Sentimental Analysis on social media

SHALIN HARIRAMANI[1], SARIKA WADHWANI[2], ADVIN MANAHAR[3]

[1, 2] *Department of Computer Science and Engineering, Amity School of Engineering and Technology, Amity University Chhattisgarh*

[3] *Assistant Professor, Department of Computer Science and Engineering, Amity School of Engineering and Technology, Amity University Chhattisgarh*

Abstract— social media are the fundamental assets to accumulate data about individuals' perspective and opinions towards various topics as they go through hours day to day on social medias and share their perspective. In this specialized paper, we show the utilization of sentimental analysis and how to associate with Twitter. In this paper,using Lexicon-Based Approach", we are fundamentally breaking down the feelings of users involving Twitter as an online stage where users post remarks in regards to their perspectives about trending topics and which can be valuable for business associations for fulfilling their client's need, for legislators to serve better to the general population and do better campaigning etc. This is done by fetching data in the real time environment. With the help of sentiment analysis (Opinion Mining) users will also be able to take better decisions and make better strategies for future development of the organization and with that they will also be able to compare their previous performance with their current performance. For bigger organizations it will also be useful for quick review for a new product or a new update. We will be applying Lexicon-based approach which is a dictionary or bag of words containing various sentiments like anger, happiness, fear, joy, etc The software which we have utilized for the execution is R-Studio which is an integral asset for analysing the information with numerous predefined techniques and libraries.

I. INTRODUCTION

These days, the time of Internet has had an impact on the way individuals express their perspectives, conclusions. It is currently essentially done through blog entries, online discussions, item survey sites, social media ,and so on. These days, a huge number of individuals are utilizing social network destinations like Facebook, Twitter, Google Plus, and so forth to express their feelings, assessment and offer perspectives about their day to day routines. Through the web-based networks, we get an intuitive media where shoppers illuminate and impact others through discussions. Web-based entertainment is creating an enormous volume of feeling rich information as tweets, announcements, blog posts, remarks, surveys, and so on. The amount of content generated by users is too vast for a normal user to analyze. So there is a need to automate this, various sentiment analysis techniques are widely used.

Sentimental analysis is fundamentally a method of distinguishing and investigating public's feelings, contemplations, thoughts towards any individual, spot or item. It educates us regardless of whether the data regarding the item is palatable before they get it. Clients can utilize this strategy to investigate items/administrations prior to making a buy. Advertisers and firms and utilize this to investigate popular assessment of their organization and items or to break down consumer loyalty. Along these lines it is a course of investigation of feelings from text, remarks, discourse, information base sources, and so forth following which assessment are grouped into classes like good, pessimistic or impartial.

Data mining has been prevalently utilized in numerous areas like showcasing, organizations, banking, finance, client connections, and so forth. In this venture we analyse the surveys about the hashtags related to trending topics as well as anticipate the achievement or disappointment of the topics which benefit different partners like entertainers, chiefs, and so forth as these expectations assist them with taking more educated choices so they can apply those choices further.

II. LITERATURE SURVEY

[1] In 2020, Kumar et al. have presented a hybrid deep learning approach named ConVNet-SVMBoVW that dealt with the real-time data for predicting the fine-grained sentiment. In order to measure the hybrid polarity, an aggregation model was developed. Moreover, SVM was used for training the BoVW to forecast the sentiment of visual content. Finally, it was concluded that the suggested ConvNet-SVMBoVW was outperformed by the conventional models.

[2] In 2020, Park et al. have designed a deep learning approach for improving performance. In order to improve the performance, two questions have come into picture. The content attention was required for being sophisticated for merging many attention results non-linearly and assumes the whole context for mentioning the complex sentences. The test results have shown that the proposed model was attained as the best performance.

[3] In 2020, Xu et al. have introduced a NB method for multi-domain and large-scale E-commerce platform product review classification of sentiment. Consequently, the parameter evaluation method was extended in NB for continuous learning fashion. Later, for fine-tuning the learned distribution on the basis of three types of assumptions, many ways were introduced for acquiring the best performance. The results have shown that the suggested model has high accuracy in Amazon product and movie review sentiment datasets.

[4] In 2020, Hassonah et al. have recommended hybrid machine learning algorithm for improving the sentiment analysis, because a classification approach was built on the basis of "Positive, Negative, and Neutral" classes with SVM classifier, at the same time two feature selection methods were merged by the MVO and Relief models. Moreover, Twitter data was employed for evaluating the proposed model. The experimental results indicated that the suggested technique was performing well than conventional techniques

[5] In 2019, Saad and Yang have aimed for giving a complete tweet sentiment analysis on the basis of ordinal regression with machine learning algorithms. The suggested model included pre-processing tweets as first step and with the feature extraction model, an effective feature was generated. The methods such as SVR, RF, Multinomial logistic regression (SoftMax), and DTs were employed for classifying the sentiment analysis. Moreover, twitter dataset was used for experimenting the suggested model. The test results have shown that the suggested model has attained the best accuracy, and also DTs were performed well when compared over other methods.

[6] In 2019, Afzaal et al. have recommended a novel approach of aspect-based sentiment classification, which recognized the features in a precise manner and attained the best classification accuracy. Moreover, the scheme was developed as a mobile application, which assisted the tourists in identifying the best hotel in the town, and the proposed model was analyzed using the real-world data sets. The results have shown that the presented model was effective in both recognition as well as classification.

[7] In 2019, Feizollah et al have concentrated on tweets related to two halal products such as halal cosmetics and halal tourism. By utilizing Twitter search function, Twitter information was extracted, and a new model was employed for data filtering. Later, with the help of deep learning models, a test was performed for computing and evaluating the tweets. Moreover, for enhancing the accuracy and building prediction methods, RNN, CNN, and LSTM were employed. From the outcomes, it was seemed that the combination of LSTM and CNN attained the best accuracy

[8] In 2018, Fang et al. have suggested multi-strategy sentiment analysis models using the semantic fuzziness for resolving the issues. The outcomes have demonstrated that the proposed model has attained high efficiency.

[9] In 2018, Mukhtar et al.have performed the sentiment analysis to the Urdu blogs attained from several domain with Supervised Machine learning and Lexicon-based models. In Lexicon-based models, a well-performing Urdu sentiment analyzer and an Urdu Sentiment

Lexicons were employed, whereas, in Supervised Machine learning algorithm, DT, KNN, and SVM were employed. The data were combined from the two soruces for performing the best sentiment analysis. Based on the tests conducted, the outcomes were shown that the Lexicon-based model was superior to the supervised machine learning algorithm.

[10] In 2018, Smadi et alhave proposed existing models on the basis of supervised machine learning algorithms for specifying the defects of feature-based sentiment analysis of Arabic hotel's review. Moreover, SVM and Deep RNN were developed and trained with word, lexical, morphological, semantic, and syntactic features. The reference dataset of Arabic hotel's review dataset was used for evaluating the proposed model. The outcomes have shown that SVM was performing well when compared over RNN

Problem Statement -

A lot of times we cannot discern the sentiment associated with the most trending hashtags that majorly affect people's lives and the image associated with a certain topic or person. During debates or media trials, often times people take to twitter to express their emotions and it is usually ambiguous with multiple reactions and perspectives. It becomes difficult to understand what the major emotion associated with the hashtag is.

Proposed Solution-

This paper helps us to easily predict what sentiment is most associated with the most trending hashtags. The solution to the given problem makes use of data taken from twitter which is then used via R libraries to analyze their sentiment based on the words used in each tweet. It is then classified based on the most expressed sentiment used with that hashtag and we represent then in 3 readable forms of a bar graph that depicts the emotions, and two word clouds out of which one shows the most associated words with the hashtag and the other one classifying the words in negative or positive.

III. METHODOLOGY

Lexicon based opinion mining is technique which uses text mining methodology to analyse data being the simplest way to analyse text format data. This approach uses R libraries such as tidytext, tidyverse, wordcloud, etc. These libraries provide us with various predefined functions which can be easily applied on unstructured data. This methodology divides the sentences accordingly and classifies them into positive and negative emotions like anger, fear, happiness, trust, etc.

Application of a lexicon is one of the two main approaches to sentiment analysis. It involves calculating the sentiment from the semantic orientation of word or phrases that occur in a text. In lexicon-based approach a piece of text message is represented as a bag of words, following this representation of the message, sentiment values from the dictionary are assigned to all positive and negative words or phrases within the message, combining function, such as sum or average, is applied in order to make the final prediction regarding the overall sentiment for the message.

Advantages of Lexicon-based approach is that we can analyse text data in no time and disadvantages of this approach is that it cannot specifically analyse the data in terms of multiple attributes.

The R libraries are than used to plot different types of graphs and word clouds to show the analysis result. The Graphs and Word cloud are drawn based on the sentiment or lexicon score of the words.
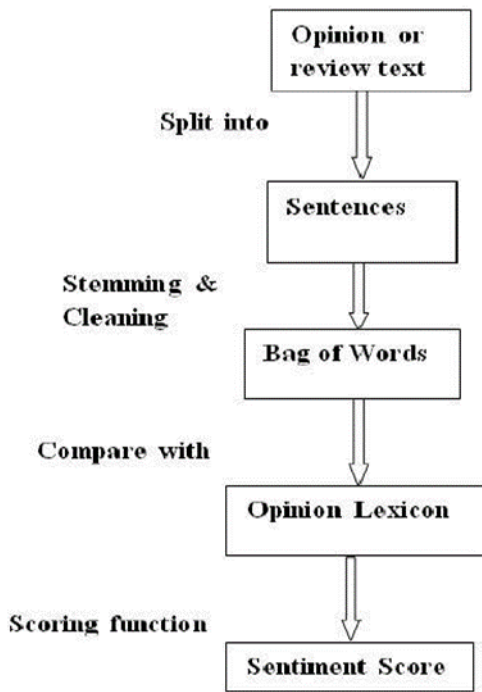
Fig1: Working of Lexicon-based approach

IV. RESULT ANALYSIS

The analysis of the hashtag "Explore India" gives us the following results. From the graphs we can infer that the most associated emotions are positivity, joy and anticipation and we also deduce that anger, fear and disgust are the least related sentiments. We also find that the words we associate the most are "journey", "travel" and "youtube" which is an interesting finding because we see how people usually go to youtube when we talk about exploring India, a lot less than travel websites like tripadvisor. There are a lot of common positive words that are linked to the given hashtag but surprisingly, the negative words associated are "freak" and "cave" which assumes that most people see caves in a negative light, most probably associated with fear or disgust.
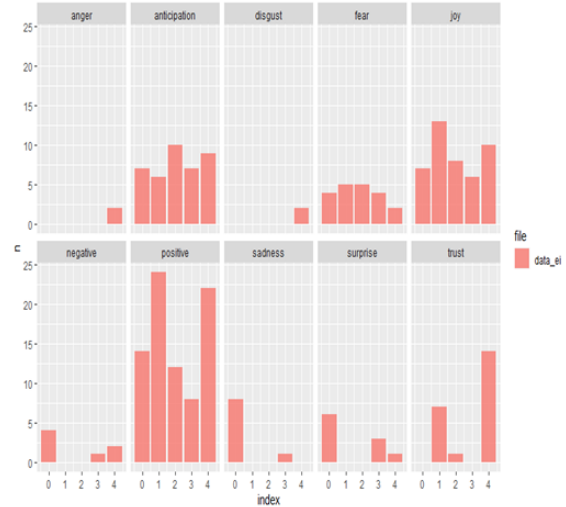
O/P 1: ggplot obtained for #ExploreInida
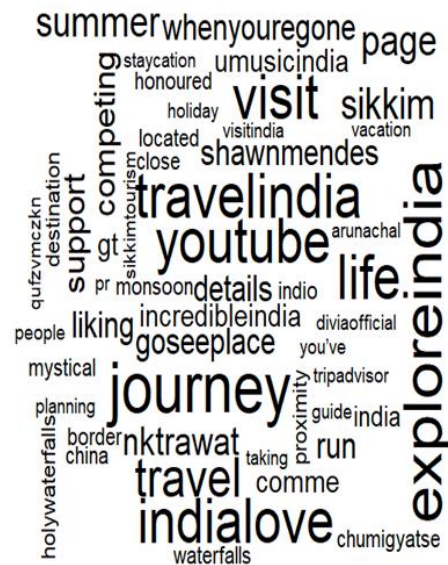
O/P 2: Simple Word Cloud

O/P 3: Positive-Negative Word Cloud

V. CONCLUSION AND FUTURE WORK

our paper is based on analysis of hashtag of trending topics in twitter, we are successful at bifercating and filtering the hashtag and the results are successfully shown undergglot, word cloud,positive negative word cloud.

The task of sentiment analysis, especially in the domain of micro-blogging, is still in the developing stage and far from complete. We propose a couple of ideas which we feel are worth exploring in the future and may result in further improved performance. Right now we have worked with only the very simplest unigram models; we can improve those models by adding extra information like closeness of the word with a negation word. The closer the negation word is to the unigram word whose prior polarity is to be calculated, the more it should affect the polarity. For example, if thenegation is right next to the word, it may simply reverse the polarity of that word and farther the negation is from the word the more minimized ifs effect should be.

One of the major future scope is that it can be performed under multiple social networking sites for better understanding of topics and accurate results.

REFERENCES

[1] Boguslavsky, I. (2017). Semantic Descriptions for a Text Understanding System. In Computational Linguistics and Intellectual Technologies. Papers from the Annual International Conference "Dialogue" (2017) (pp. 14-28).

[2] Statista, 2017, https://www.statista.com/statistics/282087/number-ofmonthly-active-twitter-users/

[3] TextBlob, 2017, https://textblob.readthedocs.io/en/dev/

[4] Rosenthal, S., Farra, N., &Nakov, P. (2017). SemEval-2017 task 4: Sentiment analysis in Twitter. In Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017) (pp. 502-518).

[5] Poria, S., Cambria, E., &Gelbukh, A. (2015). Deep convolutional neural network textual features and multiple kernel learning for utterance-level multimodal sentiment analysis. In Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing (pp. 2539- 2544).

[6] Dos Santos, C. N., &Gatti, M. (2014, August). Deep Convolutional Neural Networks for Sentiment Analysis of Short Texts

[7] Kiritchenko, S., Zhu, X., & Mohammad, S. M. (2014). Sentiment analysis of short informal texts. Journal of Artificial Intelligence Research, 50, 723-762.

[8] Balahur, A., Steinberger, R., Kabadjov, M., Zavarella, V., Van Der Goot, E., Halkia, M., ... &Belyaeva, J. (2013). Sentiment analysis in the news. arXiv preprint arXiv:1309.6202.

[9] Liu, B. (2012). Sentiment analysis and opinion mining. Synthesis lectures on human language technologies, 5(1), 1-167.

[10] Wang, H., Can, D., Kazemzadeh, A., Bar, F., & Narayanan, S. (2012, July). A system for real-time twitter sentiment analysis of 2012 us presidential election cycle. In Proceedings of the ACL 2012 System Demonstrations (pp. 115-120). Association for Computational Linguistics.

[11] Agarwal, A., Xie, B., Vovsha, I., Rambow, O., &Passonneau, R. (2011, June). Sentiment analysis of twitter data. In Proceedings of the workshop on languages in social media (pp. 30-38). Association for Computational Linguistics.

[12] Scott, J. (2011). Social network analysis: developments, advances, and prospects. Social network analysis and mining, 1(1), 21-26.

[13] Pak, A., &Paroubek, P. (2010, May). Twitter as a corpus for sentiment analysis and opinion mining. In LREc (Vol. 10, No. 2010).

[14] Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. Foundations and Trends® in Information Retrieval, 2(1–2), 1- 135.

[15] Wilson, T., Wiebe, J., & Hoffmann, P. (2005, October). Recognizing contextual polarity in phrase-level sentiment analysis. In Proceedings of the conference on human language technology and empirical methods in natural language processing(pp. 347-354). Association for Computational Linguistics.