A REVIEW IN COMPARATIVE ANALYSIS OF PATTERN BASED APPROACH FOR MULTI CLASS SENTIMENT ANALYSIS IN TWITTER USING KNN CLASSIFIER

[1]Akashdeep Dhiman, [2]Dr. Dinesh Kumar

[1]Research Scholar, [2]Associate Professor & Head,

Department of CSE

Sri Sai College of Engineering and Technology, Badhani (Pathankot)

Abstract: Sentiment analysis and conclusion mining in social netwrok show these days a hot debated issue of research. In any case, the majority of the best in class works and looks into on the programmed sentiment analysis and supposition mining of writings gathered from Social Netwrok and microblogging sites are arranged toward the twofold grouping (i.e., order into "positive" and "negative") or the ternary characterization (i.e., grouping into "positive," "negative," and "nonpartisan") of writings.

Keywords: Twitter, Sentimental Analysis, KNN

I. Introduction to Sentiment Analysis

In order to mainly, determine whether a product is satisfactory for the users before it is delivered to them, the sentiment analysis (SA) has been developed. On the basis of the requirements of the users, the markets and industries develop their products and services. The factual data is processed, searched or analyzed with the help of textual information retrieval methods. The subjective properties of the components can be presented on the basis of various textual contents within the actualities [1]. The base of sentiment analysis (SA) includes opinions, attitudes, emotions, appraisals and so on. In order to develop new applications, various challenges have been faced while applying these techniques. With the utilization of Natural Language Processing (NLP), various tweets, speech or text available on sources can be processed in sentiment analysis. There are three broader categories into which the sentiments are classified. They are "positive" or "negative" or "neutral".

II. Challenges being faced in Sentiment Analysis

There are various challenges being faced when sentiment analysis is applied in various applications. Some of these challenges are [2]:

- This technique cannot differentiate amongst the fake or spam reviews that are present within the sources. There is no comparison made amongst the quality of reviews provided. The selection of outliers and identification of reputation of reviewer is also not provided here.

- Collaborative filtering which recognizes important key points and provides innovative ideas is limited within this technique.

- The automated content analysis and behavioral analysis are combined together to form a filter bubble. This helps in providing selection of related opinions and content in a very efficient manner. This is however a risk which is available in sentiment analysis methods and results in causing various issues [3].

- The various organizations and government can afford the presence of opinion mining software which is however not available to the other general people. This means that even when people have the democracy of producing and publicizing their content, they cannot have the access to analyze it.

- The behavior as well implicit data are integrated with opinion which helps in validating and providing analysis of the data that cannot be expressed in general form.

Usability and user-friendliness of the tools are the permanent requirements which need to be maintained in these systems.

III. Literature Review

S.NO	RESEARCHER	PAPER NAME	JOURNAL	FINDINGS
1.	Mondher Bouazizi, Tomoaki Ohtsuki	A Pattern-Based Approach for Multi-Class Sentiment Analysis in Twitter.	IEEE	On the basis of experimental results achieved it is seen that the multi-class classification is achieved with 60.2% of accuracy. Thus, by providing various comparisons it is seen that the performance of proposed technique is better than the existing approaches.
2.	Ankit Kumar Soni	Multi-Lingual Sentiment Analysis of twitter data by using classification algorithms.	IEEE	In this paper, Naïve Bayes and Maximum Entropy classifiers are combined to generate one algorithm. Amongst various algorithms, the results are compared which can help in analyzing the performance of various algorithms amongst each other and show which has provide to be better. It is seen through the results achieved that the proposed technique has provided better results in comparison to other existing approaches.
3.	Aldo Hernández, Victor Sanchez, Gabriel Sánchez, Héctor Pérez	Security Attack Prediction Based on User Sentiment Analysis of Twitter Data	IEEE	Various experiments are conducted and it is seen that the proposed technique can help in providing an estimate of number of actual attacks and negative sentiments provided by the users in the form of tweets.

4.	Venkata Sasank Pagolu	Sentiment Analysis of Twitter Data for Predicting Stock Market Movements	IEEE	On the basis of various experiential results achieved at the end, it is seen that the proposed technique provides better evaluation results in comparison to existing techniques.

IV. Objectives

Following are the various objectives of this research:-

❖ To study and analyze various sentiment analysis techniques for twitter data.

❖ To propose improvement in sentiment analysis technique with KNN classification.

❖ Implement proposed technique and analyze results in terms of certain parameters.

V. Research Methodology

The proposed work is based on data classification and feature extraction for sentiment analysis. In the base paper, N-gram algorithm is being used for feature extraction and correlation factor is used for classification. The current correlation factor does not classify the features which are approximately equal. This leads to reduction in accuracy of classification and increase execution time. In this work, nearest neighbor classifier is applied which will calculate similarity using Euclidian distance and classifies the features which approximately equal. This approach increase accuracy, reduce execution time and reduce fault detection rate.
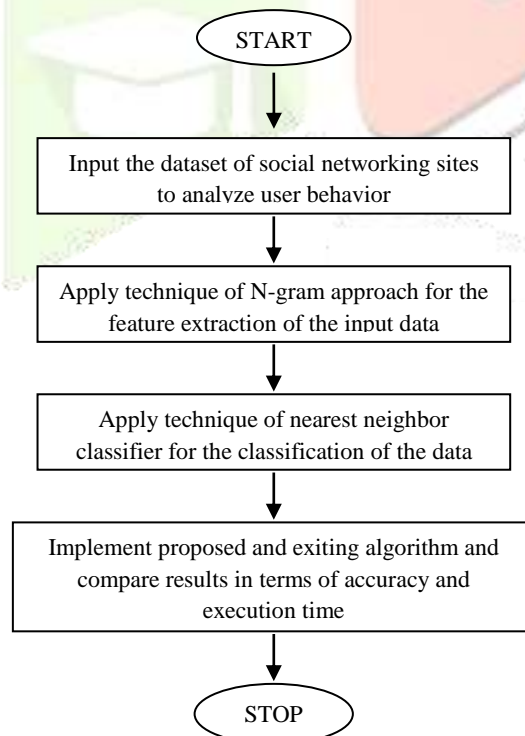
Fig 1: Proposed Flowchart

VI. Conclusion

In this paper, we have review a comparative analysis of pattern based approach for multi class sentiment analysis in twitter. How t therefore, in a implementation, we will use the results obtained for ternary classification to classify tweets into *''Positive''*, *''Negative''* and *''Neutral''*. The classified sentimental tweets (i.e., which have been classified as *''Positive''* or *''Negative''*) will then be given scores for the corresponding sentiment subclasses.

VII. References

[1] Feng S., Wang D., Yu G., Yang C., Yang N., "Sentiment Clustering: A Novel Method to Explore in the Blogosphere" Advances in Data and Web Management, vol. 5446, pp. 332-344 , 2009

[2] Andreevskaia A. , Bergler S. , "Mining WordNet For a Fuzzy Sentiment: Sentiment Tag Extraction From WordNet Glosses", Proceedings EACL-06, the 11rd Conference of the European Chapter of the Association for Computational Linguistics, pp. 209–216 ,2006

[3] Fei, Geli, et al. "A Dictionary-Based Approach to Identifying Aspects Implied by Adjectives for Opinion Mining." 24th International Conference on Computational Linguistics. 2012.

[4] Mondher Bouazizi and Tomoaki Ohtsuki, "A Pattern-Based Approach for Multi-Class Sentiment Analysis in Twitter", 2017, IEEE

[5] Ankit Kumar Soni, "Multi-Lingual Sentiment Analysis of twitter data by using classification algorithms", 2017, IEEE

[6] Aldo Hernández, Victor Sanchez, Gabriel Sánchez, Héctor Pérez, "Security Attack Prediction Based on User Sentiment Analysis of Twitter Data", 2016, IEEE

[7] Venkata Sasank Pagolu, Kamal Nayan Reddy Challa, Ganapati Panda, Babita Majhi, "Sentiment Analysis of Twitter Data for Predicting Stock Market Movements", 2016, IEEE