Twitter Classification of Complaints

Aniket Solanki[1], Harsh Tripathi[2], Harsheet Soni[3], Kunal Singh[4], Naina Kaushik[5]
[1234]*Student Computer Engineering, MCT's Rajiv Gandhi Institute of Technology, Mumbai University*
[5]*Professor, MCT's Rajiv Gandhi Institute of Technology, Mumbai University*

Abstract—This project addresses the problem of sentiment analysis in twitter; that is classifying tweets according to the sentiment expressed in them: positive, negative or neutral. Twitter is an online micro-blogging and social-networking platform which allows users to write short status updates of maximum length 140 characters. It is a rapidly expanding service with over 200 million registered users out of which 100 million are active users and half of them log on twitter on a daily basis - generating nearly 250 million tweets per day. Due to this large amount of usage we hope to achieve a reflection of public sentiment by analysing the sentiments expressed in the tweets. Analysing the public sentiment is important for many applications such as firms trying to find out the response of their products in the market, predicting political elections and predicting socioeconomic phenomena like stock exchange. The aim of this project is to develop a functional classifier for accurate and automatic sentiment classification of an unknown tweet stream.

Index Terms— Sentiments, Predicting.

I. INTRODUCTION

We have chosen to work with twitter since we feel it is a better approximation of public sentiment as opposed to conventional internet articles and web blogs. The reason is that the amount of relevant data is much larger for twitter, as compared to traditional blogging sites. Moreover the response on twitter is more prompt and also more general (since the number of users who tweet is substantially more than those who write web blogs on a daily basis).

Sentiment analysis of public is highly critical in macro-scale socioeconomic phenomena like predicting the stock market rate of a particular firm. This could be done by analysing overall public sentiment towards that firm with respect to time and using economics tools for finding the correlation between public sentiment and the firm's stock market value. Firms can also estimate how well their product is responding in the market, which areas of the market is it having a favourable response and in which a negative response (since twitter allows us to download stream of geo-tagged tweets for particular locations. If firms can get this information they can analyze the reasons behind geographically differentiated response, and so they can market their product in a more optimized manner by looking for appropriate solutions like creating suitable market segments. Predicting the results of popular political elections and polls is also an emerging application to sentiment analysis. One such study was conducted by Tumasjan et al. in Germany for predicting the outcome of federal elections in which concluded that twitter is a good reflection of offline sentiment.

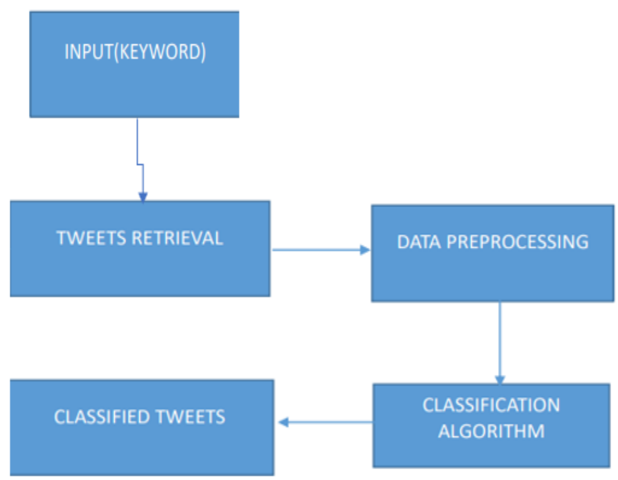


Fig 1: System Architecture

II. LITERATURE REVIEW

A. Literature Review

Due to the enormous amount of data generated on social media networks, Twitter sentiment analysis is a burgeoning field of study. An overview of the most recent studies on Twitter sentiment analysis is provided in this review of the literature.

1)"Deep Learning Techniques for Sentiment Analysis on Twitter Data: A Review" by R. Selvi et al. (2021). Deep learning techniques for sentiment analysis on Twitter are examined in detail in this work. The authors explore several pre-processing and feature extraction strategies in addition to comparing various deep learning architectures. They also emphasize the difficulties with Twitter sentiment analysis and offer suggestions for further study.

2) "Sentiment Analysis of Twitter Data: A Complete Study," by P. Pakhare et al. (2021), offers a thorough analysis of numerous methods such as lexicon-based, machine learning-based, and deep learning-based techniques for sentiment analysis on Twitter data. The authors examine the difficulties sarcasm and irony provide and make recommendations for future research topics.

3) First most recent techniques and uses of sentiment analysis on Twitter are covered by M. I. Ali et al. in their paper "Twitter Sentiment Analysis: A Survey of Techniques and Applications" from 2020. They contrast various strategies, including lexicon-based, machine learning-based, and deep learning-based ones, and offer suggestions for further research.

4) In "Sentiment Analysis on Twitter Data: A Review," R. Kumar et al. (2020), they concentrate on the various methods for pre-processing, feature extraction, and sentiment classification. The authors talk about the difficulties in performing sentiment analysis on Twitter data and make recommendations for future study areas.

The informality of tweets and the usage of irony and sarcasm, according to the literature, make Twitter sentiment analysis a difficult endeavour. Whereas recent developments in deep learning algorithms have yielded encouraging results, future research should concentrate on enhancing the precision of sentiment analysis using Twitter data.

B. Survey of Existing Systems

1) IBM Watson: A sentiment analysis service is part of IBM Watson, a collection of AI-powered products. It analyses text using machine learning techniques and assigns a sentiment score. IBM Watson may be tailored to particular domains and can evaluate text in a variety of languages.Moreover, IBM Watson offers a subjectivity score and a sentiment polarity score for text. The emotion of the text is indicated by the polarity score, which is a float value between -1 and 1. (negative to positive). The subjectivity score, which ranges from 0 to 1, reflects how subjective (opinionated) or objective the text is (factual)

2) Lexalytics: A text analytics software provider called Lexalytics provides sentiment analysis solutions, including a Twitter sentiment analysis solution. It analyses text and identifies sentiment using machine learning and natural language processing techniques. Several languages of text can be analysed using Lexalytics, which can also be tailored to certain fields.Text is given a subjectivity score and a sentiment polarity score by Lexalytics. The emotion of the text is indicated by the polarity score, which is a float value between -1 and 1. (negative to positive).

3) CoreNLP: A sentiment analysis module is part of the Java-based natural language processing toolbox known as Stanford CoreNLP. It analyses text using deep learning techniques and assigns a sentiment score. Stanford CoreNLP can be tailored to particular fields and can analyse text in a variety of languages.Text is given a sentiment polarity score and a subjectivity score using Stanford CoreNLP. The emotion of the text is indicated by the polarity score, which is a float value.

4) Vader Sentiment Analysis: The rule-based sentiment analysis program Vader (Valence Aware Dictionary and Sentiment Reasoner) employs a dictionary of words and phrases to assign a sentiment score to text. The lexicon includes positive and negative sentiment scores for words and phrases and is based on human-labeled data. Moreover, Vader considers negations, intensifiers, and other textual elements that may have an impact on emotion.

Vader has demonstrated strong performance on Twitter and other social media metrics. It features a straightforward API for text analysis and is available as a Python package.

5) TextBlob: It is a Python package that offers a straightforward API for carrying out typical natural language processing operations, such as sentiment analysis. It categorises text as either positive, negative, or neutral using a Naive Bayes classifier. The classifier can be applied to different domains and was trained on a labelled dataset of movie reviews.TextBlob also assigns a subjectivity and sentiment polarity score to each piece of text. The emotion of the text is indicated by the polarity score, which is a float value between -1 and 1. (negative to positive).

C. linitation of Existing Systems/Research gap

1) Sampling bias: Since Twitter users may not represent the entire population, the opinions posted

there might not be reflective of the general public. Compared to the average population, Twitter users tend to be younger, more urban, and politically active.
2) Limited context: Because Twitter tweets can only be 280 characters long, they frequently lack the background information required to fully comprehend the mood being expressed. An seeming negative tweet, for instance, can actually be ironic or satirical.
3) Slang and colloquial language: Twitter users frequently employ colloquial language, which can be challenging for sentiment analysis algorithms to understand. This can result in incorrect sentiment analysis findings.
4) Irony and sarcasm: Twitter users frequently employ irony and sarcasm to convey their thoughts, which sentiment analysis algorithms may find challenging to identify. Even though a tweet seems to be harsh, it could also be positive.
5) Sentiment polarity: Text is often categorised as either positive, negative, or neutral by algorithms that analyse sentiment. But sentiment polarity is frequently more nuanced than this and can change based on the situation and the speaker.
6) Minimal training data: To learn how to classify text, sentiment analysis algorithms need training data. Nevertheless, there may not be enough training data for Twitter sentiment analysis, which could result in unreliable findings.
7) Emojis and emoticons: Twitter users frequently utilise emojis and emoticons to convey emotion, which can be challenging for algorithms that analyse sentiment. Depending on the situation, the same emoji can imply something entirely different.
8) Tweets in multiple languages are possible on Twitter because it is a multilingual platform. But, it's possible that algorithms for sentiment analysis won't be able to effectively interpret tweets in languages they haven't been trained on.

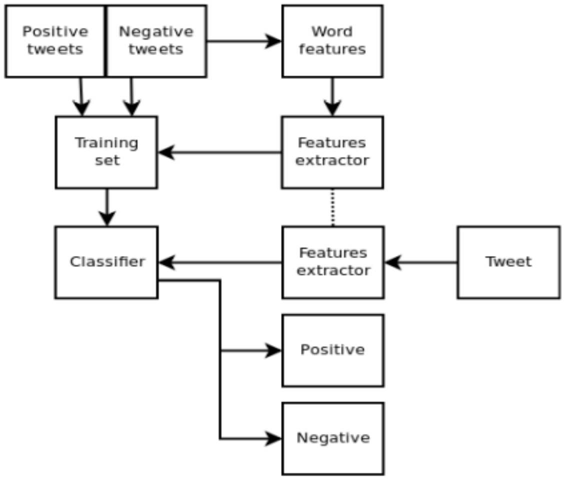
III. METHODOLOGY

Data Collection:Data gathering is the initial step in the sentiment analysis process for Twitter. The Twitter API, third-party programmes like Tweepy, or for-profit services like Gnip or DataSift can all be used for this. The information may be gathered in real-time or over a predetermined period of time, and it may comprise tweets on a certain subject or hashtag, as well as tweets from particular individuals or places.

Data preprocessing and cleaning: After the data has been gathered, it must be cleaned and preprocessed to weed out extraneous information and make it ready for analysis. There are multiple steps in this:
a. Eliminating duplicate tweets: It's critical to get rid of duplicate tweets before sentiment analysis because they can bias the results.
b. Retweets can potentially bias the results of sentiment analysis because they do not reflect the original sentiment.
c. Removing spam: Spam tweets, which can be found via automatic or manual approaches, can also skew the results.
d. Eliminating non-textual content: To draw attention to text-only content, non-textual content such as images, videos, and other media might be eliminated.
e. Text normalisation, often known as case-insensitive text or punctuation removal, is the process of transforming text to a standard format.

Classification of sentiment: Positive, negative, or neutral sentiment is assigned to each tweet based on algorithms that analyse sentiment. There are various methods for sentiment analysis, such as:
a. Rule-based strategies: These strategies utilise a set of pre-established rules to assess the tone of a tweet based on its keywords or other characteristics.
b. Machine learning techniques: Using machine learning techniques, a model is trained on a labelled dataset to discover how to categorise text according to its sentiment.
c. Hybrid methods: To increase accuracy, hybrid approaches integrate rule-based and machine learning techniques.

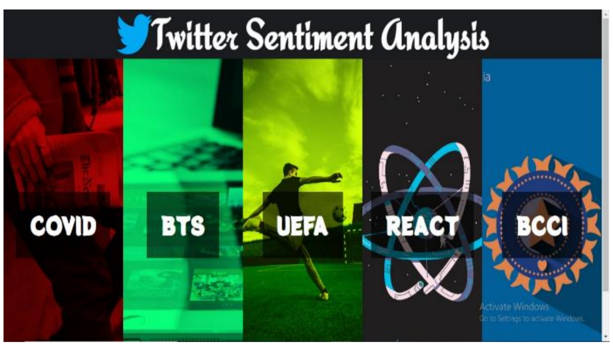
Analyzing and visualising the results is the last step in the Twitter sentiment analysis process. In order to do this, it may be necessary to discover trends and patterns in the data, identify significant influencers or subjects, and create graphs and charts to represent sentiment across time.



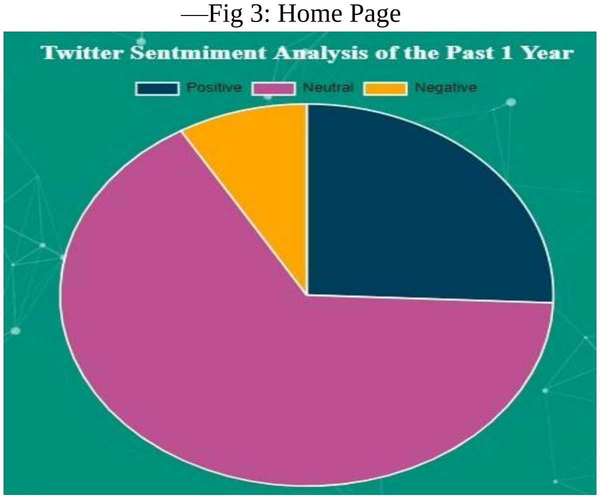
—Fig 2: System Architecture

Overall, there are a number of processes in the approach for Twitter sentiment analysis, including data collection, cleaning and preprocessing, sentiment classification, sentiment aggregation, and visualisation and analysis. Depending on the study issue and the objectives of the analysis, many approaches may be employe.

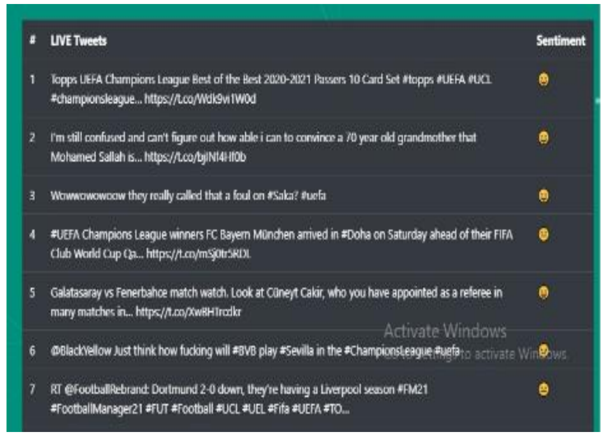
IV.RESULTS



—Fig 3: Home Page



—Fig 3:Tweet Sentiment Analysis



—Fig 4:Live Tweet Sentiment Analysis of UEFA

—Fig 4: Entered Tweet Sentimental Analysis of UEFA

V. CONCLUSION

In this paper we conclude that the proposed twitter sentimental analysis addresses the critical issues of getting the sentiments of a human from their text.The use of tweepy api of python helps to fetch live data from the twitter api and sentiment can be used for predictive analysis: Twitter sentiment analysis can be used to predict future trends or events. By analyzing the sentiment of tweets, researchers and businesses can gain insight into the likelihood of a particular event occurring or how consumers may react to a new product or service. Machine learning algorithms can be used to improve the accuracy of Twitter sentiment analysis. By training the algorithms on large datasets, they can learn to identify patterns in language and sentiment that can help to improve the accuracy of the analysis.

ACKNOWLEDGMENT

We wish to express our sincere gratitude to Dr. Sanjay U. Bokade, Principal and Prof. S. P. Khachane, Head of Department of Computer Engineering at MCT's

Rajiv Gandhi Institute of Technology, for providing us with the opportunity to work on our project, "Crowdfunding Platform Using Blockchain." This project would not have been possible without the guidance and encouragement of our project guide, Prof. Naina Kaushik. We would also like to thank our colleagues and friends who helped us in completing this project successfully.

REFERENCES

[1] A. Hassan, A. Abbasi and D. Zeng, “Twitter sentiment analysis: A bootstrap ensemble framework,” in Proc. ASE/ IEEE Int. Conf. on Social Computing, Alexandria, VA, USA, pp. 357–364, 2013.

[2] N. Yadav, O. Kudale, S. Gupta, A. Rao, and A. Shitole, “Twitter sentiment analysis using machine learning for product evaluation,” in2020 International Conference on Inventive Computation Technologies (ICICT), pp. 181–185, IEEE, 2020

[3] N. F. F. Da Silva and E. R. Hruschka, “Tweet sentiment analysis with classifier ensembles,” Decision Support Systems, vol. 66, no. C, pp. 170–179, 2014.

[4] C. Yang, K. H.-Y. Lin, and H.-H. Chen, “Emotion classification using web blog corpora,” in IEEE/WIC/ACM International Conference onWeb Intelligence (WI’07), pp. 275–278, IEEE, 2007.

[5] A. P. Patel, A. V. Patel, S. G. Butani and P. B. Sawant, "Literature Survey on Sentiment Analysis of Twitter Data using Machine Learning Approaches", *IJIRST-International journal for Innovative Reasearch in Scinece & Technology*, vol. 3, no. 10, 2017

[6] N. F. Alshammari and A. A. AlMansour, “State-of-the-art review on twitter sentiment analysis,” in 2019 2nd International Conference on Computer Applications & Information Security (ICCAIS), pp. 1–8,IEEE, 2019.

[7] J. Ho, D. Ondusko, B. Roy and D. F. Hsu, “Sentiment analysis on tweets using machine learning and combinatorial fusion,” in Int. Conf. on Dependable, Autonomic and Secure Computing, Pervasive Intelligence and Computing, Cloud and Big Data Computing, Fukuoka, Japan, Cyber Science and Technology Congress, 2019

[8] M. Gamon, “Sentiment classification on customer feedback data: noisydata, large feature vectors, and the role of linguistic analysis,” in COLING 2004: Proceedings of the 20th International Conference on Computational Linguistics, pp. 841–847, 2004

[9] S. A. El Rahman, F. A. AlOtaibi, and W. A. AlShehri, “Sentiment analysis of twitter data,” in 2019 International Conference on Computer and Information Sciences (ICCIS), pp. 1–4, IEEE, 2019

[10] A. Bermingham and A. F. Smeaton, “Classifying sentiment in microblogs: is brevity an advantage?,” in Proceedings of the 19th ACM international conference on Information and knowledge management, pp. 1833–1836, 2010

[11] B. Bhavitha, A. P. Rodrigues, and N. N. Chiplunkar, “Comparativestudy of machine learning techniques in sentimental analysis,” in2017 International Conference on Inventive Communication andComputational Technologies (ICICCT), pp. 216–221, IEEE, 2017.

[12] L. M. Rojas-Barahona, “Deep learning for sentiment analysis,”Language and Linguistics Compass, vol. 10, no. 12, pp. 701–719,2016.

[13] A. Severyn and A. Moschitti, “Unitn: Training deep convolutionalneural network for twitter sentiment classification,” in Proceedingsof the 9th international workshop on semantic evaluation (SemEval2015), pp. 464–469, 2015.

[14] M. Cliche, “Bb twtr at semeval-2017 task 4: Twitter sentiment analysis with cnns and lstms,” arXiv preprint arXiv:1704.06125, 2017.

[15] C. Baziotis, N. Pelekis, and C. Doulkeridis, “Datastories atsemeval-2017 task 4: Deep lstm with attention for message-leveland topic-based sentiment analysis,” in Proceedings of the 11thinternational workshop on semantic evaluation (SemEval-2017),pp. 747–754, 2017.

[16] M.-H. Su, C.-H. Wu, K.-Y. Huang, and Q.-B. Hong, “Lstm-based textemotion recognition using semantic and emotional word vectors,” in2018 First Asian Conference on Affective Computing and IntelligentInteraction (ACII Asia), pp. 1–6, IEEE, 2018.[28] Z. Jianqiang, G. Xiaolin, and Z. Xuejun, “Deep convolution

neuralnetworks for twitter sentiment analysis,” IEEE Access, vol. 6,pp. 23253–23260, 2018.