Examine the Behaviour of
People to Sociality Using Twitter

Akash Jadhav[1], Apeksha Jadhav[2], Reshma Bhosale[3], Ankita Chaskar[4]

BE Student, Information Technology, VPKBIET, Baramati, India[1,2,3,4]

Abstract: This Paper is review on Sentiment Analysis of Twitter data for examining the behaviour of people. Here, we focus on Twitter which is the most popular micro blogging platform. We collect tweets from the Twitter. These Tweets express emotions, feelings and point of view of people towards any particular type of subject. From the view of decision maker this collection of tweets provides Valuable source of information. Through this information we are able to examine the behaviour of people. This paper represents how we can examine the behaviour of people by performing Text Classification, Data mining, natural Language Processing and Sentiment Analysis.

Keywords: Data Mining, Sentiment Analysis, Text Classification, Natural Language Processing

I. INTRODUCTION

Now-a-days micro blogging is the most popular platform and widely used by people. This is due to nature of micro blog on which people can share their experience, post their opinion, reviews about the product. People can also share their views about any particular topic. In this paper we used one such popular micro blog called Twitter, Through the Twitter platform user can give opinion about famous personalities, products, companies, events etc. Twitter user varies from regular user to celebrities, company representative, politicians and even country presidents. That's why it is easy to collect tweets of user from different social and interests groups and also from different countries. So, we are collecting the tweets from twitter. With the help of these tweets we can analyse the behaviour of people towards a particular topic or product. So that the organizations or companies can obtain knowledge about how the users react to their products and according to that they can improve their performance. Social media mining requires human data analysts and automated software programs to sift through massive amounts of raw social media data. These patterns and trends are of interest to companies, governments and non-profit organizations as it helps them to design their strategies or introduce new programs [1]. Data mining is the computational process of discovering patterns in large data sets involving methods at the intersection of artificial intelligence, machine learning, statistics, and database systems. The overall goal of data mining process is to extract information from a dataset and transform it into an understandable structure for further use [1]. This Project aims to design a system that can analyse and summarize the data available on social media.
 The current work encompasses two main themes.
• Sentiment Analysis: The goal is to mine opinions from social media data at sentence and document level.
• Text Classification: The goal is to group similar text message that can make information more manageable. It employs a framework that can cluster similar topics that belong to the same topic and several similar topics that can belong to broader category.
Using Text classification as the basis of data mining, we thus can predict the behaviour of the people on social media. In section 2, we are discussing previous research in the field of sentiment analysis on Tweets In section 3 we are discussing comparative analysis of previous methods used for Sentiment analysis Tweets using social media mining.

II. RELATED WORK

This approach collects particular trending dataset which has tweets from random people and try to classify them accordingly. We use data mining for this approach. There are several data mining techniques that can be used for this purpose. Reza Zafarani et.al. [2] Present a novel challenges encountered in social media mining these challenges are as follows.
1. Big Data Paradox.
2. Obtaining Sufficient Samples.
3. Noise Removal Fallacy.
4. Evaluation Dilemma
Manisha Rani et.al. [5] Proposes the paradigm to extract the sentiment from famous micro blogging service, called Twitter. In this paper, data mining techniques are used to automatically classify the sentiments of Tweets taken from Twitter dataset [1].

Sriram et al. [3] classified tweets to a predefined set of generic classes such as news, events, opinions, deals, and private messages based on author information and domain specific features extracted from tweets such as presence of shortening of words and slangs, time-event phrases, opinionated words, emphasis on words, currency and percentage signs, @username at the beginning of the tweet, and @username within the tweet [1].

Sankaranarayanan et al. [4] has built a news processing system that identifies the tweets corresponding to late breaking news. Issues addressed in their work include removing the noise, determining tweet cluster of interest using online methods, and identifying relevant locations associated with the tweets [1].

Luo et. al. [7] highlighted the challenges and an efficient technique to mine opinions from Twitter tweets. Spam and wildly varying language makes opinion retrieval within Twitter challenging task. This paper looks to improving the data we collect from twitter. It explores whether the social aspects of the author have any impact on the data. It gives that by analysing the social structure of a person we can improve the retrieval performance [1].

Sameesksha Shrivastava et. al. [8], In the presented work the identified twitter data set is used to perform text analysis. The entire input data samples are required to classify in two classes namely positive and negative. The following steps were performed during the mining and subsequent analysis [1].

Shruti Wakade et. al. [6] use Weka data mining tools to extract useful information for classifying sentiment of tweets paper size. Collected from Twitter. The results of tweet mining are represented as decision trees that can be used for judging sentiment of new tweets [1].

III.   COMPARATIVE ANALYSIS OF DIFFERENT WORKS

Srno	Author	Discussion
1	Manisha Rani et.al.[5]	Proposes the paradigm to exact the sentiment from famous microblogging service.
2	Sriram et.al.[3]	Classified tweets to a predefined set of generic classes such as news, events, opinion, deals based on author information and domain feature extracted from Tweets.
3	Sankarnarayanam et.al.[4]	Built a news processing system that identifies the Tweets corresponding to late breaking news. Issued addressed in their work include removing the noise. Determining Tweet cluster of interest using online method and identifying relevant location associated with Tweet.
4	Luo et.al.[7]	Highlighted the challenges and an efficient technique to mine opinion from Twitter Tweets.
5	Sameesksha Shrivastava et.al.[8]	Presented work the identified Twitter data set is used to perform text analysis. The entire input data samples are required to classify in two classes namely positive and negative.
6	Shruti wakade et.al.[6]	Use weka data, mining tool to extract useful information for classifying sentiments of tweets, collected tweets from twitter. The results of Tweet mining are represented as decision trees that can be use for judging sentiment of new Tweets.

Fig: 1. Comparative Analysis of Different Works

IV.   PROPOSED SYSTEM ARCHITECTURE

System architecture is the conceptual model that defines the structure behaviour and more use of system. System architecture consist of four stages.
1. Data Collection
2. Pre-processing
3. Data Modelling
4. Sentiment Analysis

1) Data Collection Stage:- In this system, tweets are collected from twitter. These tweets are related to trending topics like organizations, famous personalities, music, sports, politics etc. These collected tweets contain the information such as time, text, date, country, language. These tweets are in json format.

2) Pre-processing of data:- The data that we get is in json format. This format makes it easy for humans to read the data and for machines to parse it. That data is stored in the data frame.

3) Data Modelling:- Stored data from the data frame is converted into tokens.

4) Sentiment Analysis:- Intensity of that token is calculated. According to intensity, polarity gets assign to respective token. Based on the polarity score, tweets are classified into positive, negative and neutral.
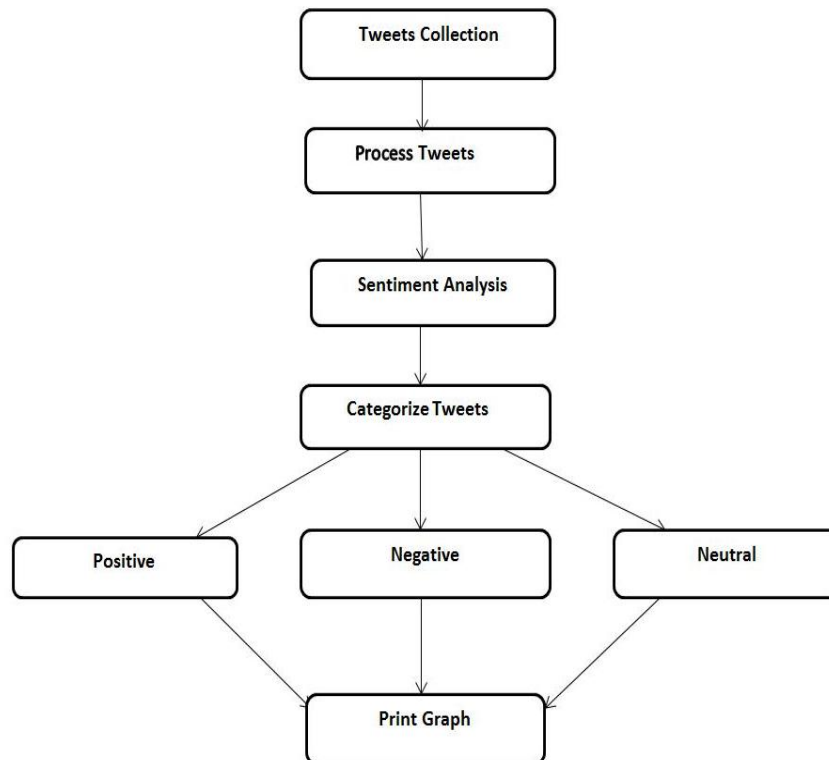

Fig: 2.System Architecture

CONCLUSION

Social media is huge source of information. User creates their service specific profile for any application that design and maintained by the social media organization. Twitter is one type of social media on which user can post their message called tweets. By using these tweets analysing of behaviour of peoples becomes easy. Analysing the views of peoples towards the product of company, examining the behaviour of organization and famous personalities can understand what people think about them with the help of this system. Once organization or company get to know what people actually think about their services according to that they can improve their services. Output of the system is in graphical format and also count of positive, negative and neutral tweets are displayed so, more detailed information can be obtained. According to this it is more useful for organizations, companies or individual people for improve their performance.

REFERENCES

[1]. Ardra, Blessy Merin Varughese, Merline Susan Joseph, Preethi Elsa Thomas, Sherly K K," Analyzing the Behavior of Youth to Sociality Using Social Media Mining" Department of Information Technology, Rajagiri School of Engineering and Technology Kochi, India

[2]. Chinemela Queen Adougo, Ovute A.O., Obochi Charles I., The influence of social media on the Nigerian youths: Aba residents experience, Quest Journals Journal of Research in Humanities and Social Service, 2015. Vol. 3, no. 3, pp.12-20.

[3]. Reza Zafarani, Mohammad Ali Abbasi and Huan Liu, Social Media Mining-An Introduction, 2014, A Book

[4]. B. Sriram, D. Fuhry, E. Demir, H. Ferhatosmanoglu, and M. Demirbas, Short text classification in twitter to improve information filtering, In Proceeding of the 33rd internationalACM SIGIR conference on Research and development in information retrieval, 2010, pp. 841842.

[5]. J. Sankaranarayanan, H. Samet, B. E. Teitler, M. D. Lieberman, and J.Sperling, Twitterstand: news in tweets, in Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, 2009, pp. 4251.

[6]. Manisha Rani, JyotiArora, Twitter Data Predicting Geolocation Using Data Mining Techniques, International Journal of Innovative Research in Computer and Communication Engineering, June 2016, vol. 4, no. 6, pp. 10446-10453

[7]. Shruti Wakade, Chandra Shekar, Kathy J. Liszka and Chien-Chung Chan.Text Mining for Sentiment Analysis of Twitter Data.

[8]. Zhunchen, L., Miles, O., Ting, W., An effective approach to tweets opinion retrieval, Springer Journal on World Wide Web, May 2015, vol. 18, no. 3, pp 545566.

[9]. Sameesksha Shrivastava, Dr. Pramod S. Nair, Mood Prediction on Tweets Using Classification Algorithm, International Journal of Science and Research, 2015, vol. 4, no. 11, pp. 295-299

[10]. Apoorv Agarwal Boyi Xie Ilia Vovsha Owen Rambow Rebecca Passonneau, sentiment Analysis of Twitter Data, [11] Alexander Pak and Patrick Paroubek. 2010. Twitter as a corpus for sentiment analysis and opinion mining. Proceedings of LREC.