International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 05 Issue: 04 | April - 2021 ISSN: 2582-3930

Sentiment Analysis Using Twitter Dataset to Predict IPL

Muthuprakash B[1], Naveen A[2], Sakthi M[3], Prof Archana M[4]
[1, 2, 3] Dept of Computer Science and Engineering [4]Assistant Professor,
Dept of Computer Science and Engineering
[1, 2, 3, 4] Adhiyamaan College of Engineering, Hosur, India.

Abstract-*Social networking site like twitter is an platform that is used by people to maintain their social relations, managing contact, sharing thoughts and emotions by tweets etc. Between many social media site one of the most used social media i.e. twitter, that allows users to post the various activities from their life. As we already know that twitter post is become most favorite platform for users to share their thoughts. But there are also some limitation on twitter for posting the tweets like their is word limit while posting the tweets so to detect opinion mining of the peoples related to any particular event. Tweet post by user helps to recognize the thoughts of people behind that tweet like the IPL 2016.Paper presents, and gives the uses and evaluation of a machine learning models as positive or negative sentiments on twitter data that is tweets. The collection of tweets which get processed after applying the algorithm on that which is depend on the hashtag such as (#IPL and #IPLTEAM) which can be done through using of Twitter 's API (Application Programming Interface) service. They analyze performance of Random* Forest algorithm with already implemented supervised machine learning algorithms with respect to its accuracy, precision, sensitivity etc.

Keywords*-Sentiment, Follow, SENTDIFF, polarity, TwitterAPI, ,NaïveBayes, ,Support Vector Machine.*

I. INTRODUCTION

Social media is an internet-based form of communication. Social media platforms allow users to have conversations, share information and create web content. There are many forms of social media, including blogs, micro-blogs, wikis, social networking sites, photo-sharing sites, instant messaging, video-sharing sites, podcasts, widgets, virtual worlds, and more. Twitter is a micro-blogging site that allows people to post updates in 140 characters or less. Departments looking to engage their audience at a high frequency and have the resources to respond promptly should consider using Twitter. UCM maintains the twitter field, and frequently posts information about campus events, university research, student accomplishments, USF news stories and more. Twitter is open source so their is no restriction while creating the twitter account anyone who wants to create the twitter account they can create it. With twitter, the steps is that you ―follows people‖ and have people ―follow you back‖. When you write a message to post, your followers see It in news field, and when you visit twitter account, you see the messages of people who you're following back. You can also ―retweet‖ messages which means you want to repeat what someone else has said because you want to your followers read it, and you can name people in your tweets by adding @ symbol where you put their twitter name, or "handle" after the @ sign so that they're told via their account that you've mentioned their name in them. If you want to communicate with people without it being public, you can send short private messages to them. You can also use the keywords or phrases in your tweets by using a hash tag start of them (eg. #data) and then that lets twitter monitor ―trending topics among them‖ – keywords that are being talked about by a lot of people. If someone clicks on a hashtag such as '#data they'll be taken back to a page

you can also ―retweet‖ messages which means you want to repeat what someone else has said because you want to your followers read it, and you can name people in your tweets by adding @ symbol where you put their twitter name, or "handle" after the @ sign so that they're told via their account that you've mentioned their name in them. If you want to communicate with people without it being public, you can send short private messages to them. You can also use the keywords or phrases in your tweets by using a hash tag start of them (eg. #data) and then that lets twitter monitor ―trending topics among them‖ – keywords that are being talked about by a lot of people. If someone clicks on a hashtag such as '#data they'll be taken back to a page

Displaying other tweets with the hashtag #data so they can see what other people are saying on the topic. With the purpose of advertising of any things, you can post a tweetregarding that thing so that it comes to knows to people who may be interested in your products. With a little bit of a push, this can then become a trending topic. This is the ways that the old spice videos became a viral successfully – it wasn't just PR orthe fact that they're very nice and well put all this ideas together, it was

paying money to the twitter accounts API to push them up to promote their products . SM sites are basically known for Information dissemination ,opinion/sentiment expression, and product reviews.

News alerts, breaking news, political debates and government policy and various other topics are also discussed and analysed on SM sites. However, while some opinions on SM guides users and other organizations to make powerful and beneficial decisions, which helps to organization to boost the growth of their organization. Users opinions/sentiments on SM such as Twitter, Facebook, YouTube and Yahoo etc are basically positive, negative or neutral [2]. Recent studies of the opinion mining have proven that with Twitter social media it is possible to get people‘s views regarding particular product from their profiles in compare to traditional ways of getting the information about particular views. Furthermore, authors of proposed an algorithm for the extract the emotions from tweets posted by the user while including a large number of data for the sentiment analysis.

To recognize social communities with influential impact, a novel method was proposed by and evaluated by assigns metric depend value to each of the user‘s emotional posts. Together, the data involvement includes the analysis of IPL related sentiments collects from the Twitter profiles, with various types of sentiment analyzers. In addition to that, it presents the proper validation of results comes out from each analyzer with the help of machine-learning classifiers. Our analysis is depend on the comparison of the different sentiment analyzers and evaluates the results with the various classifiers.

II. LITERATURE REVIEW

Using one of the factors such as the number of positivenegative, and neutral tweets posted about parties, the effect size of thesetweets (the number of re-tweets), or the number of people who posted these tweets. However, no study was found that used all of these factors together. The goal of this study is to develop a new approach that takes into account all of the described factors and contributes to the literature in this context. A new multifactor model for the IPL result prediction based on Twitter data has been developed for this purpose. This new model uses the Page Ranks centrality of users in Twitter and sentiment scores. Moreover, the forecasting is performed on the medians of multiscore distributions by using the autoregressive fractionally integrated moving average model (FARIMA). The model was tested by attempting to predict the results of the USA 2020 IPLs in November, which had not yet taken place when the first version of this article was written. Also, a comparison has been made with three alternative estimation approaches in the literature and polls. Analyzes were made for more than 10 million tweets collected between September 1, 2020, to November 2, 2020. As a result of the analysis, it has been observed that the newly developed multifactor method gives better results not only than other methods in the literature but also polls. The USA November 2020 presidential IPL results had estimated by only a 1,52%.

The IPL result prediction using sentiment analysis recent year, social media has provided end users a powerful platform to voice their opinions. Opinion of people matters a lot to analyze how the propagation of information impacts the lives in a large-scale network like Twitter. Businesses need to identify the polarity of these opinions in order to understand user orientation and thereby make smarter decisions. One such application is in the field of politics, where political entities need to understand public opinion to determine their campaigning strategy. Twitter is indeed used extensively for political deliberation. We find that the mere number of messages mentioning a party reflects the IPL result. This data is used to predict outcome of IPL by using sentiment analysis. Sentiment analysis of the tweets determine the polarity and inclination of vast population towards specific topic, item or entity. Popular text classification algorithms like Naive Bayes and SVM are Supervised Learning Algorithms which require a training data set to perform Sentiment analysis. These algorithms are utilized to build classifier and classified the test data as positive, negative and neutral. A two stage framework can be formed to create a training data from the mined Twitter data and to propose a scalable machine learning model to predict.IPL is conducted to view the public opinion, where group of people choose the candidate by using votes, many methods are used to predict result. many agencies and media companies conduct pre poll survey and expert views to predict result of IPL. We use twitter data to predict outcome of IPL by collecting twitter data and analyze it to predict the outcome of the IPL by analyzing sentiment of twitter data about the candidates. We used lexicon-based approach with machine learning to find emotions in twits and predict sentiment score.

III. EXISTING SYSTEM

Social media has provided end users a powerful platform to voice their opinions. Opinion of people matters a lot to analyze how the propagation of information impacts the lives in a large-scale network like Twitter. Businesses needto identify the polarity of these opinions in order to understanduser orientation and thereby make smarter decisions. One such application is in

the field of politics, where political entities need to understand public opinion to determine their campaigningstrategy. Twitter is indeed used extensively for politicaldeliberation. We find that the mere number of messagesmentioning a party reflects the IPL result. This data is used topredict outcome of IPL by using sentiment analysis.Sentiment analysis of the tweets determine the polarity and inclination of vast population towards specific topic, item or entity. Popular text classification algorithms like Naive Bayes and SVM are Supervised Learning Algorithms which require a training data set to perform Sentiment analysis. These algorithms are utilized to build classifier and classified the test data as positive, negative and neutral. A two-stage framework can be formed to create a training data from the mined Twitter data and to propose a scalable machine learning model to predict the IPL result.

IV. PROPOSED SYSTEM

The proposed model was divided into sub-tasks to make the workflow organized. The first task was to collect the data for our model that could then be processed and worked upon, using twitter streaming API, which gave us the data in JSON format which we converted to CSV format. After the cleaning and pre-processing of tweets, we ran some preliminary analysis (most common words, bigrams, hashtags) to know the basic facts about the data and story generation from the available data. The next step was to perform the sentiment analysis using Text Blob on the cleaned tweets, and analyse and predict the IPL outcomes based on positive/negative tweet percentage. After this we applied community detection algorithm on the data set considering users as nodes and retweets as the edges between them and then identified community leaders to find community orientation, and also found the average distance between the nodes to confirm small world phenomenon in the graphs. In this way our model will predict IPLs.

V. MODULES

A. DATA COLLECTION:

The data collection step is the initial phase in the resarch,where data is collected from the twitter is showing the result in 10 recent tweets.The data collection step is the initial phase in the research, where data is collected from twitter. There are two methods on how to connect and collect tweets from Twitter. The first method is by searching tweets matching to the keywords. The second method is by collecting all the tweets provided by Twitter through streaming API, or all the tweets in a specific language, or all the tweets in a specific location then put all of them into the database.

B. DATA TRAINING:

In this module, the training module requires a feature extraction to be done before hand. This will allow the algorithm to identify the required features, which are required to identify the object, in this case the sentiment of the post. Now comes the training part, for training we use random forest algorithm as it proven to give best results in datasets such as ours. The training phase will use the training part of the dataset and not the testing part. The result will be saved in a model, which is achieved by NLTK library of python. The training phase will be done for 10000 epochs.

C. ALGORITH APPLYING :

Random forest algorithm can be used for both classification and regression tasks in this module. It provides high accuracy through cross validation. Logistic regression analysis is used to examine the association of independent variables with one dichotomous dependent variable. SVM uses a technique called the kernel trick to transform data and then based on these transformations it finds on optimal boundary between the possible outputs.

D. PREDICTION:

The proliferation of social media in the recent past has provided end users a powerful platform to voice their opinions.Businesses (or similar entities) need to identify the polarity of these opinions in order to understand user orientation and thereby make smarter decisions. One such application is in the field of politics, where political entities need to understand public opinion and thus determine their campaigning strategy. Sentiment analysis on social media data has been seen by many as an effective tool to monitor user preferences and inclination. Popular text classification algorithms like Naive Bayes and SVM are Supervised Learning Algorithms which require a training data set to perform Sentiment analysis. The accuracy of these algorithms is contingent upon the quantity as well as the quality (features and contextual relevance) of the labeled training data. Since most applications suffer from lack of training data, they resort to cross domain sentiment analysis which misses out on features relevant to the target data. This, in turn, takes a toll on the overall accuracy of text classification. In this paper, we propose a two stage framework which can be used to create a training data from the mined Twitter data without compromising

on features and contextual relevance. Finally, we propose a scalable machine learning model to predict the IPL results using our two stage framework.

VI. CONCLUSION

Logistic regression and SVM are supervised learningalgorithms used to classify data according to parties. Most of the researches have extracted only twitter data but we can also use other social media sites like Facebook and Instagram to fetch data. The data used so far is in the form of words and sentiment analysis is applied on it to determine the polarity of the word. But applying sentiment analysis on sentences it may provide better results than on words. Final result can be obtained by comparing Sentiment Percent of various team obtained by using above algorithms.

VII. RESULTS

	Tweets
0	DO NOT MISS: The match-defining over from Shah...
1	A few smiles & a handshake as we say good ...
2	Upstox Most Valuable Asset of the Match betwee...
3	READ - RCB produced a monumental comeback with...
4	In the end, it's all about the SpiritOfCricket...
...	...
195	Catchin' up \n\n45 | | VIVOIPL
196	The stage is set for Match 5 of VIVOIPL \n\nKK...
197	Hello and welcome to Match 5 of the VIVOIPL \n...
198	In the VIVIOIPL season opener & on his deb...
199	In first game of the season, _27 \nblazed his...

200 rows × 1 columns

Fig 1. Show 200 collection of data.

	Tweets	Subjectivity	Polarity	Analysis
0	DO NOT MISS: The match-defining over from Shah...	0.571429	0.446429	Positive
1	A few smiles & a handshake as we say good ...	0.350000	0.250000	Positive
2	Upstox Most Valuable Asset of the Match betwee...	0.500000	0.500000	Positive
3	READ - RCB produced a monumental comeback with...	0.650000	0.212500	Positive
4	In the end, it's all about the SpiritOfCricket...	0.000000	0.000000	Negative
...	...	...	...	...
195	Catchin' up \n\n45 | | VIVOIPL	0.000000	0.000000	Negative
196	The stage is set for Match 5 of VIVOIPL \n\nKK...	0.000000	0.000000	Negative
197	Hello and welcome to Match 5 of the VIVOIPL \n...	0.900000	0.800000	Positive

Fig.2. Prediction of tweets

Fig.3. Print 10 positive tweets.

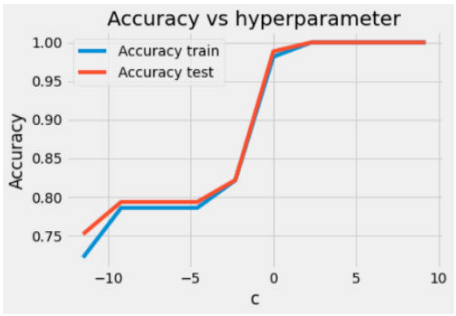

Fig.4. Performance table.

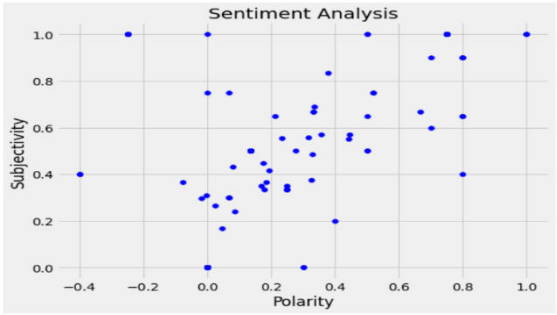

Fig.5. Subjectivity and polarity analysis

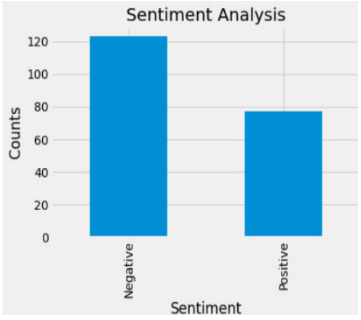

Fig 6. Plot the counts of positive and negative

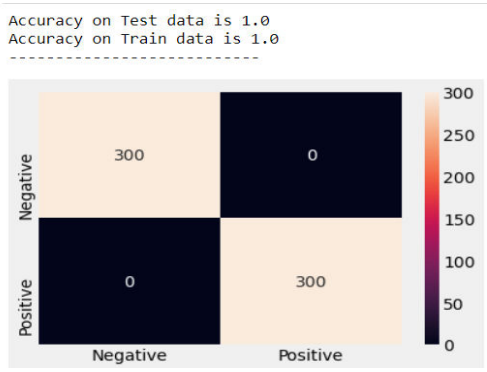

Fig.7. Plot the test and train data

REFERENCES

[1] Ms. Farha Nausheen et al., "Sentiment Analysis to Predict IPL Results Using Python", Proceedings of the Second International Conference on Inventive Systems and Control (ICISC 2018).

[2] Jyoti Ramteke et al., "IPL Result Prediction Using Twitter Sentiment Analysis", 2016 International Conference on Inventive Computation Technologies(ICICT).

[3] Alexander Pak, et al., "Twitter as a Corpus for Sentiment Analysis and Opinion Mining", Proceedings of the Universit e de Paris-Sud, Laboratoire LIMSI-CNRS.K. Elissa, "Title of paper if known," unpublished.

[4] AndranikTumasjan et al., "Predicting IPLs with Twitter: What 140 Characters Reveal about Political Sentiment", Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media.

[5] Parul Sharma et al., "Prediction of Indian IPL Using Sentiment Analysis on Hindi Twitter", 2016 IEEE International Conference on Big Data (Big Data).

[6] PriteeSalunkhe et al., "Twitter Based IPL Prediction and Analysis", International Research Journal of Engineering and Technology (IRJET) Volume: 04 Issue: 10 | Oct -2017.

[7] Amandeep Kaur et al., "Sentiment Analysis On Twitter Using Apache Spark", Carleton University Project Report.

[8] Widodo Budiharto et al., "Prediction and analysis of Indonesia Presidential IPL from Twitter using sentiment analysis", J Big Data (2018) 5:51.

[9]. Alexandre Trilla, Francesc Alias, Sentence-Based Sentiment Analysis for Expressive Text-to-Speech, IEEE TRANSACTIONSON AUDIO, SPEECH, AND LANGUAGE PROCESSING, VOL.21, NO. 2, FEBRUARY 2013.

[10] Rui Xia, Feng X, ChengqingZong, Qianmu Li, Yong Qi, Tao Li,Dual Sentiment Analysis: Considering Two Sides of One Review,IEEE TRANSACTIONS ON KNOWLEDGE AND DATAENGINEERING, VOL.27, NO. 8, AUGUST 2015.

[11].Rasheed M. Elawady, Sherif Barakat, Nora M.Elrashidy,"Different Feature Selection for Sentiment Classification, "International Journal of Information Science and Intelligent System, 3(1): 137-150, 2014.

[12]. Duyu Tang, Bing Qin, Furu Wei, Li Dong, Ting Liu, Ming Zhou,"A Joint Segmentation and Classification Framework for Sentence Level Sentiment Classification,"IEEE/ACM TRANSACTIONSON AUDIO, SPEECH, AND LANGUAGE PROCESSING, VOL.23, NO. 11, NOVEMBER 2015.