SYNOPSIS

ON

IPL SCORE PREDICTION

Submitted By:

Name: - Devesh Srivastav-201500217

Name: - Shreyash Gaur B- 201500674

Name: - Siddhant Yadav F-201500689

Submitted To:

MS.Neelam Soni Mam

INTRODUCTION

Sports have gained much importance in both national and international level. Cricket is one such game, which is marked as the prominent sports in the world. T20 is one among the forms of cricket which is recognized by the International Cricket Council (ICC). Because of the short duration of time and the excitement generated, T20 has become a huge success. The T20 format gave a productive platform to the IPL, which is now pointed asthe biggest revolution in the field of cricket. IPL is an annual tournament usually played in the month of April and May. Each team in IPL representsa state or a part of nation in India. IPL has taken the T20 cricket's popularityto sparkling heights.

1. It is the most attended cricket league in the world and in the year 2010, IPL became the first sporting event to be broadcasted live. Till date, IPLhas successfully completed 11 seasons from the year of its inauguration.
2. Currently, there are 8 teams that compete with each other, organized in a round robin fashion during the stages of the league. After the completion of league stages, the top 4 teams in the points table are eligible to the playoffs. In playoffs, the winner between 1st and 2nd teamqualifies for the final and the loser gets another opportunity to qualify for the finals by playing against the winner between 3rd and 4th team. In the end, the 2 qualified teams play against each other for the IPL title.
3. The significance is that IPL employs television timeouts and therefore there is no time constraint in which teams as to complete the innings. This game is exceedingly unpredictable because at each phase of the game, the momentum changes to one of the teams between the two. Many times, the results will be decided in the last ball of the match wherethe game gets really closer. Considering all these aspects, there is immense interest among the viewer to make

predictions either at the beginning of the match or during the match.

4. IPL games can’t be easily predicted only by making use of statistics andteams past match’s data. Forecasting from the previous data is highly personalized and requires remarkable expert decisions.

MOTIVATION OF THE PROJECT

The history of machine learning and technology have always been intertwined. Artistic revolutions which have happened in history were made possible by the tools to make the work. We are entering an age wheremachine learning is becoming increasingly present in almost every field.

As the audience of IPL is increasing daily, people are looking at trending technologies like data science, big data to deal with predictions. So we gathered the data from the past seasons and made an analysis on the data. We focused on the factors that are affecting the match winning and startedto predict the match winner using those features.

MODELS FOR PREDICTING RUNS

A. Proposed Solution :

The main objective of this work is to predict the runs using labeled data. A largeamount of data has to be analyzed in order to get a reliable accuracy and hence the first step in achieving the purpose is to gather data of all IPL matches. Data that provides detailed information about each IPL match was collected. Then themandatory data was extracted and refined. The attributes are the gathering of IPL match details which uses less amount of domain knowledge. We decided tomodel the problem as a supervised learning problem since labels can be attachedto match details. The labels are discrete and hence this work can be modeled asa classification problem.

B. Data Collection and Data Preprocessing :

We collected data from dataworld.com which contains details about 577 matches with 21 attributes. Thedataset has been divided into training and testing datasets. The logic behind partitioning the dataset is to give a clear outline of each innings. In this analysis,match details with respect to the 1st innings of all the matches is considered as the training dataset and match details regarding 2nd innings of all matches is considered to be the testing dataset. Match id, inning, batting team, bowling team, over, ball, batsman, bowler, runs etc. are the attributes in the dataset. Current Run Rate (CRR), Required Run Rate (RRR), Wickets taken, Batsman Strike Rate, Bowler Average is the additional attributes that are computed in order to achieve the goal. Runs are considered to be the independent feature and the remaining attributes are taken as the dependent features in the dataset. We have taken runs as the labeled data for predicting the outcome of the match.

C. Calculating Attributes Run Rate:

It is the number of runs that a team scores in one over (6 balls). It

is obtained by dividing the total number of runs scored at some point of time by the number of overs played.

Run Rate = Total number of runs scored/Number of over's bowled

1. Required Run Rate: It is the number of runs per over the batting side must score in order to win the current match [6]. Required Run Rate = Total runs required to win/Total over's left

2. Batsman Strike Rate: It is the average number of runs scored per 100 balls faced [6]. Batsman Strike Rate = (Runs scored/Total balls faced) *100.
3. Bowler Average: It is the number of runs conceded by a bowler per wicket taken. Bowler Average = Runs conceded/Wickets taken

Comparative Analysis

- Native Bayes Classifier.
- Decision Tree.
- Random Forest.

Future Scope

At present, the data is limited to match and score. It doesn't have details about the players and their stats. There is a great scope for applying this concept to the players and their stats data and can findthe batting order and bowling order of a particular match. It will be helpful to franchise people who are at decision making level.

Software Specification:

1. Software: ➢ Python Version 3.0 or above
2. ➢ Flask Framework
3. Operating System: Windows 10 3.
4. Tools: Web Browser (Google Chrome or Firefox)
5. Python Libraries: NumPy, pandas, sklearn, matplotlib, pickle.

Hardware Requirements:

- RAM: 4 GB or above
- Storage: 30 to 50 GB
- Processor: Any Processor above 500MHz