

# Recognizing Sentiment Prediction on Twitter Data

Nishu Sethi, Neha Bhateja, Navya Sethi, Sakshi Sinha

**ABSTRACT**– Escorted by the wide spread of Internet today, people have found a new way of expressing their opinions. It is a platform with a variety of information where an individual can also view the opinions of others. This is continuously growing and becoming an important factor in decision making for various organisations, businesses and even for Politics. In this paper we have chosen the most popular social media platform i.e. Twitter for our Sentiment Analysis. Eventually, Acknowledging the opinions beyond the tweets is of great concern. The fundamental aim of Sentiment Analysis is to reason feelings and ideas of individuals. We have made data analysis with tweets related to a topic and thereby classified their polarity using different machine learning algorithms.

**KEYWORDS**– Sentiment Analysis, Opinion Mining, Literature Review, Supervised Machine Learning.

## I. INTRODUCTION

Accompanied by the immeasurable widening of web innovation, the figure of people revealing their perspective and opinions across social- media platforms are also increasing. Micro blogging websites is the most important source of information about almost all variety of topics. This is due to the fact that a majority of people post their view-point on diversity of subjects that they can publish their likes and dislikes about any new place they visit or any new product they buy or any new dish they eat. People take a lot of interest in discussing current affairs and politics online too. Sentimental scrutiny [1] is the task of deriving the quality information from the data. In other words, it is the process of deriving structured data from given unstructured data. This is used to measure opinion [2] of the customer, upon feedback and product reviews.

**Manuscript received May 14, 2020**

**Nishu Sethi**, Department of Computer Science, Amity University, Gurgaon, Haryana, India (email: nsethi@ggn.amity.edu)

**Neha Bhateja**, Department of Computer Science, Amity University, Gurgaon, Haryana, India

**Navya Sethi**, Department of Computer Science, Amity University, Gurgaon, Haryana, India

**Sakshi Sinha**, Department of Computer Science, Amity University, Gurgaon, Haryana, India

Unstructured data does not only refer to the tables and figures from the organization but it also consists of information from the web i.e. chats, mails, text files, word files, E-Commerce websites and social media sites.

On structured data analytics operation can be easily performed and the results can be obtained with much ease. But in case of unstructured data from E-mail, Twitter etc., it is quite difficult to conclude the output because of various problems such as unspecific data. In this paper, we look at one such popular micro blogging site called Twitter.

Sentiment analysis on tweets is considered good due to the following reasons.

1. Abstract nature of tweets.
2. It enables real time analysis.
3. A big range of data is present for the analysis

## II. LITERATURE REVIEW

Chunxu Wu [3] implemented a framework for context-dependent opinions which are impotent to be identified by WordNet. Such a framework identifies the sentiment of opinion using Semantic measures. This framework helps to govern the inclination of opinions when the information provided is limited and was effective. Paroubek and Pak [4] extracted data from twitter using Twitter APIs and emoticons to perform Sentiment Analysis. The frequency distribution for each term was also evaluated during the analysis. The three classification algorithms such as Conditional Random Fields, Multinomial Naive Bayes and Support Vector Machine were built on the training datasets. Later then, performance of different features was analysed. Xia et al. [5] intended to collate different classification algorithms and various feature sets to create another efficient and strong classifier. Therefore, three classifiers namely Naive Bayes, Maximum Entropy, Support Vector Machine and two feature types Word-relations and Part-Of-Speech (POS) were used for analysis. As a result, it was noted that ensemble methods and various classifiers showed significant improvements. Saif et al. [6] bought in the concept of extracting features and defining the polarity of tweets. Saif et al. [6] came up with the scheme of consolidating Semantic accompanied by Unigram and Part-of-Speech (POS) features. With the results, it was demonstrated that semantic features were able detect the sentiment of the tweets in a better way. In the end the accuracy of each model was measured. Po-Wei Liang et.al [7] extracted data from Twitter using APIs. Naive Bayes classification model was implemented to determine the polarity of tweets. As a result, the accuracy was measured. Khan et al. [8] introduced a new framework for opinion mining. A Hybrid scheme was used which

included Emoticon analysis and SentiWord Net analysis with an intensified classifier to determine the polarity. In the end, the proposed framework resulted with an average harmonic mean value. Hagen, Matthias et al. [9] worked on Ensemble model known as “Webis”. For this model, four Classifiers were combined all together. A score was also generated for each classifier to further make predictions. The “Webis” classifier was the winner in the SemEval-2015 so it is considered as a base classifier. Vishal et al. [10] performed opinion mining. Methodology employed was Vocabulary-Based and Machine-Learning algorithms. Some classifiers that were used were Maximum Entropy, Support Vector Machine and Naive Bayes. All limitations while performing Sentiment analysis were noted. Tripathy et al. [11] completed Sentiment analysis using Supervised Learning Algorithms. Reviews were categorised according to their polarity using the classification algorithms. The workflow consisted of pre-processing the data. Then reviews were converted to numeric matrix using which vectors were generated. In the end, performance of each classifier was measured. Fouad et al. [12] attempted to fetch data from Twitter and classifying them basis polarity. The classifiers that were utilised are Support Vector Machines, Logistic Regression and Naive Bayes. The extracted set of tweets were broken into two different Datasets namely Testing and Training Dataset. Each classifier was trained using the training dataset and results were recorded respectively. Hetu et al. [13] has done Sentiment Analysis on Twitter Data by considering a downloaded dataset. Emotions of people were discovered based on positive and negative reviews. Accuracy was then recorded. According to the proposed model, the accuracy recorded was high and the dataset considered was large.

**III. PROPOSED METHODOLOGY**

Figure 1 demonstrates the proposed model which will be implemented for Sentiment Analysis.

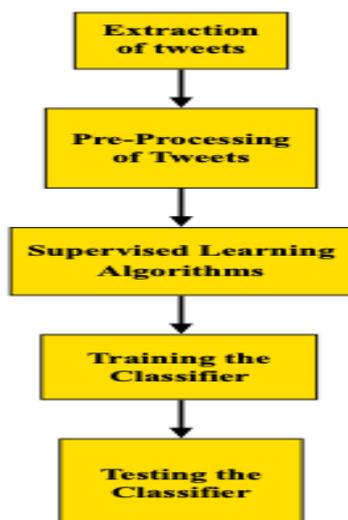


Fig 1: Proposed Framework

The tweets extracted from Twitter will be pre-processed removing stop-words, special and numeric characters. Additionally, Supervised Learning Algorithms [14] could be utilized to train the model and classify tweets basis polarity. Further the model could be evaluated and analysed basis its prediction.

**VI. RESULTS AND DISCUSSION**

Along with Sentiment Analysis, Polarity Percentage can be calculated with the proposed model. Efficiency of each classifier can also be evaluated. Distinct Learning Algorithms can also be utilized to compare each algorithm in terms of performance. With the help of matrix and graphical representation, the results can be visualized. For example, if two learning algorithms are to be compared basis their performance then the result analysis could be the course of action:

If total numbers of tweets fetched are 60 then following matrix displays number of tweets for three different sentiments.

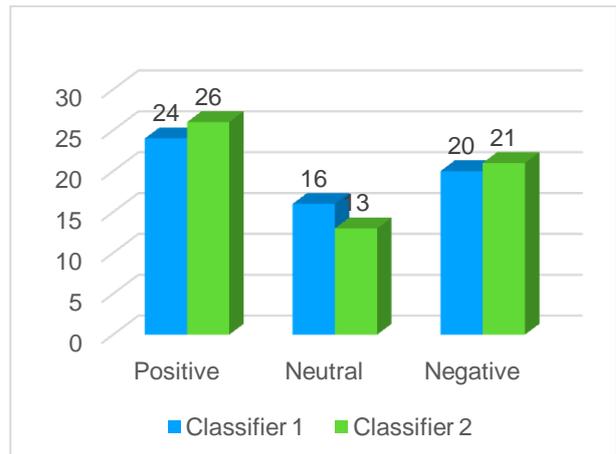


Fig 2: Graphical Representation Analyzed Tweets

**V. CONCLUSION AND FUTURE SCOPE**

Sentiment Analysis is an emerging and interesting way for the application of Natural Language Processing in making automated conclusions about texts and their meaning. Sentimental analysis for deriving the sentiments has gained a lot of importance for many organisations and individuals today.

Table 1: Table for Sentiment Analysis of Tweets

Classifier	Positive	Neutral	Negative
Classifier 1	24	16	20
Classifier 2	26	13	21

Sentiment analysis is an emerging field decision making and is developing at a very fast pace. Our aim through this project is to analyze the sentiments on a topic using tweets related to them which are extracted from the Twitter and determine the nature of the topic in terms of polarity i.e. how much positive, negative or neutral it is in people’s belief. Distinct Machine Learning Algorithms could be employed to perform the sentiment analysis. However, relying on individual models does not generate a very high accuracy so we continued with the top few algorithms and generated the model. Many researches are present in literature for understanding the sentiments of the texts. Still, there is a huge scope for improvement of these models that already exist. The models can be improved with using different approaches and common-sense knowledge.

## REFERENCES

- [1] T. Wilson, J. Wiebe, and P. Hoffman. Recognising Polarity and performing Sentiment Analysis of Contextual phrases, 2005.
- [2] Harb, M. Roche, P. Poncelet, M. Plantie, F. Troussel and G. Dray performed "Web Opinion Mining: How to extract opinions from blogs" in the 5<sup>th</sup> International Conference on Soft Computing as Trans-disciplinary Science and Technology: pp:211-217, 2008: ACM.
- [3] C. Wu, X. Wang and L. Shen, "A new method of using contextual information to infer the semantic orientations of context dependent opinions" in Computational Intelligence and Artificial Intelligence; International Conference, AICI'09, Vol:4; IEEE, pp:274-278, 2009.
- [4] P. Paroubek and Pak, "Twitter as a corpus for Sentiment Analysis and Opinion Mining"; LREc, Vol:10, 2010.
- [5] R.Xia, C.Zong and S.Li, "Ensemble of feature Sets and Classification Algorithms for Sentiment Classification, "Information Sciences"; Vol:181, no.6, pp:1138-1152, 2011.
- [6] H.Saif, H.Alani and Y.He, "Semantic Sentiment Analysis of Twitter", in International Semantic Web Conference, Springer; pp: 508-524, 2012.
- [7] Po-Wei Liang and Bi-Ru Dai, "Opinion Mining on Social Media Data" in 14<sup>th</sup> IEEE International Conference: Mobile Data Management, Milan, Italy, pp:91-96, ISBN:978-149673-6068-5, June 3-6, 2013.
- [8] F.H. Khan, U. Qamar and S. Bashir, "Twitter Opinion Mining Framework using Hybrid Classification Scheme" in Decision Support Systems; Vol:57, pp:245-257, 2014.
- [9] M. Hagen, B. Stein, M. Buchner and M.Potthast "Webis, an Ensemble for Twitter Sentiment Detection" in the 9<sup>th</sup> International Workshop on Semantic Evaluation 2015, pp:582-589.
- [10] V. Kharde and P. Sonawane, "Sentiment analysis of twitter data: A survey of techniques," arXiv preprint arXiv:1601.06971, 2016.
- [11] A.Tripathy and S.K.Rath, experimented with "Supervised Machine Learning Techniques for Sentiment Classification Of Reviews" in "International Journal Of Rough Sets and Data Analysis (IJRSDA), Vol:4, no.1, pp:56-74, in 2017.
- [12] M.M. Fouad, A.S. Mashat and T.F. Gharib, "Efficient Twitter Sentiment Analysis with Feature Selection and Classifier Ensemble", in International Conference: Advanced Machine Learning Technologies and Applications; Springer; pp:516-527, 2018.
- [13] Hetu Bhaysar and Richa Manglani, "Sentiment Analysis Of Twitter Data using Python" in International Research Journal Of Engineering And Technology (IRJET), e-ISSN: 2395-0056 p-ISSN:2395-0072, Mar 2019.
- [14] K.P.Murphy, "Naive Bayes classifiers" Vol. 18, in the University of British Columbia, 2006.