IPL Match Analysis and Predict Match Results

Gowri Srinivasa
Computer Science and Engineering
PES University
Bangalore, India
gsrinivasa@pes.edu

Ashwin Ashok, Debanik Mishra, Gajendra K.S
Computer Science and Engineering
PES University
Bangalore, India
ashwinashok1998@gmail.com, mishra.debanik@gmail.com, gajendra2ks@gmail.com

Abstract—Cricket is the second most watched sport in the world after soccer, and enjoys a multi-million dollar industry. There is remarkable interest in simulating cricket and more importantly in predicting the outcome of cricket match which is played in three formats namely test match, one day international and T20 match. The complex rules prevailing in the game, along with the various natural parameters affecting the outcome of a cricket match present significant challenges for accurate prediction. Several diverse parameters, including but not limited to cricketing skills and performances, match venues and even weather conditions can significantly affect the outcome of a game. There are number of research paper on pre-match prediction of cricket match. Many papers on building a prediction model that takes in historical match data as well as the instantaneous state of a match, and predict match results. Knowing that cricket match with shorter version match result keep on changing every ball. So, it is important to predict the outcome of the match on every ball. In this paper, I have developed a model that predicts match result on every ball played. Using Duckworth- Lewis formula match outcome will be predicted for live match. For every ball bowled a probability is calculated and probability figure is plotted. For betting industry this model and the probability figure will be very useful for bettor in deciding which team to on and how much to bet.
From casual to ardent cricket fans, the Indian Premier League (IPL) has captured the world's attention in the last decade. It's the shortest format in the game and it's unpredictable nature is what makes it so appealing to the fans.

Index Terms—K-means,

I. INTRODUCTION

Cricket was one of the first sports to use statistics as a tool for illustration and comparison. Although compared to other sports, there has not been much statistical modeling work done for cricket.

Cricket is the sport that is immensely popular in India. The recent format of Twenty20 cricket which was first introduced in 2003 by the England and Wales Cricket Board (ECB), has gained huge recognition in India as well, like several other major cricket playing countries. With India winning the inaugural Twenty20 world cup in 2007 in South Africa, a massive interest in this format of cricket was generated in India.

Soon Subhash Chandra, the promoter of Essel group, started his own private Twenty20 cricket league for Zee TV in 2007 (Malcolm, Gemmell and Mehta, 2009). The name of that tournament was Indian Cricket League (ICL). The league was a six-team competition in its first year (2007) which had expanded to eight in its second year (2008). The tournament drew high profile names to play in its fixtures (Kitchin, 2008). However, the game of cricket got a new dimension in April 2008, when BCCI initiated the Indian Premier League (IPL). It is a Twenty20 cricket tournament being played among eight domestic teams, named after eight Indian cities/states, and owned by franchises (Mitra, 2010). The franchises formed their teams by competitive bidding from a collection of Indian and international players and the best of Indian upcoming talent (Saikia and Bhattacharjee, 2011). Each player has a base price fixed by the IPL authorities. However, there is no upper limit for the bid price. The valuation of players obtained through auction and the availability of players' performances have allowed researchers to infer on different aspects of this format of the game.

II. RELATED WORK

Shubhra Singh and Parmeet Kaur [9] present a data visualization and prediction tool using HBase to keep the data related to IPL (Indian Premier League) cricket matches and players. This data is then visualizes the past performance of players' performance. The data also predicts the outcome of a match through various machine learning approaches.

The tool is used to evaluate the performance of players. This tool provides a visualisation of players' performances. Using IPL T-20 variables related to statistics of batsmen and bowlers, a number of apt variables have been identified that have elucidative power over auction values. Further, several predictive models are also built for predicting the result of a match, based on each player's past performance as well as some match related data.The developed models can help

decision makers during the IPL matches to evaluate the strength of a team against another. The tool employs HBase, a distributed, open source and non-relational database for storing the data. HBase is increasingly being used for hosting of tables with billions of rows and millions of columns. It allows automatic and configurable sharding of tables for scalability of applications, The contributions of the presented work are as follows:

- To provide the statistical analysis of players based on different characteristics.
- To predict the performance of a team depending on individual player statistics.
- To successfully predict the outcome of IPL matches.

Preeti Satao et al [8] built a prediction system that takes in historical match data, player performance as well as the scores predicted by spectator, and predicts future match events culminating in a victory or loss. Our system predicts match outcome by analyzing pre-stored match data using simple but effective K-means clustering algorithm. They described their system and algorithms and finally present quantitative results, demonstrating the performance of our algorithms in predicting the number of runs scored, one of the most important determinants of match outcome.

By using unsupervised learning algorithms, our approach learns a number of features from T-20 cricket dataset which consists of complete records of all games played since the beginning of IPL in the year 2009. For every match that is being played our system predicts the player's score by checking scores from database and also taking into account the scores predicted by fans watching the match. The system outputs a range of probable score that the player will make on that particular round.

Kalpdrum Passi and Niravkumar [4] attempts to predict the performance of players as how many runs will each batsman score and how many wickets will each bowler take for both the teams. Both the problems are targeted as classification problems where number of runs and number of wickets are classified in different ranges. They used naïve bayes, random forest, multiclass SVM and decision tree classifiers to generate the prediction models for both the problems. Random Forest classifier was found to be the most accurate for both the problems.

They have predicted the players' performance in One Day International (ODI) matches by analyzing their characteristics and stats using supervised machine learning techniques. For this, they predict batsmen's and bowlers' performance separately as how many runs will a batsman score and how many wickets will a bowler take in a particular match.

Fahad Munir et al [3] used decision tree algorithm to design our forecasting system by depending on the previous data of matches played between the teams. This system will help the teams to take major decision when the match is in progress such as when to send which batsman or which bowler to bowl in the middle overs. It significantly expands the exposure of research in sports analytics as it was previously bound between some other selected sports.

The aim is to prepare a model which will predict the result of a T20 cricket game while the match is in progress. Our main objective is to combine pre-game data and in-game data in order to design a good predictive model. Understanding the different attributes is also needed in order to get more accuracy in result.

Hemanta Saikia, Dibyojyoti Bhattacharjee and Hoffie Lemmer [7] analyze and predict the performance of bowlers in IPL, using artificial neural network. Based on the performance of bowlers in the first three seasons of IPL, the paper tries to predict the performances of those bowlers who entered in the league in its fourth season as their maiden IPL venture. The performances of these bowlers in IPL-IV are predicted, and the external validity of the model is tested using their actual performance in IPL-IV. This prediction can help the franchises to decide which bowler they should target for their team.

Gunjan Kumar [2] applies descriptive and predictive analysis of soccer matches and player performances. He attempts to identify the important attributes of player's performance. He performed series of classifications with three different strategies and other machine learning algorithms, list the performance metrics for all the four playing positions which approximates the attributes which is used to find the most influential performance metrics of the players for predicting the match outcome. He created 27 different team rating attributes using 8 attributes aggregated over the players ratings and used weighted average method to aggregate team ratings over past match outcome which helped in predicting match outcomes are better characterised by the ratings for the current match.

Deep Prakash, C Patvardhan and Sushobhit Singh [6] implement a machine learning based approach is used to create new index, named as Deep Performance Index (DPI), that reflects the performance of the batsmen and bowlers on a deeper analysis of the requirements of T20 cricket, the recursive feature elimination algorithm used to extract meaningful features and their relative importance towards designing the DPI which captures performance related data for both batsmen and bowlers.

The same authors [5] have developed a Deep Mayo Predictor model for the predicting the outcomes of the matches of IPL season 9 played in the year 2016. Their model has three components which are based on multifarious considerations emerging out of the deep analysis of T20 cricket. The prediction accuracy obtained is high as the model was able to correctly predict the outcomes of 39 matches out of the 56 matches played in the tournament.

Their design of the model for predicting the winners of individual matches is based on few considerations. One approach being to check whether the team as a whole is better than the other team or not. Second approach, was to create a model free estimator based on machine learning on neural network techniques that takes values of individual performance parameters aggregated at the team level and trained to out the result based on these parameters. Third approach was based on comparing the players one on one at corresponding batting

positions in the teams with respect to both batting and bowling strengths.

Madan Gopal Jhawar and Vikram Pudi [1] model the individual player's batting and bowling performances. Player's independent factors have also been considered in order to predict the outcome of the match, showing k-Nearest Neighbors (k-NN) algorithm yeilds better results when compared to other classifiers. Taking two other features, toss decision and the venue of the match along with team strength, they used supervised learning techniques to predict the winner of the match. Their approach lies in addressing the problem as a dynamic one and using the participating players as the key feature in predicting the winner of the match.

They use the player potentials to provide relative dominance one team has over the other.

REFERENCES

[1] M. G. Jhawar and V. Pudi, "Predicting the outcome of odi cricket matches: A team composition based approach," in *European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECMLPKDD 2016 2016)*, 2016.

[2] G. Kumar, "Machine learning for soccer analytics," *University of Leuven*, 2013.

[3] F. Munir, M. K. Hasan, S. Ahmed, S. Md Quraish *et al.*, "Predicting a t20 cricket match result while the match is in progress," Ph.D. dissertation, BRAC University, 2015.

[4] K. Passi and N. Pandey, "Increased prediction accuracy in the game of cricket using machine learning," *arXiv preprint arXiv:1804.04226*, 2018.

[5] C. D. Prakash, C. Patvardhan, and C. V. Lakshmi, "Data analytics based deep mayo predictor for ipl-9," *International Journal of Computer Applications*, vol. 152, no. 6, pp. 6–10, 2016.

[6] C. D. Prakash, C. Patvardhan, and S. Singh, "A new machine learning based deep performance index for ranking ipl t20 cricketers," *International Journal of Computer Applications (0975–8887) Volume*, 2016.

[7] H. Saikia, D. Bhattacharjee, and H. H. Lemmer, "Predicting the performance of bowlers in ipl: an application of artificial neural network," *International Journal of Performance Analysis in Sport*, vol. 12, no. 1, pp. 75–89, 2012.

[8] P. Satao, A. Tripathi, J. Vankar, B. Vaje, and V. Varekar, "Cricket score prediction system (csps) using clustering algorithm."

[9] S. Singh and P. Kaur, "Ipl visualization and prediction using hbase," *Procedia computer science*, vol. 122, pp. 910–915, 2017.