A review: Sentiment Analysis and Prediction for Cricket Tweets

Pranali J Phulare [1],   Dr. Sachin N Deshmukh [2]

[1] M. Tech Student, Department of Computer Science and Information Technology,
Dr. Babasaheb Ambedkar Marathwada University, Aurangabad, (M.S), India
[2] Associate Professor & HOD, Department of Computer Science and Information Technology,
Dr. Babasaheb Ambedkar Marathwada University, Aurangabad, (M.S), India
Email – [1]pranaliphulare@gmail.com,   [2]sndeshmukh@hotmail.com

Abstract: *Twitter has grown to become one of the most prominent social media platforms, and its popularity is growing exponentially each day as the number of tweets reaches millions. A common Web information mining research topic is sentiment analysis of online reviews. Emotion dictionaries or machine learning are used in the classic method of text sentiment analysis. This paper provides the review on sentiment analysis on different areas like Cricket tweets, products tweets, movie reviews, COVID-19 etc. On Twitter, we also talked about the general issues and applications of sentiment analysis.*

Key Words: *Web information mining, Twitter, text sentiment analysis, Cricket tweets.*

1. INTRODUCTION:

People all over the world have become increasingly passionate about expressing their emotions and thoughts on the internet in recent years, thanks to the rapid expansion of the Internet. As a result, sentiment analysis has emerged as a critical research area in Natural Language Processing. Traditional database solutions are incapable of processing millions of information; hence big data is the newest buzzword. Capturing data, analysing data, searching, storing, sharing visualization, and maintaining data privacy are all challenges addressed by big data. Twitter is used in this paper to do predictive analysis and extract values from data [15]. Using a standard relational database to process huge data will take more time, money, and resources. To improve the big data architecture, big data can be combined with data lakes, machine learning, and artificial intelligence to construct analytical algorithms. The data received from tweets can be used to categories tweets into good and negative categories. The most prevalent strategies for extracting feelings from textual input are used lexicons and machine learning approaches. Regular texts as well as texts with a lot of noise in the data can benefit from context-based sentimental analysis [10, 13].

There are three primary types of sentiment analysis approaches:

(1) Machine learning algorithms such as SVM, KNN and Nave Bayes.

(2) Giving them positive and negative polarity and categorizing them as unigrams or n-grams

(3) Sentiment lexicons are used to categories words as positive, negative, or neutral. Few people even add their level of emotion to each word [1]. It's utilized in a variety of applications, including cricket score or win forecasts, stock market forecasting, product review collection, and political forecasting based on public mood [15].

2. Challenges in Sentiment Analysis:

It's a difficult undertaking to conduct a sentiment analysis. Some of the difficulties encountered in Twitter Sentiment Analysis are listed below [16].

i.   Sarcasm Detection:-

Sarcastic sentences use positive words to express a negative view about a target in an unusual way. For example, "Nice scent. You have to take a shower in it." Although the sentence contains solely positive words, it conveys a negative message [17].

ii.   Domain dependence:-

In different realms, the same language or phrase can have distinct meanings. For example, the phrase "unpredictable" has a favourable connotation in the realm of movies, plays, and the like, but it has a bad connotation when applied to a vehicle's steering [18].

iii.   Identifying text's subjective elements:-

Sentiment-bearing content is represented by subjective portions. In some cases, the same term can be viewed as subjective, while in others, it can be treated as objective. This makes it difficult to distinguish between objective and subjective text.

iv.   Entity Recognition:-
     a.   Separating material concerning a certain entity and then analysing sentiment towards it is necessary.
     b.   "I despise Microsoft, but I enjoy Linux," for example. A simple bag-of-words method labels it as neutral, but it has a distinct attitude toward both entities in the statement.

v.   Internationalization:-
     The current research focuses primarily on English content, however Twitter has a diverse user base from all over the world [19, 20].

vi.   Applying sentiment analysis to Facebook messages:-
     Due to the limits imposed by the Facebook graph API and security policies in accessing data, there has been less effort on sentiment analysis on Facebook data.

3. LITERATURE SURVEY:

In this review of the literature, provide recent work on sentiment analysis and opinion mining that is relevant to my research. A CNNs (Convolution Neural Networks) model integrated with SVM text sentiment analysis was proposed by Yuling Chen and Zhi Zhang [2]. In this study, the benefits of CNNs and SVM are combined to construct a text sentiment analysis model using CNNs and SVM. The fully connected classification layer's classification performance for nonlinear separable data is limited, despite the fact that a convolution neural network can successfully extract meaningful feature representation from input samples. The experimental data in this article was the NLPCC2014 emotional analysis evaluation task data set based on deep learning technology. This was created to test the performance of sentiment analysis technology based on deep learning. There are 10,000 training data points and 2500 test data points, including 5000 positive and negative emotional polarity data points and 1250 positive and negative emotional polarity data points. We ran a comparison experiment with the standard CNN-based text sentiment analysis strategy and the NLPCC-SCDL-best technique to see how effective the suggested method is. The results of the experiments show that the proposed method improves text sentiment classification accuracy when compared to traditional CNN, and that sentiment analysis using CNNs and SVM is useful.

P. D. Turney [3] proposed the PMI-IR method for classifying reviews on movies, banks, travel destinations, and automobiles, and he achieved a 74 percent accuracy rate, despite the fact that movie reviews were difficult to analyze due to the prevalence of unpleasant phrases.

Varsha Sahayak et al. [4] suggested a mechanism for determining social media sentiment. NB, SVM, and maximum entropy are some of the unique machine learning algorithms used in the suggested system. The model was created using the WEKA and SentiView data visualization tools, which are commonly used.

RazaUl Mustafa et al. [5] conducted study into forecasting the outcome of a cricket match using tweets from the Twitter social media platform. The data was gathered from an IPL match from 2014 and a World Cup match from 2015. They are given some language characteristics in order to predict three essential things: (1) Number of followers on social media (2) Number of tweets (3) Predicting a score by classifying tweets as positive, negative, or neutral.

Based on movie reviews and blog entries, V. K. Singh et al. [6] established a method for sentiment analysis. They looked at two types of data from Libya and Tunisia, respectively: movie reviews and blog postings. Their major purpose was to compare the performance of the SentiWordNet technique to two alternative machine learning approaches. They used SVM and NB for sentiment classification. It was revealed that using Naive Bayes and SVM algorithms produced better results than using the SentiWordNet technique. It was also revealed that while the performance levels of SVM and NB for both movie reviews and blog posts were not the same, SentiWordNet's performance was the same for both types of datasets.

On the top three mobile brands, Samsung, BLU, and Apple, ZeeniaSingla et al. [7] employed the Statistical and Sentiment Analysis framework, SVM, for Cross-validation. Samsung receives the most positive response from statistical analysis. Customer happiness, quality, and rating were all higher for higher-priced products. By emotion orientation, these brands received positive ratings with positive sentiments of trust, delight, anticipation, and surprise. On categorized data, SVM used to cross validation gave 84.87 percent precision. This system can be developed more targeted in the future for the benefit of the industry. To mine user feedback and needs and deliver meaningful information to the designer, allowing the designer to improve the product and reduce the time it takes to gather requirements, reducing the cost of surveys, interviews, and market research.

Anh-Dung Vo et al. [8] proposed extracting and summarizing product facets from a large number of reviews in various product domains using two stage systems: first, knowledge extraction using NLP tools (DP, CR, NER) for the extraction of syntactic knowledge and entail opinion-facet connection; and second, opinion-facet connection extraction using NLP tools (DP, CR, NER) for the extraction of syntactic knowledge and entail opinion. Sentiment Analysis is the second stage, which is based on the knowledge gained in the first. The proposed system receives F1-scores of 0.714 for product camera reviews and 0.774 for laptop reviews. Facet extraction yielded satisfactory results for the laptop product. Facet extraction yielded satisfactory results in the case of the laptop device. This method may be used to extract product

facets from big datasets and can be used to tackle difficult problems such as sarcastic statements and implicit opinion inferences.

Yan Wan et al. [9] provided Fine-grained Sentiment Analysis with inherent features based on POS criteria on Product reviews and created a specialized market analysis to assist manufacturers and consumers on online reviews of one dress from the taobao.com site. By combining implicit features, the precision of fine-grained sentiment analysis is increased. The generality of the sentiment potency analysed is also improved. This approach assists producers in determining what aspects of their products could be improved and discovering niche markets. Customers also learn about the benefits and drawbacks of the product they want to buy, allowing them to make an informed selection. As a result, the system improves the accuracy of analysis.

The research published by Samuel et al. [11] attempted to undertake a textual analysis of US tweets during the peak of the COVID-19 outbreak in order to track the progression of dread sentiment that formed in response to the virus's rapid spread and its repercussions. Twitter was used to collect the data, which was based on crisis scenario keywords. Two machine learning approaches, Nave Bayes and Logistic regression, were used to conduct the analysis. The Nave Bayes (NB) classifier obtained 91 percent classification accuracy with short-length tweets, compared to 74 percent for logistic regression.

Alhajji et al. [12] used NB to execute Arabic sentiment analysis of tweets with a Python module called Natural Language Toolkit (NLTK). They saved time by using a categorized database of Arabic tweets with sentiment classifications based on the emoji lexicon. 50 percent of the data was favourable, and 50 percent was bad. A total of 47,000 tweets were used in the training process, with 11,000 being used in the testing process. The testing results obtained with a unigram NB had an accuracy of 0.89, precision of 0.92, recall of 0.86, and F-score of 0.89. Stemming, normalization, tokenization, and stop word removal were used as pre-processing procedures. With the NB classifier, achieved results scored 82.1 percent for precision and 81.4 percent for recall. With an 86 percent precision and recall, SVM came out on top.

Sumayh S. Aljameel et al. [14] constructed a model based on machine learning predictive models. They used Support Vector Machine (SVM), Knearest Neighbours (KNN), and Nave Bayes (NB), as well as the N-gram feature extraction methods for machine learning predictive model. The results show that in the Term Frequency–Inverse Document Frequency (TF-IDF) model, utilising the SVM classifier in conjunction with bigram outperformed other models by 85 percent. The south had the highest level of awareness of COVID-19 containment measures, while the centre had the lowest, according to the results of the awareness prediction.

Table 1: Comparative analysis of different techniques used

Sr. No.	Reference	Method	Accuracy
1	Peter D. Turney et al. [3]	PMI-IR Method	74 Percent accuracy rate
2	Zeenia Singla et al. [7]	Support Vector Machine (SVM) Classifier	84.87 Percent precision
3	Samuel, J et al. [11]	Bayes (NB) Classifier	91 Percent classification accuracy
4	Behdenna, S. et al. [12]	Support Vector Machine (SVM) With the NB classifier	82.1 Percent for precision and 81.4 percent for recall
5	Sumayh S. Aljameel et al. [14]	Support Vector Machine (SVM) Classifier	85 Percentage
6	Sandeep Nigam et al. [21]	Support Vector Machine (SVM) Classifier	82.59 Percentage
7	Alec Go et al. [22]	Support Vector Machine (SVM) Classifier	81 Percentage
8	Jagdish Chandra Patni et al. [23]	k-means++ over k-means	76.8 Percentage
9	Kalpdrum Passi and Niravkumar Pandey [24]	Random Forest	90.74% for predicting runs scored by a batsman and 92.25% for predicting wickets
10	Nitish Ranjan Bhowmik et al. [25]	Support Vector Machine (SVM) Classifier	82.2Percentage

4. Applications Of Sentiment Analysis:

Sentiment analysis offers a wide range of applications in a variety of fields. Such as:-

➢ Reviews from Websites (On practically any topic, the Internet now provides a significant collection of reviews and feedback. This covers product reviews, political input, service complaints, and so on. About a particular product or service is needed a sentiment analysis system capable of extracting sentiments accordingly. It will assist us in automating the process of providing comments or ratings for a certain product, item, or other object.

➢ Sub-component Technology (In recommender systems, a sentiment predicting system can be useful. Items with a lot of negative comments or low ratings will not be recommended by the recommender system.)

➢ In Business Intelligence (For many firms, the success or failure of their product is determined by online opinion. As a result, sentiment analysis is critical in the corporate world. Businesses also want to extract sentiment from online evaluations to enhance their products and, as a result, their reputation and customer happiness.)

➢ Sentiment Analysis has benefited current sociological research as well as other fields such as medicine and sports by showing trends in human emotions, notably on social media.

➢ Smart homes are expected to be the technology of the future.

➢ There has been a lot of research recently on the Internet of Things (IoT). IoT would also benefit from sentiment analysis. For example, the home might change its atmosphere to provide a calming and peaceful environment based on the user's present attitude or emotion.)

5. CONCLUSIONS:

This paper includes various papers supporting different techniques applied on online reviews in order to find polarity or to classify those review into various sentiments. Finally, sentiment analysis has been shown to be a valuable source of information mining, particularly in circumstances when a big amount of data about the public is required, such as examining public behaviour toward various scenarios.

REFERENCES :

1. Bharati S. Kannolli and Prabhu R. Bevinmarad, "Analysis and Prediction of Sentiments for Cricket Tweets Using Hadoop", International Research Journal of Engineering and Technology (IRJET) Volume 04, Issue 10, Oct - 2017.
2. Yuling Chen and Zhi Zhang, "Research on text sentiment analysis based on CNNs and SVM",13th IEEE Conference on Industrial Electronics and Applications (ICIEA), IEEE, 2018.
3. Peter D. Turney, "Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews," Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL), Philadelphia, July 2002, pp. 417-424.
4. VarshaSahayak, VijayaShete, Apashabi Pathan,"Sentiment Analysis on Twitter Data," International Journal of Innovative Research in Advanced Engineering (IJIRAE), Issue 1, Volume 2, January 2015, pp. 178-183.
5. RazaUl Mustafa, M. Saqib Nawaz, M. Ikram UllahLali, Tehseen Zia, Waqar Mehmood, "Predicting The Cricket Match Outcome Using Crowd Opinions On SocialNetworks: A Comparative Study Of Machine Learning Methods,"Malaysian Journal of Computer Science. Vol. 30(1), 2017, pp. 63-76.
6. V.K. Singh, R. Piryani, A. Uddin P. Waila,"Sentiment Analysis of Movie Reviews and Blog Posts," 3rd IEEE International Advance Computing Conference (IACC), IEEE, 2013, pp. 893-898.
7. Zeenia Singla, Sukhchandan Randhawa, and Sushma Jain, "Statistical and Sentiment Analysis of Consumer Product Reviews", 8th International Conference on Computing, Communication and Networking Technologies (ICCCNT), IEEE, 2017.
8. Anh-Dung Vo ,Quang-Phuoc Nguyen , and Cheol-Young Ock, "Opinion_Aspect Relations In Cognizing Customer Feelings Via Reviews", IEEE Access (Volume: 6), IEEE, 2018.
9. Yan Wan, HongzhuruiNie, TianguangLan, Zhaohui Wang, "Fine-grained Sentiment Analysis Of Online Reviews", 12th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD), IEEE, 2015.
10. DurgeshSamariya, Ajay Matariya, DhwaniRaval, L.D. Dinesh Babu, Ebin Deni Raj and Bhavesh Vekariya, "A Hybrid Approach for Big Data Analysis of Cricket Fan Sentiments in Twitter", Proceedings of International Conference on ICT for Sustainable Development, Advances in Intelligent Systems and Computing 408, 2016.
11. Samuel, J. Nawaz Ali, G.G.M.; Mokhlesur Rahman, M.; Esawi, E.; Samuel, Y., "COVID-19 Public Sentiment Insights and Machine Learning for Tweets Classification", arXiv, 2020.
12. Behdenna, S. Barigou, F. Belalem, G., "Sentiment Analysis of Arabic Tweets: Opinion Target Extraction Sentiment Analysis of Arabic Tweets: Opinion Target Extraction", Int. Symp.Model. Implement. Complex System 2018.

13. S Anjali, V Aswini, M Abirami, "Predictive Analysis With Cricket Tweets Using Big Data", International Journal of Scientific & Engineering Research, Volume 6, Issue 10, October-2015.

14. Sumayh S. Aljameel, Dina A. Alabbad, Norah A. Alzahrani, Shouq M. Alqarni, Fatimah A. Alamoudi, Lana M. Babili, Somiah K. Aljaafary and Fatima M. Alshamrani, "A Sentiment Analysis Approach to Predict an Individual's Awareness of the Precautionary Procedures to Prevent COVID-19 Outbreaks in Saudi Arabia", International Journal of Environment Research and Public Health 2021, 18, 218.

15. Tarun R R, Sahana J S, Sadvik B S, Shashank S, Prof. Mahesh T R, "Context Based Sentiment Analysis Of Twitter Using Hadoop Framework", International Journal of Computer Science and Mobile Computing, Vol.8 Issue.5, May- 2019, pg. 193-202.

16. R. Xia, C. Zong, and S. Li, "Ensemble of feature sets and classification algorithms for sentiment classification, "Information Sciences: an International Journal, vol. 181, no. 6, pp. 1138–1152, 2011.

17. Vishal A. Kharde and S.S. Sonawane, "Sentiment Analysis of Twitter Data: A Survey of Techniques", International Journal of Computer Applications, Volume 139 – No.11, April 2016.

18. V. M. K. Peddinti and P. Chintalapoodi, "Domain adaptation in sentiment analysis of twitter," in Analyzing Microtext Workshop, AAAI, 2011.

19. Liu, S., Li, F., Li, F., Cheng, X., &Shen, H..Adaptive containing SVM for sentiment classification on tweets. In Proceedings of the 22nd ACM international conference on Conference on information & knowledge management (pp. 2079-2088). ACM, 2013.

20. Pan S J, Ni X, Sun J T, et al. "Cross-domain sentiment classification via spectral feature alignment". Proceedings of the 19th international conference on World Wide Web. ACM, 2010: 751-760.

21. Sandeep Nigam, Ajit Kumar Das and Rakesh Chandra Balabantaray, "Machine Learning Based Approach To Sentiment Analysis", International Conference on Advances in Computing, Communication Control and Networking (ICACCCN2018), IEEE, 2018.

22. Alec Go, Lei Huang, and Richa Bhayani, "Twitter sentiment analysis", Final Projects from CS224N for Spring 2008/2009 at The Stanford Natural Language Processing Group, 2009.

23. Jagdish Chandra Patni, Ravi Tomar, Mahendra Singh Aswal, Ankur Dumka, "Prediction Based on Sentiment Analysis of Game", International Journal of Current Engineering and Scientific Research (IJCESR), Volume-4, Issue-10, 2017.

24. Kalpdrum Passi and Niravkumar Pandey, "Increased Prediction Accuracy In The Game of Cricket Using Machine Learning", International Journal of Data Mining & Knowledge Management Process (IJDKP) Vol.8, No.2, March 2018

25. Nitish Ranjan Bhowmik, Mohammad Arifuzzaman, M. Rubaiyat Hossain Mondal, M. S. Islam, "Bangla Text Sentiment Analysis Using Supervised Machine Learning with Extended Lexicon Dictionary", Natural Language Processing Research,Vol. 1(3-4), 2021, pp. 34–45.