Analysis of Various Sentiment Analysis Techniques

[1]Prateema Sharma, [2]Dr. Neelam Duhan, [3]Ms. Shilpa Sethi

[1]Student, [2]Assistant Professor, [3]Assistant Professor
[1]Information Technology,
[1]YMCAUST, Faridabad, India

Abstract: The sentiment analysis is the technique which is applied for the sentiment analysis. The technique of classification is generally applied for the sentiment analysis. There are various classification techniques which are implemented for sentiment analysis. The classification techniques are like SVM, KNN, and Naïve Bayes. In this review paper, various sentiment analysis techniques are reviewed and analyzed in terms of certain parameters.

Keywords **- SVM, KNN, Naïve Bayes.**

I. INTRODUCTION

With the advent in technology and its growth large amount of data is generated on daily basis by millions of people. There is change in expressing thoughts, views and opinions nowadays due to the internet. Blog posts, online forums, product review websites, social media and many more are the platforms using which user can communicate and remains connected with everyone [1].In order to express emotions, opinion and share views about their daily lives people are using social network sites like Facebook, Twitter, and Google and so on. It is an optimal platform by which user can interact, discuss their issues, happiness and achievements by connecting to the online communities. In the form of tweets, status updates, blog posts, comments, reviews a large amount of data is created by the social media. It provides an opportunity for businesses by which they can linked with their customers for advertising. People are now relying on the internet for making their decisions hence they searched the user-generated content over online to solve their issues. Therefore, various sentiment analysis techniques has been utilized as the amount of content generated by users is too vast for a normal user to analyze.

Sentiment analysis is the process in which feelings, thoughts, and emotion for a particular event of every individual is analyzed [2]. Positive, negative or neutral are the three attributes to determine the stage of opinion using sentiment analysis. It shows that sentiment analysis can be interpreted as a task of a classification in which each sentiment represents a new category. The computer science and artificial intelligence field Natural Language Processing (NLP) plays an important role in dealing with interaction between human and computer language. This field is particularly of use to merchants, stock traders, and in election works. Due to change in marketing level competition and the needs of the people a lot of research is needed in sentiment analysis.

The main objective of the textual Information retrieval techniques is to extract the present data by processing, searching or analyzing it. The factual data present mainly focus on processing, searching or analyzing the retrieval textual information technique. It's very challenging to develop new applications due to huge growth of available information gathered from blogs and social networks [3]. Twitter is mainly used to tweet ongoing messages using microblogging and social networking sites. New challenges have been created due to different unique characteristics present in tweets that also shape the different domains and methods conveyed in sentiment analysis. The sentiment analysis on Twitter is mainly discussed in his paper. User expressed their views and opinions on micro-blogging website such as Twitter for the extraction of sentiments of users. The generalized architecture of sentiment analysis is shown below:

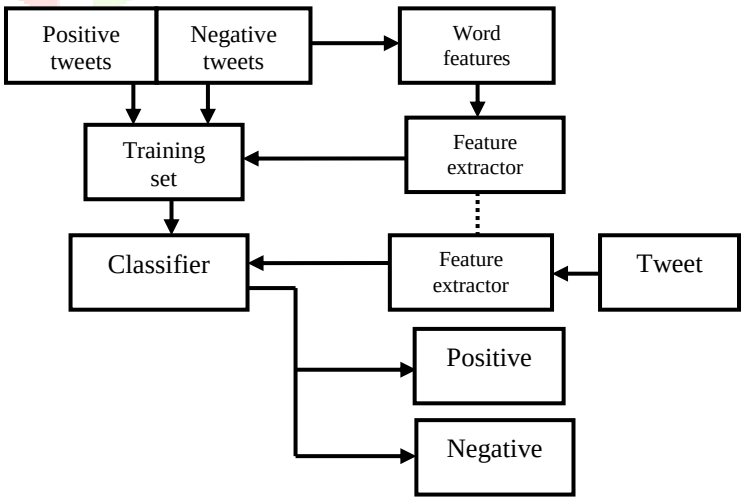

Fig.1 Sentiment Analysis Architecture

1.1 Benefits of using sentimental analysis:

There are various areas in which benefits are provided by the sentiment analysis such as business and organization [4]. It will allow user to identify potential of their products to create more influence on social media. Hence, it will be used to identify

negative threads that are emerging online regarding your business, thereby allowing you to be practical in dealing with it more quickly. Some of the benefits are as follow:

a. Marketing strategy: Social sites has been utilized by most of the companies in order to promote their brands and services, this is the place where customers provide full review about existing brand that how their brand is perceived by targets customers. Marketing campaign after short span of time act as a tactical point to know what customer need. In order to target customers, they can fit campaign even more useful by doing continuous sentiment analysis.

b. Crisis management: An essential role is played by the social-media in determining what are the current issues from which customers had gone through due to this constant monitoring on social-media is necessary to mitigate the damage of online communications crisis. Some factors led to crisis such as bad quality, serious social issues such as environmental harm, child-labor or unacceptable customer-service. If the customer complaints didn't manage on yearly basis, it can go viral and lead to huge crisis that might not be able to handle in near future.

c. Enlarge product quality: This can help in finding the customer opinions about their needed products, quality and features. Product quality is not only judged by how well they performed but also by how nice and in which form they are presented [5]. To enhance their product quality and how well it will be presented can be derived from the targets opinions. First step to do that is by conducting a structured and a planned survey and second, gathering information's that are going on related to their brand in public social platforms.

d. Improve customer service: The customers can essentially become micro-influencers in case if customer purchases product and keep them loyal. The utmost priority of the brand that a best customer service should be provided and keep their current customers happy [6]. Sentiment analysis helps to eliminate negative negotiations and provide a real-time analysis so that work can be done accordingly. Giving faster response to customer complaints will satisfied the customer at first place.

1.2 Approaches for sentiment analysis

There are mainly two techniques for sentiment analysis for the Twitter data:

a. Lexicon based approach
b. Machine learning approach

Lexicon Based Approach: opinion lexicon for sentiment classification is based on the insight that the polarity of a piece of text can be obtained on the grounds of the words which compose it. This approach is of two types: positive or negative opinions. Positive opinion is used to show the desired stage and negative stage is used to show undesired stage. Opinion lexicon is also known as opinion phrases or idioms.

- **Dictionary-Based Approach:** The resources used are lexicographical, initial method is to collect the seeds of the sentiment words and their orientation to find their antonyms and synonyms to expand their set. When no more words left this iteration stops [7]. This approach is unable to extract opinions with domain-specific orientations which are counted as its disadvantage.
- **Corpus-Based Approach:** It can help to find domain-specific opinion words and their orientation if a corpus from only the specific domain is used in the discovery process. It resolves the problem of dictionary-based approach and also explore the idea of intra-sentential and inter-sentential sentiment. Despite showing domain dependent words it showed the same word having different contexts even in the same domain.

Machine Learning Approach: Machine learning approach has been used to solve the sentence classifications problems that totally based on the algorithms. It trains a text classifier on a human-labeled training dataset [8]. Two approaches were used that is supervised learning approach and unsupervised learning approach. Bulk of labeled training document is known as supervised learning and further it is of two types: Naïve Baye's algorithm and Maximum Entropy Classifier.

- **Unsupervised learning:** This method depends upon clustering. They are not provided with correct information about the target and also not contain any category.
- **Supervised learning:** In this learning, labels are provided to the model during the process as it is based on labeled dataset. In order to obtain desired output, these labeled dataset are trained in order to tackle situation when they encounter during decision making.

II. LITERATURE REVIEW

Shreya Ahuja, et.al (2017) presented a platform using which people from different locations can share their thoughts and view on a single platform known as social media [9]. This platform can be anything from social media platform such as twitter, facebook, Instagram, blogs and many more. Nowadays, the user on the twitter is growing rapidly approximately 500 million. All the information posted on the twitter is very useful which can be extracted easily using opinion mining. Using this platform, it becomes easy to extract opinion about any news or launch of a product. It is also beneficial in determining the emotion, opinion, attitude by discovering from a natural text. This analysis help in differentiating tweets into positive and negative sentiment.

Jaishree Ranganathan, et.al (2017) presented a new optimized system which provides the speed and efficiency is the main objective of this paper. They implemented the Specific Action Rule discovery based on Grabbing strategy (SARGS) algorithm in order to generate meta-actions. In this paper, they performed the comparison of meta-actions generating algorithmic which is implemented in Apache Spark driven system and conventional Hadoop driven system using the Twitter social networking data. After comparison, they evaluated the results. On the basis of the quantity of Sentimental Analysis of social networking data,

they performed analysis. It also tests the total time taken by both the systems and their subcomponents for the data processing [10]. As per obtained results, it is concluded that as compared to Hadoop more computational time is utilized by Spark system and for the implementation of meta-action generation methods they utilizes the MapReduce.

Huma Parveen, et.al (2016) presented with the advent in the technology and its growth, the use of social networking sites has been increased rapidly. This social networking is the reason in the generation of the large amount of data. With the help of these microblogging websites, it becomes easy for millions of people to express their views and opinion easily. User expressed their views and opinions on micro-blogging website such as Twitter for the extraction of sentiments of users [11]. In this paper, sentiment analysis was performed on tweets in order to predict opinions on business intelligence. The available dataset on the twitter website of the movie data has been processed using Hadoop Framework in terms of reviews, feedback, and comments. On the basis of the positive, negative and neutral sentiments, they analyze the results of sentiment analysis on Twitter data.

Mitali Desai, et.al (2016) explained it becomes easy for user to express their views, opinions, trends and issues on any topic with the help of World Wide Web which emerged as optimal way [12]. In order to make decisions, these platforms have been widely utilized in various fields such as advertising, political polls, scientific surveys, market prediction and business intelligence. With the help of different mediums such as internet large amount of user-generated data is present. The issue related to the mining of the sentiments from online available data and this data is categorized further in positive, negative and neutral. First, the present highly unstructured data on Twitter is classified in this paper using the sentiment analysis. Second, they evaluate various sentiment analysis in order to carry out sentiment analysis on Twitter data.

Rincy Jose, et.al (2016) presented the use of analyzer in order to determine the sentiments by implementing a real-time Twitter sentiment analyzer using classifier ensemble approach. The machine learning classifiers was combined with lexicon based classifier in this paper. They discussed the advantages of these three classifiers such as SentiWordNet classifier, naive Bayes classifier and hidden Markov model classifier. This was done to classify the political data accurately [13]. Hence, for determining the political sentiment from real-time tweets, they developed a novel accurate sentiment classifier in this paper. The obtained results from comparison were done by plotting graphs on the basis of the obtained results from the sentiment analysis. This classifier has been utilized to compare two new released movies by the sentiment analysis on Twitter data. As per performed experiments, they concluded that proposed classifier improves the accuracy as compared to other classifiers.

Pulkit Garg, et.al, (2017) presented with the increase in the number of people coming online for online sharing social media has become a platform. In this paper, they studied post-terror attack tweets by extracting it from Twitter [14]. The flow data posted on Twitter is used to study factors like last retweet, number of retweets and number of favorites. Maximum number of retweets indicates maximum reach. It creates widespread reaction on the social media. Governments are concentrating on digitalizing the whole nation. Due to increase in number of people, huge data is generated. Author discussed the Uri Terror Attacks that show more negative tweets tend to survive as compare to positive tweets, although their amount is low. It will lead to public unrest if people start targeting a community and provide negative information. Misleading information, the trends of retweets and number of favorites are the future scope to study its flow and survival [14].

Zahra Rezaei et.al, (2017) presented large amount of data is generated on daily basis due to the widespread use of social networking sites by which they share their opinions and views using microblogging website. Due to increase in the twitter messages continuously and arriving at very high speed, therefore, prediction of sentiments can be done using algorithms in real-time and under limited time. In the mining the data, Hoeffding tree algorithm has been utilized as the powerful tool. Also an appropriate approach in order to obtain splitting attribute as data in twitter follows data stream model. They utilize the filtering and wrappers techniques in order to enhance the performance of the methods [15]. For the implementation of an algorithm, a twitter language common preprocessing task was used. Hoeffding tree has the high process time, in order to eliminate this Mcdiamid tree algorithm is used. McDiarmid tree performance is better than Hoeffding tree but the accuracy of both are same. Due to large amount of twitter data, minimum processing time is required for the sentimental analysis therefore, McDiarmid Tree is better than Hoeffding tree.

Vishal A. Kharde, et.al (2016) presented the process of the opinion mining using machine learning and lexicon-based approaches, they done a survey and comparative study on the existing techniques. For this purpose they utilize the cross-domain and cross-lingual methods together for the evaluation of metrics [16]. As per obtained results, it is predicted that highest accuracy is present in SVM and naive Bayes classifier subpart of machine learning methods. Due to this accuracy, they are regarded as the baseline learning methods. In some cases, lexicon-based methods are very effective as it requires human-labeled document efforts. Hence, the accuracy of the obtained result is more if the data is cleaner. Better sentiment accuracy is provided by the bigram model as compared to existing methods. Therefore, in order to improve the accuracy of sentiment classification and adaptive capacity to variety of domains and different languages, they combine the machine learning method into opinion lexicon method.

Table 1 Table of Comparison

Author's Name	Year	Description	Outcomes
Shreya Ahuja, Gaurav Dubey	2017	A platform is presented using which people from different locations can share their thoughts and view on a single platform known as social media. This platform can be anything from social media platform such as twitter, facebook, Instagram, blogs and many more.	It becomes easy to extract opinion about any news or launch of a product. It is also beneficial in determining the emotion, opinion, attitude by discovering from a natural text. This analysis help in differentiating tweets into positive and negative sentiment.
Jaishree Ranganathan, Allen S. Irudayaraj, Angelina A. Tzacheva	2017	A new optimized system presented which provides the speed and efficiency is the main objective of this paper. They implemented the Specific Action Rule discovery based on Grabbing strategy (SARGS) algorithm in order to generate meta-actions.	As per obtained results, it is concluded that as compared to Hadoop more computational time is utilized by Spark system and for the implementation of meta-action generation methods they utilizes the MapReduce.
Huma Parveen, Prof. Shikha Pandey	2016	Presented with the advent in the technology and its growth, the use of social networking sites has been increased rapidly. These social networking are the reason in the generation of the large amount of data.	On the basis of the positive, negative and neutral sentiments, they analyze the results of sentiment analysis on Twitter data.
Mitali Desai, Mayuri A. Mehta	2016	Explained it becomes easy for user to express their views, opinions, trends and issues on any topic with the help of World Wide Web which emerged as optimal way.	First, the present highly unstructured data on Twitter is classified in this paper using the sentiment analysis. Second, they evaluate various sentiment analysis in order to carry out sentiment analysis on Twitter data.
Rincy Jose, Varghese S Choorali	2016	The use of analyzer in order to determine the sentiments by implementing a real-time Twitter sentiment analyzer using classifier ensemble approach. The machine learning classifiers was combined with lexicon based classifier in this paper.	As per performed experiments, they concluded that proposed classifier improves the accuracy as compared to other classifiers.
Pulkit Garg, Himanshu Garg, Virender Ranga	2017	In this paper, they studied post-terror attack tweets by extracting it from twitter. The flow data posted on Twitter is used to study factors like last retweet, number of retweets and number of favorites.	It will lead to public unrest if people start targeting a community and provide negative information. Misleading information, the trends of retweets and number of favorites are the future scope to study its flow and survival.
Zahra Rezaei, Mehrdad Jalali	2017	In the mining the data, Hoeffding tree algorithm has been utilized as the powerful tool. Also an appropriate approach in order to obtain splitting attribute as data in twitter follows data stream model	McDiarmid tree performance is better than Hoeffding tree but the accuracy of both are same. Due to large amount of twitter data, minimum processing time is required for the sentimental analysis therefore, McDiarmid Tree is better than Hoeffding tree.
Vishal A. Kharde, S.S. Sonawane	2016	Presented the process of the opinion mining using machine learning and lexicon-based approaches, they done a survey and comparative study on the existing techniques. For this purpose they utilize the cross-domain and cross-lingual methods together for the evaluation of metrics.	Better sentiment accuracy is provided by the bigram model as compared to existing methods. Therefore, in order to improve the accuracy of sentiment classification and adaptive capacity to variety of domains and different languages, they combine the machine learning method into opinion lexicon method.

III. CONCLUSION

In this work , it is concluded that sentiment analysis is the technique which is applied to analyze sentiments of the input data. The classification technique are generally applied for the sentiment analysis. In this work, various sentiment analysis techniques are reviewed and analyzed in terms of accuracy and execution time. In future, novel technique will designed for the sentiment analysis.

IV. REFERENCES

[1] R. Parikh and M. Movassate, "Sentiment Analysis of User- GeneratedTwitter Updates using Various Classi_cation Techniques",CS224N Final Report, 2009.

[2] Go, R. Bhayani, L.Huang. "Twitter Sentiment classification using Distant Supervision". Stanford University, Technical Paper,2009.

[3] L. Barbosa, J. Feng. "Robust Sentiment Detection on Twitter from Biased and Noisy Data". COLING 2010: Poster Volume,pp. 36-44.

[4] Bifet and E. Frank, "Sentiment Knowledge Discovery inTwitter Streaming Data", In Proceedings of the 13th International Conference on Discovery Science, Berlin, Germany: Springer,2010, pp. 1-15.

[5] Agarwal, B. Xie, I. Vovsha, O. Rambow, R. Passonneau, "Sentiment Analysis of Twitter Data", In Proceedings of the ACL 2011Workshop on Languages in Social Media,2011 , pp. 30-38.

[6] Dmitry Davidov, Ari Rappoport." Enhanced Sentiment Learning Using Twitter Hashtags and Smileys". Coling 2010: Poster Volume pages 241{249, Beijing, August 2010.

[7] A. Sarlan, C. Nadam and S. Basri, "Twitter Sentiment Analysis", in Int. Conf. on Information Technology and Multimedia, 2014, pp. 213-216.

[8] B. Gokulakrishnan, P. Plavnathan, R. Thiruchittampalam, A. Perera and N. Prasath, "Opinion Mining and Sentiment Analysis on aTwitter Data Stream", in Int. Conf. on Advances in ICT for Engineering Regions, 2012, pp. 182-188.

[9] Shreya Ahuja, Gaurav Dubey, "2017 2nd International Conference on Telecommunication and Networks", IEEE, 2017.

[10] Jaishree Ranganathan, Allen S. Irudayaraj, Angelina A. Tzacheva, "Action Rules for Sentiment Analysis on Twitter Data using Spark", 2017 IEEE International Conference on Data Mining Workshops.

[11] Huma Parveen, Prof. Shikha Pandey, "Sentiment Analysis on Twitter Data-set using Naïve Bayes Algorithm", IEEE, 2016.

[12] Mitali Desai, Mayuri A. Mehta, "Techniques for Sentiment Analysis of Twitter Data: A Comprehensive Survey", International Conference on Computing, Communication and Automation, 2016.

[13] Rincy Jose, Varghese S Choorali, "Prediction of Election Result by Enhanced Sentiment Analysis on Twitter Data using Classifier Ensemble Approach", IEEE, 2016.

[14] Pulkit Garg, Himanshu Garg, Virender Ranga, "Sentiment Analysis of the Uri Terror Attack Using Twitter" Department of Computer Engineering National Institute of Technology Kurukshetra , vol. 8, pp. 1-4, 2017.

[15] Zahra Rezaei, Mehrdad Jalali, "Sentiment Analysis on Twitter using McDiarmid Tree Algorithm" 7th International Conference on Computer and Knowledge Engineering, vol. 4, pp. 1-4, 2017.

[16] Vishal A. Kharde, S.S. Sonawane, "Sentiment Analysis of Twitter Data: A Survey of Techniques", International Journal of Computer Applications, Volume 139 – No.11, April 2016.