IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

SENTIMENT ANALYSIS BASED TOP ORGANIZATIONS RATING USING MACHINE LEARNING

Arun Kumar Gautam[1], Zainab Khan[2]

[1]M. Tech, Dept. of CSE, B N College of Engineering & Technology, (AKTU), Lucknow, India

[2]Assistant Professors, Dept. of CSE, B N College of Engineering & Technology, (AKTU), Lucknow, India

Abstract— In the digital age, user-generated content such as reviews, feedback, and social media posts significantly influences organizational reputation and customer decisions. This research proposes a machine learning-based sentiment analysis framework to evaluate and rank top organizations based on public sentiment. By leveraging Natural Language Processing (NLP) techniques, the system extracts and analyzes textual data from diverse online sources to classify sentiments as positive, negative, or neutral. Various machine learning models, including Support Vector Machines (SVM), Random Forest, and Naïve Bayes, are employed and compared for sentiment classification accuracy. The results are further aggregated to compute sentiment-based scores for organizations, enabling a fair and data-driven ranking. The proposed approach provides insights into customer perception, enhances decision-making for stakeholders, and offers a scalable solution for real-time reputation monitoring.

Keyword: Sentiment Analysis, Machine Learning, Organization Ranking, Natural Language Processing, Reputation Management, Text Classification.

I. INTRODUCTION

In today's data-driven world, the opinions and sentiments expressed by individuals on digital platforms play a crucial role in shaping the public image and credibility of organizations. With the widespread use of social media, review sites, and online forums, organizations are constantly being evaluated by users based on their products, services, and overall experiences. This shift has created a need for intelligent systems that can automatically interpret and quantify these opinions to assess organizational performance and public perception.

Sentiment analysis, also known as opinion mining, is a Natural Language Processing (NLP) technique that analyzes textual data to determine the sentiment expressed—positive, negative, or neutral. When combined with machine learning algorithms, sentiment analysis becomes a powerful tool for extracting meaningful insights from large volumes of unstructured data. In the context of organizational evaluation, this technology enables the identification of trends, strengths, and areas for improvement by analyzing user feedback.

This study presents a sentiment analysis-based approach to rate and rank top organizations using machine learning techniques. By collecting user-generated reviews and comments from various platforms, preprocessing the data, and applying supervised learning models such as Support Vector Machine (SVM), Naïve Bayes, and Random Forest, the system can effectively classify sentiments and derive a comprehensive rating system. This rating not only reflects public opinion but also provides a transparent, data-informed basis for comparing organizational performance.

The proposed system aims to assist customers in making informed decisions, help organizations understand their market perception, and support researchers and policymakers with a reliable sentiment-based evaluation framework. As businesses increasingly rely on digital feedback, such sentiment-driven analytical tools are becoming essential for competitive analysis, strategic planning, and reputation management.

The rest of the work is sorted out as pursues: Section 2 exhibits the primary ideas identified with Sentiment Analysis examined in a few late works. Segment 3 demonstrates the data description. Area 4 exhibits a few techniques that have been performed so as to think about the Web administrations remarked on the past segment, and also the outcomes got. At last, Section 5 points out a few ends.

II. LITERATURE SURVEY

Sentiment analysis has emerged as a critical area in the field of Natural Language Processing (NLP) and machine learning due to the exponential growth of opinionated content on the internet. Researchers have extensively explored various techniques for extracting sentiment from textual data and applying it to domains such as product reviews, political analysis, and organizational reputation.

Pang et al. (2002) laid the groundwork for sentiment classification using machine learning by comparing Naïve Bayes, Maximum Entropy, and Support Vector Machines (SVM) on movie reviews, establishing that machine learning outperforms traditional rule-based approaches [1]. Since then, the application of sentiment analysis has diversified. Liu (2012) provided a comprehensive survey of sentiment analysis, highlighting the importance of opinion mining for understanding consumer behavior and market trends [2].

In the context of organizational analysis, Sarlan et al. (2014) demonstrated the application of sentiment analysis on Twitter data to evaluate public perception of events and institutions, using lexical approaches and machine learning techniques [3]. Similarly, Pak and Paroubek (2010) utilized Twitter corpora to train classifiers, showing that social media can be a rich source for sentiment-based insights [4].

More recent studies have incorporated deep learning for sentiment analysis. For instance, Zhang et al. (2018) implemented Convolutional Neural Networks (CNNs) for sentiment classification and reported improved accuracy compared to traditional models [5]. However, for scalable and interpretable systems like organizational ranking, classical machine learning models are still widely preferred due to their lower computational cost and ease of deployment.

Research by Medhat et al. (2014) reviewed various sentiment classification methods and noted that hybrid approaches combining lexicon and machine learning techniques yield better performance [6]. On the other hand, Tripathy et al. (2015) compared the performance of SVM, Naïve Bayes, and Logistic Regression on movie reviews and found SVM to be superior in handling high-dimensional feature spaces, which is relevant when working with large corpora of reviews [7].

Furthermore, Mukherjee et al. (2017) explored aspect-based sentiment analysis to rate institutions based on specific features (e.g., management, infrastructure, faculty), which is directly applicable to multi-faceted organizational evaluations [8].

Overall, these studies underscore the growing significance of sentiment analysis for extracting actionable insights and the effectiveness of machine learning techniques in building robust classification systems. Building on this foundation, the current study aims to develop a machine learning-based framework for rating top organizations using aggregated public sentiment from multiple online platforms.

III. PROBLEM STATEMENT

In the modern digital era, organizations are continuously evaluated by customers and stakeholders through online reviews, social media comments, and feedback platforms. However, manually analyzing and interpreting this vast amount of unstructured textual data is time-consuming, error-prone, and inefficient. Moreover, traditional rating systems often fail to capture the nuanced sentiments expressed in user feedback, leading to biased or incomplete evaluations of organizational performance.

There is a lack of automated, intelligent systems that can efficiently process user-generated content and provide reliable sentiment-based ratings for organizations. Without such a system, stakeholders—including customers, investors, and regulatory bodies—struggle to make informed decisions based on public opinion.

Therefore, there is a need for a robust sentiment analysis framework that leverages machine learning techniques to classify and quantify user sentiments from diverse online sources and generate accurate, data-driven organizational ratings. This study addresses the challenge by developing a scalable and effective sentiment analysis-based model to rank top organizations using real-world textual data.

IV. OBJECTIVE

The proposed thesis major objective is to give top organization rating based on people sentiment classification. This thesis introduces a new approach to get the rating of top organization which is more accurate and efficient than existing system. It is using people feedback to generate the rating of top organization and providing analysis report and comparison chart of feedbacks.

V. PROPOSED ALGORITHM

- **Preprocessing**

The preprocessing mainly used to filter the data based on Machine Learning (ML) concept, means it is removing unnecessary part of data from input data. Here it will remove word like is, am, are, was, were, will, will be etc. Apart from this it will remove all special character from sentence. If any URL is mention in sentence that is also removed from sentence.

- **Self-Learning and word standardization System**

The proposed system of the thesis using maximum entropy algorithm for self learning of people feedback sentiment to provide rating of top organizations. In this algorithm, first we have to pass keyword (meaning full words) it generate numerical values for all keyword and based on threshold value in create categories range for positive, negative and neutral sentiment. Based on these categories machine self recognize the sentiment of the word.

VI. TECHNIQUES

A couple of works try to show the different frameworks associated with Sentiment Analysis. A huge part of them bundle the works from the point of view of the unmistakable applications/challenges that can be found in SA as in (Pang and Lee, 2008) and (Liu and Zhang, 2012). Various works like (Tsytsarau and Palpanas, 2011) or (Feldman, 2013) are revolved around the essential purposes of SA. As such, Feldman packs all works under five essential social affairs: report level notion examination, sentence-level inclination assessment, point of view based inclination examination, comparable supposition assessment and, end jargon getting (Feldman, 2013). In addition, on the other hand, Tsytsarau and Palpanas generally base on end all out, feeling spam and intelligent irregularities examination, especially associated with Web organizations, for example, microblogs or spouting data, among others (Tsytsarau and Palpanas, 2011). They present four one of a kind centers worried past endeavors to arrange Sentiment Analysis techniques: AI, word reference based, truthful and semantic. Possibly, the most intriguing work from the point of view of the SA frameworks is (Medhat et al., 2014), which shows a refined course of action of doubtlessly comprehended SA systems (see Fig. 1) including new examples, for instance, Emotion Detection (Rao et al., 2014), Building Resources and Transfer Learning.

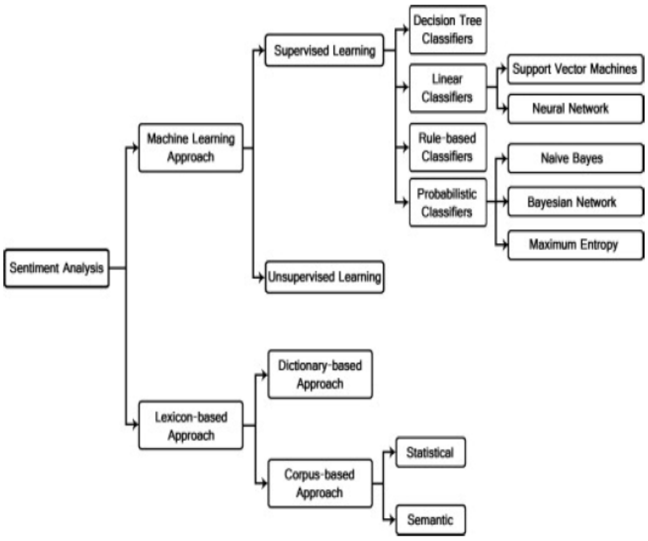

Figure 1: Sentiment classification techniques (Medhat et al., 2014)

VII. RESULTS

Now days, all top organization doing survey so they can get know their rating in between other similar organizations. Clients first see the rating of the organization while joining them and from customer point of view organization rating shows the growth of the organization. This proposed thesis concept introduce a data mining technique to provide rating of organization based on customer feedbacks and plot the comparison graph of different organization, for this data mining analysis machine learning (ML) is used for data pre-processing and for self learning maximum entropy artificial intelligence (AI) algorithm is used, which is help to machine to self decide word meaning sense like positive, negative or neutral.
Table 1 shows a couple of the insights acquired.

TABLE I. STATISTICS ON THE SENTIMENTS EXTRACTED FROM TWEETS

College	**Ratio of positive to negative tweets**	**Average positive sentiment**
NIT	1.51	2.84
IIT	1.62	3.12
AIIMS	1.78	5.24

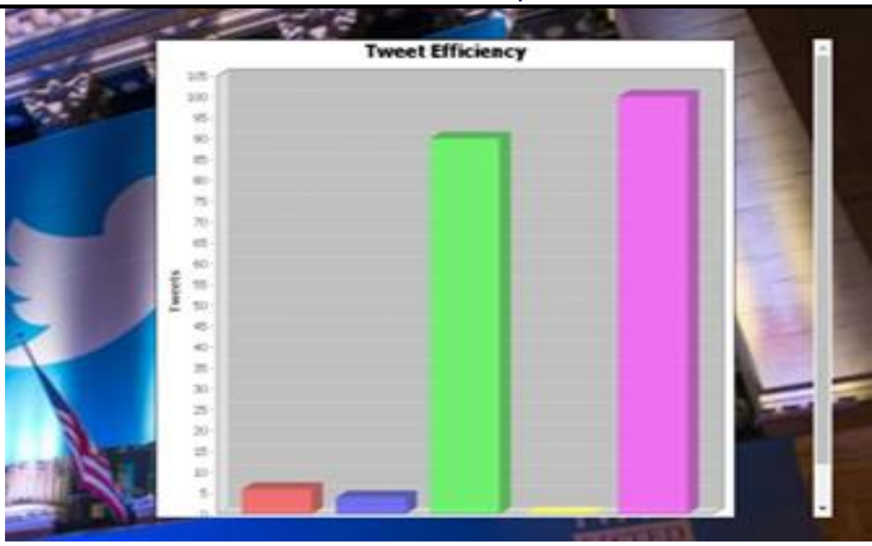

Figure 2. Shows the tweets efficiency graph.

VIII. CONCLUSIONS

Top organization rating based on people feedback data sentiment analysis is a convincing strategy for orchestrating the evaluations figured by people as for any point, organization or thing. Motorization of this task makes it less requesting to deal with the enormous proportion of data gave by individual's criticism. Stood out from the current system, taking into account the edge thought, cycle thought, watchwords are broadened .Based on the all-inclusive words adequacy of the structure is extended and give the rating of top association in India.

In future proposed system can be used in mobile app, or it can be used on social media application. In proposed system using excel data for data analysis, but in future input data can be getting from web application which is running on internet.

REFERENCES

[1] Pang, B., Lee, L., & Vaithyanathan, S. (2002). Thumbs up? Sentiment classification using machine learning techniques. In Proceedings of the ACL-02 conference on Empirical methods in natural language processing (pp. 79-86).

[2] Liu, B. (2012). Sentiment analysis and opinion mining. Synthesis Lectures on Human Language Technologies, 5(1), 1-167.

[3] Sarlan, A., Nadam, C., & Basri, S. (2014). Twitter sentiment analysis. International Conference on Information Technology and Multimedia (pp. 212-216). IEEE.

[4] Pak, A., & Paroubek, P. (2010). Twitter as a corpus for sentiment analysis and opinion mining. In LREC.

[5] Zhang, Y., & Wallace, B. C. (2018). A sensitivity analysis of (and practitioners' guide to) convolutional neural networks for sentence classification. In Proceedings of the 8th International Conference on Learning Representations.

[6] Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. Ain Shams Engineering Journal, 5(4), 1093-1113.

[7] Tripathy, A., Agrawal, A., & Rath, S. K. (2015). Classification of sentiment reviews using n-gram machine learning approach. Expert Systems with Applications, 57, 117-126.

[8] Mukherjee, S., Joshi, M., & Bhattacharyya, P. (2017). Aspect-based sentiment analysis: A survey. ACM Computing Surveys (CSUR), 50(3), 1-39.

[9] Dan Klein and Christopher D. Manning. 2003. Accurateunlexicalized parsing. Proceedings of the 41st Meeting of the Association for Computational Linguistics, pages 423–430.

[10] Alessandro Moschitti. 2006. Efficient convolution kernels for dependency and constituent syntactic trees. In Proceedings of the 17th European Conference on Machine Learning.

[11] Alexander Pak and Patrick Paroubek. 2010. Twitter as a corpus for sentiment analysis and opinion mining. Proceedings of LREC.

[12] B. Pang and L. Lee. 2004. A sentimental education: Sentiment
analysis using subjectivity analysis using subjectivity summarization based on minimum cuts. ACL.

[13] P. Turney. 2002. Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews. ACL.

[14] Abbasi, A., Chen, H., Salem, A., 2008. Sentiment analysis in multiple languages. ACM Transactions on Information Systems 26, 1{34.

[15] Baccianella, S., Esuli, A., Sebastiani, F., 2010. SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining, in: Calzolari, N., Choukri, K., Maegaard, B., Mariani, J., Odijk, J., Piperidis, S., Rosner, M., Tapias, D. (Eds.), Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10), European Language Resources Association (ELRA), Valletta, Malta. pp. 2200{2204.

[16] Banea, C., Mihalcea, R., Wiebe, J., 2010. Multilingual subjectivity: are more languages better?, in: Proceedings of the 23rd International Conference on Computational Linguistics (COLING '10), pp. 28{36.

[17] Bar-Haim, R., Dinur, E., Feldman, R., Fresko, M., Goldstein, G., 2011. Identifying and following expert investors in stock microblogs, in: Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP '11), pp. 1310{1319.

[18] Barbosa, L., Feng, J., 2010. Robust sentiment detection on Twitter from biased and noisy data, in: Proceedings of the 23rd International Conference on Computational Linguistics (COLING '10), pp. 36{44.

[19] Beineke, P., Hastie, T., Manning, C., Vaithyanathan, S., 2004. Exploring Sentiment Summarization, in: Shanahan, J.G., Wiebe, J., Qu, Y. (Eds.), Proceedings of the AAAI Spring Symposium on Exploring Attitude and Affect in Text Theories and Applications, AAAI Press. pp. 1{4.

[20] Boiy, E., Moens, M.F., 2008. A machine learning approach to sentiment analysis in multilingual Web texts. Information Retrieval 12, 526{558.

[21]Bollen, J., Mao, H., 2011. Twitter Mood as a Stock Market Predictor. Journal of Computational Science 44, 91{94.

[22] Cambria, E., Schuller, B., Xia, Y., Havasi, C., 2013. New Avenues in Opinion Mining and Sentiment Analysis. IEEE Intelligent Systems 28, 15{21.

[23] Cao, Q., Duan, W., Gan, Q., 2011. Exploring determinants of voting for the helpfulness of online user reviews: A text mining approach. Decision Support Systems 50, 511{521.

[24] Castellanos, M., Dayal, U., Hsu, M., Ghosh, R., Dekhil, M., Lu, Y., Zhang, L., Schreiman, M., 2011. LCI: a social channel analysis platform for live customer intelligence, in: Proceedings of the 2011 international conference on Management of data - SIGMOD '11, ACM Press, New York, New York, USA. pp. 1049{1058.

[25] Chen, B., Zhu, L., Kifer, D., Lee, D., 2010. What is an Opinion About? Exploring Political Standpoints using Opinion Scoring Model, in: Proceeedings of AAAI Conference on Artificial Intelligence (AAAI-2010), pp. 1007{1012.

[26] Chenlo, J.M., Losada, D.E., 2014. An empirical study of sentence features for subjectivity and polarity classification. Information Sciences 280, 275{288.

[27] Deerwester, S.C., Dumais, S.T., Landauer, T.K., Furnas, G.W., Harshman, R.A., 1990. Indexing by Latent Semantic Analysis. Journal of the American Society of Information Science 41, 391{407.

[28] Esuli, A., Sebastiani, F., 2006. Determining Term Subjectivity and Term Orientation for Opinion Mining, in: Proceedings of the 11th Conference of the European Chapter of the Association for Computational Linguistics (EACL06), pp. 193{200.

[29] Feldman, R., 2013. Techniques and applications for sentiment analysis. Communications of the ACM 56, 82{89.

[30] Filatova, E., 2012. Irony and Sarcasm: Corpus Generation and Analysis Using Crowdsourcing, in: Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC-2012), Istanbul, Turkey. pp. 392{398.

[31] Finn, A., Kushmerick, N., 2006. Learning to classify documents according to genre: Special Topic Section on Computational Analysis of Style. Journal of the American Society for Information Science and Technology 57, 1506{ 1518.

[32] Ganesan, K., Zhai, C., Han, J., 2010. Opinosis: a graph-based approach to abstractive summarization of highly redundant opinions, in: Proceedings of the 23rd International Conference on Computational Linguistics (COLING '10), pp. 340{348.

[33] Ganesan, K., Zhai, C., Viegas, E., 2012. Micropinion generation: An Unsupervised Approach to Generating Ultra-Concise Summaries of Opinions, in: Proceedings of the 21st international conference on World Wide Web - WWW '12, ACM Press. pp. 869 { 878.
Garcia Esparza, S., OMahony, M., Smyth, B., 2012. Mining the real-time web: A novel approach to product recommendation. Knowledge-Based Systems 29, 3{11.

[34] Gerani, S., Carman, M., Crestani, F., 2012. Aggregation Methods for Proximity-Based Opinion Retrieval. ACM Transactions on Information Systems 30, 1{36.

[35] Go, A., Bhayani, R., Huang, L., 2009. Twitter sentiment classification using distant supervision. Technical Report. Standford University.

[36] Groh, G., Hau_a, J., 2011. Characterizing Social Relations Via NLP-Based Sentiment Analysis, in: Adamic, L.A., Baeza-Yates, R.A., Counts, S. (Eds.), ICWSM, The AAAI Press. pp. 502{505.

[37] Guo, L., Wan, X., 2012. Exploiting syntactic and semantic relationships between terms for opinion retrieval. Journal of the American Society for Information Science and Technology 63, 2269{2282.

[38] He, Y., Lin, C., Alani, H., 2011. Automatically extracting polarity-bearing topics for cross-domain sentiment classification, in: Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies (HLT '11), pp. 123{131.