© 2017 IJSRSET | Volume 3 | Issue 3 | Print ISSN: 2395-1990 | Online ISSN : 2394-4099
Themed Section: Engineering and Technology

# Sentiment Analysis : A key factor in Social Media

**Dipesh Pratap Singh Tilara, Chandan Kumar**
Bansal Institute of Engineering & Technology, Lucknow, India

## ABSTRACT

Reviews play a crucial role in determining the growth of the product. At present, analyzing people's review from social networking site is in trend as people now-a-days are interested more in knowing other person's opinion about any particular object or issue. It deals with finding the opinion, identifying them and classifying opinion and the sentiment behind them from text. It is a field in which people are finding more interest, but at the same time it is very complex to analyze the sentiments as people can have different opinion about an entity at different situations. Facebook, Yahoo, Twitter are such social sites, which provide large number of reviews on a particular object. These reviews play an important role in the marketing of a particular product. These reviews can change the market either in the favor of the product or downgrade the rating of that product. Reviews are the base of creating positive, negative or neutral wave in the market about any particular product, which in a way making it easy for both consumers and sellers of the product to understand each other's need. In this paper, we are focusing on sentiment analysis of data in the form of tweets on different issues collected from Twitter.

**Keywords :** Twitter, Entity, Sentiment Analysis, Naïve Bayes, Maximum Entropy, Random Forest.

## I. INTRODUCTION

In recent time, interest of people in internet and online marketing has changed the focus of e-commerce from seller to buyer's view. Customers can share their opinion about any entity on social networking sites which provides information to other customers about that particular entity. Analysis of Sentiment shows the actual sentiment of a person through the text which he or she is posting in the form of a social media message. This information sharing can be important because the sentiment of a person can affect other persons, connected to him. People can also make a change in the sentiment of persons which belong to their social network. Sentiment analysis has become very important, especially for small business persons who are looking to leverage data. There are many different companies that are spending money on sentiment analysis of their customers so that they can find out what user like and what they want, which conclusively helps them to improve the products and services that they have to offer and also increases their demand among consumers. Users are very influenced by these social networks so they named it on their own as: Social Media Marketing (SMM). The big difference between social media and face-to-face communication is that, with their social network, people have a more chances to affect many people of their connections all at once. "Word-of-mouth (WOM) marketing communication strategies have always been around [1] to us but traditional WOM, which was the oral form of communication [2], has involved in a new form of communication, which is known as electronic Word-Of-Mouth (e-WOM) communication."

Sentiment analysis or Opinion mining is becoming increasingly important as number of organization are curious to know what their consumers are thinking at the exact moment and it's being shared via social blogs, posts on Facebook or Twitter and review forums, etc.. Business owners can use websites, such as Twitter, Facebook and many others, to find out what is in the demand or trend, what people like at present and do not like so that failure and negativity of any product can be avoided and to motivate the customers to buy their product or take their services.

Sentiment analysis tools can be divided into two categories:

1. An Automated sentiment analysis system is where "an algorithm processes the text sharing and determines

its overall sentiment, generally dropping comments as either positive or negative."

2. Manual or human sentiment analysis requires "the intervention of a human element into the analysis and is required to dissect abbreviations, sarcasm, emotions, slang etc. and determine the true expressed sentiment."

The purpose of this paper is to analyze the sentiment and opinions expressed by peculiar people from different area of the world from tweets on various issues. This paper presents the result by applying classification on tweets of different issues like Political media, Global warming, Coachella festival and some Hollywood movies to classify as positive, negative and neutral statements using three algorithms which are Naïve Bayes, Maximum Entropy and Random Forest.

## II. Literature Review

Social networking sites and micro-blogs are becoming very important communication medium today amid the users. Twitter, Facebook and Yahoo have become most popular social sites. Tweets are short messages of the length 140 charactersin which people write about their views, activities, news, opinions, emotions, etc., on Twitter which reach and affecta large number of people on the internet. Twitter has become the most popular micro-blogging site today with more than 350 million monthly active users and more than 500 million tweets per day (March2016). Twitter supports around 35+ languages and 85% of Twitter users are active on mobile.

**Akcora et al.** [3] in their experiment, try to "identify the emotional pattern and the word pattern that claims to change the public opinion, using Twitter data. To identify the breakpoint, researchers use Jaccard's similarity of two successive intervals of words."

**Sun et al.** [4] study "fan pages on Facebook to understand diffusion trees."

**Kwak et al.** [5] compare "the number of followers, page-ranks and number of re-tweets as three different measures of influence. Their finding is that the ranking of the most influential users differed depending on the measure."

**Cha et al.** [6] compare "the number of followers, thenumber of re-tweets and the number of mentions as measures and conclude the most followed users do not necessarily score the highest on other measures."

**Weng et al.** [7] compare "the number of followers and page ranks with a modified page rank measures that accounted for topic and conclude that the ranking is dependent on the influence measure."

**Patil et al.** [8] presents "novel approach from text entered by users on social networking sites for emotion estimation. This paper is also helping users to express their feeling/emotion by generating images according to their feeling/emotion in text."

**Tanbeer et al.** [9] their experiment helps social network users to"find a group of friends who are significant across multiple domains in social networks."

**Pak et al.** [10] created "a twitter corpus by automatically collecting tweets using Twitter API and automatically annotating those using emoticons. Using that corpus, they built a sentiment classifier based on the multinomial Naive Bayes classifier that uses N-gram and POS-tags as features. The training set is less efficient if it contains only tweets having emoticons."

## III. Research Approach and Methods

In this paper, model is trained using Naïve Bayes, Maximum Entropy and Random Forest method and this model is used to classify the data with positive, negative and neutral label. The steps of experiment and how to collect useful data from raw data, is shown below. The length of tweet is fixed to 140 characters, it can't be extended. Due to this, people use short form of words or in other words, acronyms and symbols to express their emotions and attitude. Therefore, it is very tough to classify the tweets into those three classes, i.e., positive, negative and neutral statements. The proposed work which helps to achieve the objective of paper is:

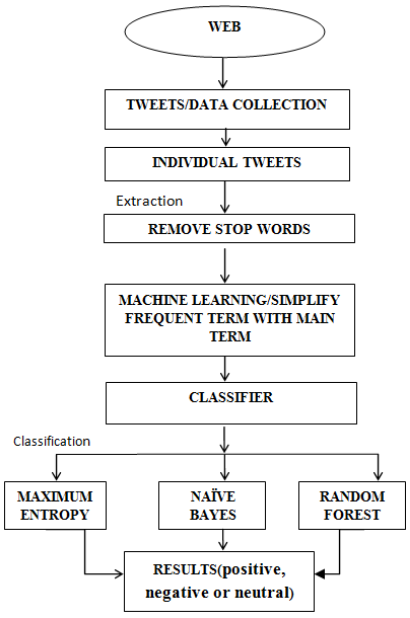


**Figure 1**. Proposed work

## IV. Data Collection

Usually, per day 550 million tweets are posted on Twitter and are stored in Twitter corpus. Many Twitter APIs are available where tweets can be accessed. Users can collect the data from many available websites that provides the reviews of movies in the form of tweets. There are some sites like 'sentiment140.com' and 'boxofficemojo.com' for the tweets' access. These sites contain tweets related to any topic. There is a sample of sentiment in all three categories in the following table:

## V. Results

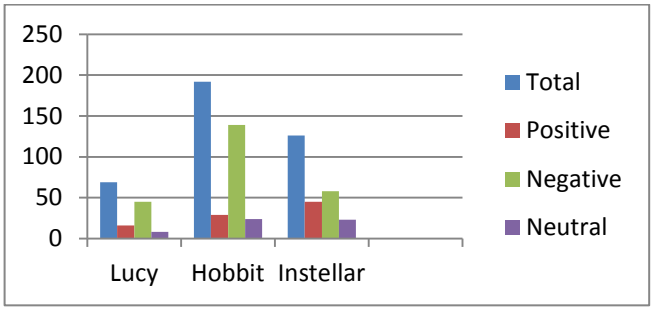


**Figure 2**. Sample tweets

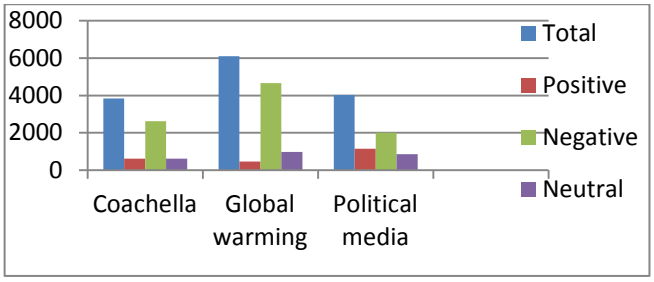


**Figure 3**. Sentiments of tweets using Maximum Entropy

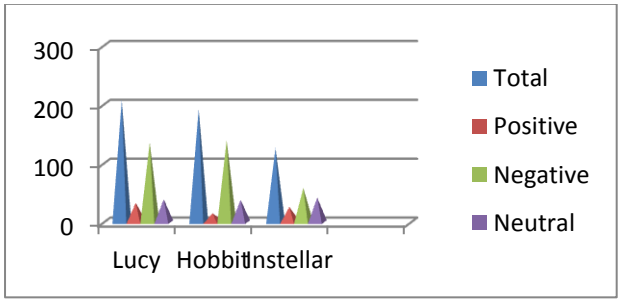


**Figure 4**. Sentiments of tweets using Naïve Bayes

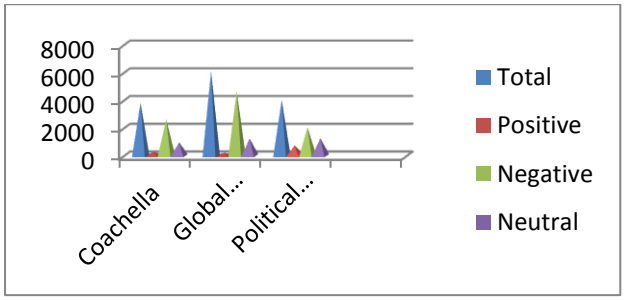


**Figure 5**. Sentiments of tweets using Random Forest

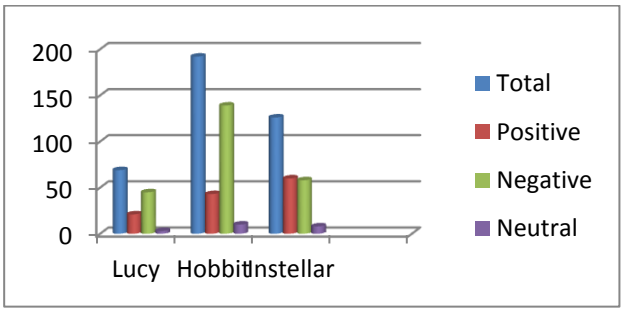


**Figure 6**. Sentiments in % by Maximum Entropy

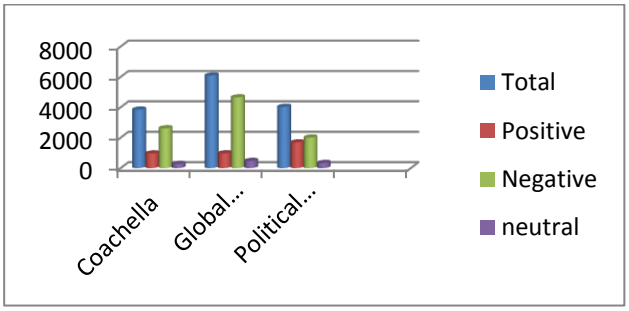


**Figure 7**. Sentiment in % by Naïve Bayes

The Twitter data is collect in the raw format with username, time stamp, Geo-location and actual message. For collecting the actual message from raw data, we will be applying text mining and after that analyze the sentiment of tweets.

| MOVIE NAME | POSITIVE (%) | NEGATIVE (%) | NEUTRAL (%) |
|---|---|---|---|
| LUCY | 23 | 65 | 12 |
| HOBBIT | 15 | 72 | 13 |
| INSTELLAR | 36 | 46 | 18 |
| COACHELLA | 16 | 68 | 16 |
| GLOBAL WARMING | 8 | 76 | 16 |
| POLITICAL MEDIA | 29 | 50 | 21 |

| SENTIMENT'S POLARITY | EXAMPLE OF TWEETS |
|---|---|
| POSITIVE | McDonald's burger tastes awesome. |
| NEGATIVE | The billing process of McDonald's is poor. |
| NEUTRAL | Seating arrangements were normal. |

| MOVIE NAME | POSITIVE(%) | NEGATIVE (%) | NEUTRAL (%) |
|---|---|---|---|
| LUCY | 16 | 65 | 19 |
| HOBBIT | 8 | 72 | 20 |
| INSTELLAR | 21 | 46 | 33 |
| COACHELLA | 7 | 68 | 25 |
| GLOBAL WARMING | 3 | 77 | 20 |
| POLITICAL MEDIA | 18 | 51 | 31 |

## VI. Conclusion and Future Scope

The message expressed in Twitter can be related to the human behavior, nature, personality and attitude. People can express their opinionin positive, negative or neutral way. If a sentiment is neither positive nor negative then it is neutral sentiment. Classification of tweets into different category specifies the opinions of people on various issues. People of different regions and cultures express their opinion about these issues differently according to their culture and choice. I will be working on this to expand my work by including higher number of tweets and more number of testing models.

## VII. REFERENCES


[1]. UmeshRaoHodeghatta, Sentiment Analysis of Hollywood Movies on Twitter, 2013 IEEE/ACM International Conference on Advances in Social Networking Analysis and Mining.

[2]. Arndt, J. Role of Product-Related Conversations in the Diffusion of a New Product. Journal of Marketing Research, 4, 291-295. 1967.

[3]. Akcora, C. G., Bayir, M. A., Demirbas, M., and Ferhatosmanoglu, H., "Identifying breakpoints in public opinion", In ACM Proceedings of the First Workshop on Social Media Analytics, July, pp. 62-66. 2010.

[4]. Sun, E., Rosenn, I., Marlow, C. and Lento, T. (2009), "Gesundheit! modeling contagion through Facebook news feed", In Proc. Of International AAAI Conference on Weblogs and Social Media, May, pp. 22.

[5]. Kwak, H., Lee, C., Park, H. and Moon, S. (2010), "What is twitter, a social network or a news media?", In ACM Proceedings of the 19thInternational Conference on World Wide Web, April, pp. 591-600.

[6]. Cha, M., Haddadi, H., Benevenuto, F. and Gummadi, K. P., "Measuring user influence in Twitter: the million follower fallacy", In 4th international AAAI Conference on Weblogs and Social Media (ICWSM), May, Vol.14, No.1, pp. 8.. 2010.

[7]. Weng, J., Lim, E. P., Jiang, J. and He, Q. (2010), "Twitterrank: finding topic sensitive influential twitterers", In Proceedings of the Third ACM International Conference on Web Search and Data Mining, February, pp. 261-270.

[8]. Patil,T. ; "Automatic Generation of Emotions for Social Networking Websites Using Text Mining", Computing, Communication and Networking Technologies(ICCCNT), 2013 Fourth International Conference, July 2013

[9]. Tanbeer, S.K. ; "Finding Groups of Friends Who are Significant Across Multiple Domains in Social Networks", Computational Aspects of Social Networks(CASoN), 2013 Fifth International Conference, August 2013.

[10]. A. Pakand P. Paroubek, "Twitter as a corpus for sentiment analysis and opinion mining," in Proceedings of LREC, vol.2010,2010.