

*A RESEARCH PAPER ON  
“ANALYSIS OF IPL MATCH RESULTS USING MACHINE LEARNING  
ALGORITHMS”*

**DEPARTMENT OF COMPUTER SCIENCE  
BIRLA INSTITUTE OF TECHNOLOGY  
MESRA-835215, RANCHI  
Noida Campus**

**AUTHOR  
ARPIT GUPTA**

## ABSTRACT –

Indian Premier League (IPL) is a high-profile T20-type cricket game in India involving top celebrities who finance teams of finest players in the country for usually an interstate championship. Since high-profile persons buy or sell teams with huge financial resources, every one often needs to watch over or predict the public opinion about various teams involved. Public opinions usually come in the form of comments and posts on social media platforms. In this paper, Twitter posts (tweets) related to IPL teams have been analysed to classify them in positive or negative sentiments for facilitating trends understanding in this new and promising type of game.

Sentiment analysis is one of the newest techniques in NLP (AI and ML application). Sentiment analysis is used in many applications for confirmation and analysis. This paper discusses adding feature extraction and combining results to improve performance, from translating thinking theory to thinking algorithms, from the first step of thinking analysis to evaluating thinking prediction. A brief description of complexity.

based neural network sensitivity classifiers is given along with a suitable analysis.

Our hypothesis is that machine learning can be used with high accuracy to separate emotions from Twitter messages. We propose a method to analyse the opinions of the fans and interact with the game. We use information stored on Twitter in the form of messages. Instead of relying on a single machine learning algorithm to predict the outcome of a sporting event, we use at least two machine learning algorithms to compare reality. We use modern classification techniques such as logistic regression and random forest and do a comparative study based on all cricket tweets. The results of the project are provided in the form of a web page using logistic regression to analyse and predict live tweets and random forest to analyse and predict old tweets.

# INTRODUCTION –

## INDIAN PREMIER LEAGUE(IPL)

At the national and international levels, sports have grown significantly in significance. One such game that is recognised as the most popular sport worldwide is cricket. One of the cricket formats recognised by the International Cricket Council is T20(ICC). T20 has achieved great success because of the brief time frame and the excitement it has created. The T20 format served as a fruitful foundation for the IPL, which is currently hailed as the largest advancement in cricket. IPL is an annual competition that is typically held in April and May. In the IPL, each team symbolises a state. T20 cricket's popularity has soared thanks to the IPL.

Social media platforms allow users to exchange and share information and create web content. Social media comes in many forms: blogs, microblogs, wikis, social networking sites, photo-sharing sites, instant messaging, video-sharing sites, podcasts, widgets, and virtual worlds. Twitter is a microblogging site where people can post updates up to 140 characters long. When considering Twitter, departments want a resource that can interact frequently with their audience and respond quickly. UCM has a Twitter account and frequently posts about campus events, campus studies, student achievement, USF news, and more.

The International Premier League (IPL) is the most watched cricket league in the world, and in 2010, it became the first athletic event to be televised live. 13 seasons of the IPL have already been successfully completed since its inception. At the moment, eight clubs square off against one another in a round-robin format throughout the league's phases. The top four clubs in the standings after the league stages are finished are qualified for the playoffs. The winner of the match between the first and second-place teams advances to the championship game, while the losing team is given another chance to do so by taking on the winner of the match between the third and fourth-place teams. The two qualifying teams ultimately competed against one another.

In this study, we looked into a number of factors that might influence an IPL match's outcome while calculating the runs for each ball by using the batsman's previous ball's runs as the labelled data. The proposed forecasting model uses KNN and SVM to achieve the goal of the given challenge. There have not been many studies done in this area of IPL prediction. In our study, we discovered that the work done so far for assessing and forecasting the results of the match is based on data mining. By using the runs the batsman scored on the previous ball as the observed data, our work is novel in that we predict runs for each ball and then check to see if our prediction fits the desired model.

## Keyword –

INDIAN PREMIER LEAGUE, IPL, MACHINE LEARNING, SENTIMENT ANALYSIS, CRICKET, TWITTER, SPORTS, PREDICTION SYSTEM SOCIAL MEDIA

## **PRAPOSED SYSTEM –**

- Naive Bayes classifiers - There are many classifiers for text classification. Naive Bayes classifier is one of them. Naive Bayes classifiers perform well on many complex real-world problems. Develops an eigenvalue-based model that provides tags for questions. There are many algorithms that work the same way, where the value of one property is independent of the value of another property for a different class.
- Support Vector Machine - The support vector machine is nothing but a kernel-related machine and is a significant improvement of the machine learning algorithm. The Support Vector Machine is a supervised learning set that can be easily applied for classification purposes. It shows an updated version of the nonlinear generalized portrait algorithm developed by Vladimir Vapnik. The algorithm adopted by the support vector machine is based on statistical learning theory and dimensionality. A Support Vector Machine [SVM] classified the data by evaluating the N-dimensional hyperplane, which effectively separated the data into two different classes. Even a that doesn't use this option can have very good performance of support vector machines.

## **METHODOLOGY**

### **DATA ACQUISITION –**

This step aims to spot and acquire all data-related problems. In this step, we would like to spot the various data sources, as data can be collected from multiple sources like files and databases. The size and quality of the collected data will determine the efficiency of the output. The more data points, the more accurate the prediction will be.

### **Data Cleaning –**

This step aims to understand the nature of the data that we used to work and to know the characteristics, format, and quality of data. The information needs cleaning and converting into a usable format. It's the method of cleaning the data points, selecting the variable to use, and converting the data into a proper format suitable for analysis within the next step. It's one of the foremost vital steps of the entire process. We used a method called data wrangling that is used for selecting the features to use in the model, converting the acquired data in the dataset to a format that would be suitable for next steps and cleaning of data points.

### **TESTING DATA –**

Test case When an IPL prediction model is trained on a set of parameters, the model is tested using different parameters. In this evaluation step, the model is checked for accuracy and accuracy by providing measurement data. All training methods need to be validated to find the best model to use. After fitting our model to the training data, we use this model to predict the values of the test data. The estimated values of this test data are used for model comparison and accuracy calculations.

## **Future of sentiment analysis –**

Sentiment analysis is a unique and powerful tool for businesses that want to measure their brand's behaviour, thoughts, and feelings. Until now, most strategic planning has been done almost entirely by companies and brands using social media, survey responses, and other user-generated content. By observing and analysing customer sentiment, these brands can understand customer behaviour and ultimately better serve their audiences through the products, services, and information they provide.

The future of sentiment analytics will continue to go far beyond the number of likes, comments and shares to reach and truly understand relationships on social media, and they tell us about the customer behind the screen. The forecast also shows wider use of sentiment analysis, a tool that companies will continue to use but individuals, governments, non-profits, educational institutions, and many organizations will see in the public eye.

## **Conclusion –**

Mumbai Indians are the most successful team in the IPL. Mumbai Indians won most of the shots. Offense won more total games (559) than defence (391). The biggest win while defending the title was 146 runs (the Mumbai Indians beat the Delhi Daredevils by 146 runs on 6 May 2017 at Feroz Shah Kotla Stadium in Delhi). Mumbai city hosts the most IPL matches. Scoring a win gives the opponent a small advantage (52% chance to win). Five Indian players are in the top ten of the IPL rosters. S. Sundaram Ravi officiated most IPL matches on the field. Eden Gardens hosts the most IPL matches.

## Reference –

- Manish Bhelande, Aniket Sanadhya, Mrugang Purao, Abhay Waldia and Vikas Yadav, "Identifying Controversial News using Sentiment Analysis", *Imperial Journal of Interdisciplinary Research (IJIR)*, vol. 3, no. 2, 2017.
- Andreas Adamides and Kampakis, "University College London", *Using Twitter to predict football outcomes*, 2014.
- Galavotti, L., Sebastiani, F., & Simi, M. (2000). Feature selection and negative evidence in automated text...
- Kennedy, Alistair, & Inkpen, Diana (2005). Sentiment classification of movie and product reviews using contextual...
- Luciano Barbosa and Junlan Feng, "Robust sentiment detection on twitter from biased and noisy data", *Proceedings of the 23rd International Conference on Computational Linguistics: Posters*, pp. 36-44, 2010.
- Kevin Gimpel, Nathan Schneider, Brendan O'Connor, Dipanjan Das, Daniel Mills, Jacob Eisenstein, et al., Part-of-Speech tagging for twitter: annotation features and experiments, Pittsburg, USA: CMU.
- Jain, J., et al.: Sentiment analysis using supervised machine learning. *Imp. AI J. Interdiscip. Res.* 2.6 (2016)
- Ch Sai Abhishek, Ketaki V Patil, P Yuktha, Meghana K S, MV Sudhamani, "Predictive Analysis of IPL Match Winner using Machine Learning Techniques", *International Journal of Innovative Technology and Exploring Engineering (IJITEE)* ISSN: 2278-3075, Volume-9 Issue-2S, December 2019.