TWITTER SENTIMENT ANALYSIS

Rahul Tiwari*1, Asim Ahamad*2

*1,2Student, Department of Computer Science and Engineering, Shri Ramswaroop MemorialCollege of Engineering and Management, Lucknow, Uttar Pradesh, India.

ABSTRACT

Twitter sentiment analysis refers to the process of using natural language processing and machine learning techniques to analyze the sentiment (positive, negative, or neutral) of tweets on Twitter. This analysis can be used to understand public opinion on various topics, such as products, events, and political issues. The sentiment analysis process typically involves data preprocessing, feature extraction, model training, and evaluation. The accuracy of the sentiment analysis model can be improved by using techniques such as word embeddings, sentiment lexicons, and deep learning algorithms. Twitter sentiment analysis has numerous applications, including brand management, customer service, and market research. Keywords: Sentiment analysis, Twitter data, Natural language processing(NLP), Text preprocessing.

I.   INTRODUCTION

Twitter sentiment analysis is a powerful technique used to analyze and understand the sentiment or opinions expressed in tweets on the popular social media platform, Twitter. With millions of users sharing their thoughts, ideas, and emotions in short messages called tweets, sentiment analysis on Twitter provides valuable insights into public sentiment, attitudes, and trends. The goal of Twitter sentiment analysis is to automatically classify tweets as positive, negative, or neutral, based on the sentiment conveyed by the text. This process involves natural language processing (NLP) techniques, machine learning algorithms, and text mining to extract meaningful information from the tweets. By analyzing the sentiment of tweets, individuals, businesses, organizations, and researchers can gain valuable insights and make informed decisions. For example, companies can use sentiment analysis to monitor brand perception, identify customer opinions, and track the success of marketing campaigns. Governments can utilize sentiment analysis to understand public sentiment towards policies or events. Researchers can analyze sentiment to study public opinion on various topics, such as politics, entertainment, or social issues. The process of Twitter sentiment analysis typically involves several steps. First, the tweets are collected either through the Twitter API or by utilizing data scraping techniques. Then, the collected tweets undergo pre-processing, which includes tasks such as removing noise (e.g., URLs, hashtags, emojis), tokenization, stemming, and removing stop words. Next, the pre-processed tweets are passed through machine learning or deep learning models, such as Naive Bayes, Support Vector Machines (SVM), Recurrent Neural Networks (RNN), or Convolutional Neural Networks (CNN). These models are trained on labeled datasets, where human annotators have manually labeled tweets as positive, negative, or neutral. During the training phase, the models learn to recognize patterns and features in the text that indicate sentiment. Once the model is trained, it can be used to predict the sentiment of new, unseen tweets. The output of the sentiment analysis is typically a sentiment score or a classification label indicating whether the tweet is positive, negative, or neutral. There are various challenges in performing Twitter sentiment analysis. These challenges include dealing with noisy and informal language used on Twitter, understanding sarcasm and irony, handling misspellings and abbreviations, and addressing biases in the training data.

II.   LITERATURE REVIEW

A. Park, A., & Paroubek, p. (2010)- Twitter as a Corpus for Sentiment Analysis and Opinion Mining. In Proceedings of the 7th Conference on International Language Resources and Evaluation (LREC 2010). This influential paper introduced Twitter as a valuable corpus for sentiment analysis and opinion mining. The authors explored various linguistic features and machine learning techniques for sentiment classification on Twitter data.

B. Go, A., Bhayani, R., & Huang, L. (2009)- Twitter Sentiment Classification Using Distant Supervision. In Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP 2009). This paper proposed a semi-supervised approach using distant supervision for sentiment classification on Twitter. They leveraged emoticons and the presence of positive/negative terms to train a sentiment classifier.

C. Agarwal, A., Xie, B., Vovsha, I., Rambow, O., & Passonneau, R. (2011)- Sentiment Analysis of Twitter Data. In Proceedings of the Workshop on Languages in Social Media (LSM 2011). The authors conducted a comprehensive study on sentiment analysis of Twitter data. They compared different feature sets, classifiers, and techniques for sentiment classification and also explored the impact of training data size on classification accuracy.

D. Thelwall, M., Buckley, K., & Paltoglou, G. (2012)- Sentiment in Twitter Events. Journal of the Association for Information Science and Technology, 63(1), 109-118. This study focused on analyzing sentiment in Twitter data related to specific events or topics. The authors proposed a method to detect bursts of sentiment in time series data and examined how sentiment evolves over time during events.

E. Nguyen, D., Gravel, R., Trieschnigg, D., & Meder, T. (2013)- How Old is this Tweet? A Study of Twitter Lifetime. In Proceedings of the International Conference on Social Informatics (SocInfo 2013). This research investigated the temporal aspect of sentiment in tweets by analyzing the lifetime of tweets and its impact on sentiment analysis. The study highlighted the importance of considering tweet age for accurate sentiment analysis.

F. Bermingham, A., & Smeaton, A. F. (2011)- Classifying Sentiment in Microblog Posts. In Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2011). The authors explored the use of lexical and syntactic features, as well as sentiment lexicons, for sentiment classification in microblog posts. They also compared different machine learning algorithms for sentiment analysis on Twitter data.

G. Agarwal, A., Xie, B., Vovsha, I., Rambow, O., & Passonneau, R. (2013)- Sentiment Analysis of Twitter Data. ACM Transactions on Intelligent Systems and Technology, 4(2), 21. This comprehensive survey paper provides an overview of sentiment analysis techniques applied to Twitter data. It covers different aspects, including data collection, preprocessing, feature extraction, sentiment classification, and evaluation methodologies.

III.   METHODOLOGY

The methodology of Twitter sentiment analysis using live data access can be summarized in the following steps:

1. Connect to the Twitter API to access live tweet data in real-time.

2. Preprocess the data by removing irrelevant information, converting text to lowercase, and cleaning using techniques such as stop word removal, stemming, and lemmatization.

3. Extract features from the text using techniques such as bag-of-words, n-grams, or word embeddings.

4. Train a machine learning model to classify the sentiment of the tweets using algorithms such as Support Vector Machines, Random Forests, or Neural Networks.

5. Continuously update the model with new data as it becomes available to improve its accuracy.

6. Use the model to classify the sentiment of new tweets in real-time and gain insights into public opinion as it evolves.
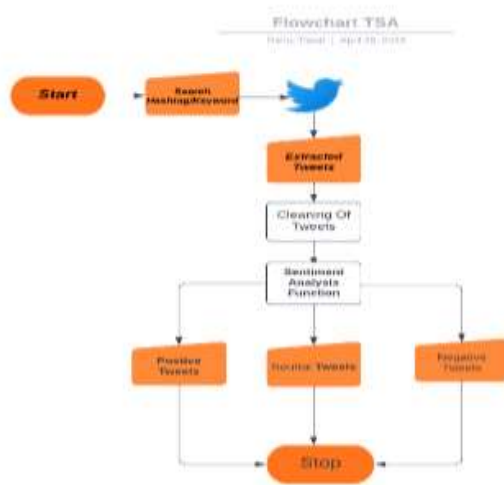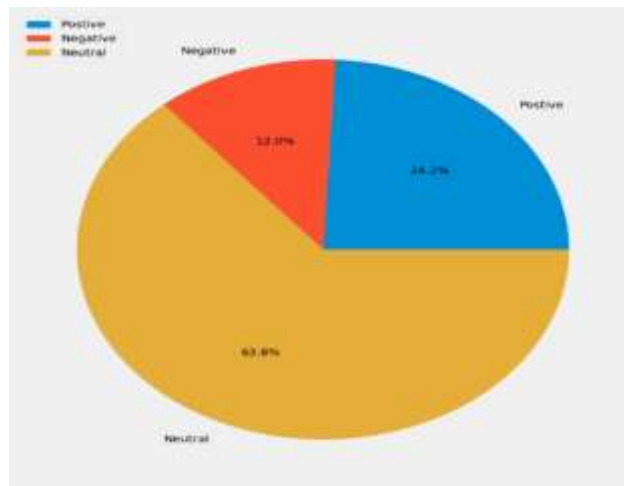
IV.   VISUALIZATION

A. Scatter plot
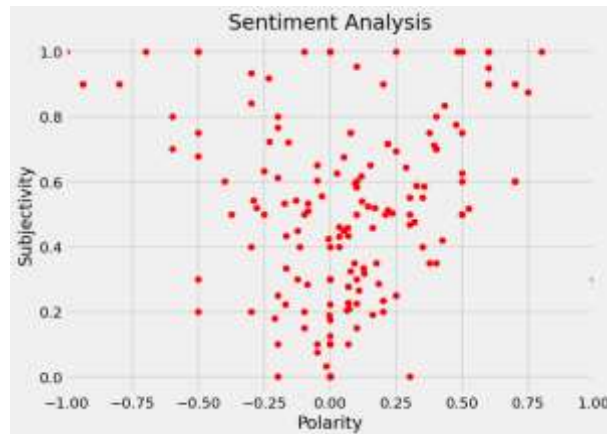
B. Pie Chart plot

V.   CONCLUSION

Twitter sentiment analysis is a powerful tool for gaining insights into public opinion on various topics. By analysing the sentiment of tweets related to a particular topic, analysts can understand how people feel about it, identify emerging trends, and make data-driven decisions. The methodology of Twitter sentiment analysis involves collecting a large dataset of tweets, pre-processing the data, extracting features, training a machine learning model, evaluating its performance, and using it to classify the sentiment of new tweets. With the rise of social media and its impact on public opinion, Twitter sentiment analysis is becoming increasingly important for businesses, governments, and organizations to stay informed and make informed decisions. However, it is important to note that the accuracy of the analysis depends on the quality of the data and the machine learning model used. Therefore, analysts must carefully select the appropriate techniques and continuously update the model with new data to improve its accuracy.

VI.   REFERENCES

[1] Pak, A., & Paroubek, P. (2010). Twitter as a corpus for sentiment analysis and opinion mining. In Proceedings of the Seventh Conference on International Language Resources and Evaluation (LREC'10) (pp. 1320-1326).

[2] Agarwal, A., Xie, B., Vovsha, I., Rambow, O., & Passonneau, R. (2011). Sentiment analysis of twitter data. In Proceedings of the Workshop on Languages in Social Media (LSM 2011) (pp. 30-38).

[3] Zhang, H., Guo, Q., & Zhang, Y. (2018). Twitter sentiment analysis: A survey. Journal of Information Processing Systems, 14(4), 901-912.

[4] Go, A., Bhayani, R., & Huang, L. (2009). Twitter sentiment classification using distant supervision. CS224N Project Report, Stanford, 1(12), 2009.

[5] Wang, H., Can, D., Kazemzadeh, A., Bar, F., & Narayanan, S. (2012). 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).

[6] Bifet, A., Holmes, G., Pfahringer, B., & Frank, E. (2010). Mining twitter with weka. Journal of Machine Learning Research, 11(Feb), 2531-2535.

[7] Kouloumpis, E., Wilson, T., & Moore, J. D. (2011). Twitter sentiment analysis: The good the bad and the OMG!. ICWSM, 11(538-541), 164.

[8] Aggarwal, C. C., & Zhai, C. (2012). Mining text data (Vol. 9). Springer Science & Business Media.

[9] Cambria, E., & Hussain, A. (2012). Sentic computing: The common sense paradigm for natural language processing. In Handbook of natural language processing (pp. 343-364). Springer, New York, NY.

[10] Jansen, B. J., Zhang, M., Sobel, K., & Chowdury, A. (2009). Twitter power: Tweets as electronic word of mouth. Journal of the American Society for Information Science and Technology, 60(11), 2169-2188.