement

- Laptop or PC - Minimum 4GB RAM
- Processor version - greater than i3
- Hard Disk - 4GB Storage (minimum)

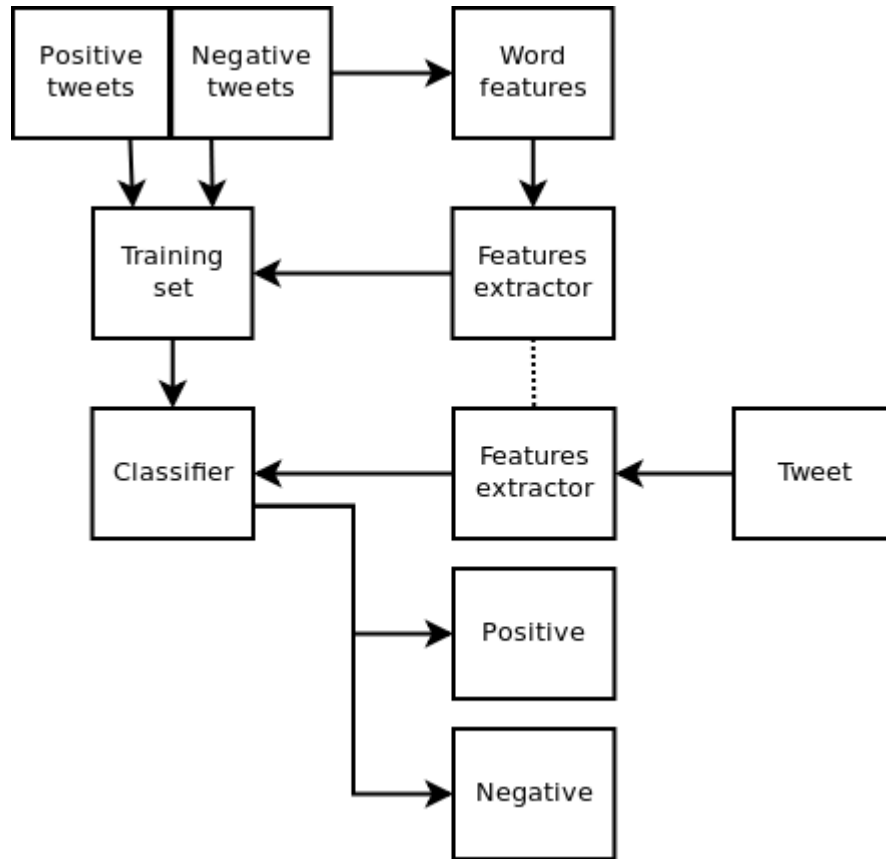
### 2. Software Requirement

- Unix or linux -Windows 7,Windows 8,window 10
- Languages - python 3 or upper version with NLTK package
- XAMPP

### 3. Database Requirement

- MySQL

## 7. Flowchart



## 8. Scope of Project

Without having to read thousands of client comments at once, sentiment analysis allows you to determine how customers feel about various aspects of your organization. It is hard for one individual to read all of the comments if you receive thousands of feedback per month.

The sentiment analysis will continue to delve deeper, far beyond the surface of likes, comments, and shares, with the goal of genuinely comprehending the value of social media interactions and what they reveal about the people behind the screens..

## 9. Conclusion

We demonstrated a technique for real-time Twitter sentiment analysis during the ongoing 2012 presidential election in the United States. To gain a complete and accurate view of the online political scene, we employ the Twitter "firehose" and expert-curated criteria and keywords. Our real-time data processing infrastructure and statistical sentiment model assesses public attitude changes as new political events and news emerge. The architecture and approach are general, thus they can easily be applied to other domains (for instance, we used the system for gauging sentiments about films and actors surrounding Oscar nomination and selection).

**Reference**

- [1] I. King, J. Li and K. T. Chan, "A Brief Survey of Computational Approaches in Social Computing", in Proc. of Int. Joint Conf. on Neural Network, 2009, pp. 2699-2706.
- [2] S. R. Barahate and V. M. Shelake, "A Survey and Future Vision of Data mining in Educational Field", in Proc. 2nd Int. Conf. on Advanced Computing Communication Technology, 2012, pp. 96-100.
- [3] Bing Liu, N. Indurkha and F. J. Damerau, Handbook of Natural Language Processing, Second Edition, 2010, pp. 1-3860-68.
- [4] M. Dredze, "How Social Media Will Change Public Health", IEEE Intelligent Systems, 2012, pp. 1541-1672.
- [5] G. Siemens and P. Long, "Penetrating the fog: Analytics in learning and education", Educause Review, 2011, vol. 46, no. 5, pp. 30-32