1 INTRODUCTION

1.1 Overview A brief description about your project:

The game of cricket is played in various formats, i.e., One Day International, T20 and Test Matches. The Indian Premier League (IPL) is a Twenty-20 cricket tournament league established with the objective of promoting cricket in India and thereby nurturing young and talented players. The league is an annual event where teams representing different Indian cities compete against each other. It was started by the Board of Control for Cricket in India (BCCI) and has now become a giant, remunerative cricket venture. The teams for IPL are selected by means of an auction. Players' auctions are not a new phenomenon in the sports world. However, in India, selection of a team from a pool of available players by means of auctioning of players was done in Indian Premier League (IPL) for the first time. Due to the involvement of money, team spirit, city loyalty and a massive fan following, the outcome of matches is very important for all stake holders. This, in turn, is dependent on the complex rules governing the game, luck of the team (Toss),the ability of players and their performances on a given day. Various other natural parameters, such as the historical data related to players, play an integral role in predicting the outcome of a cricket match. A way of predicting the outcome of matches between various teams can aid in the team selection process. However, the varied parameters involved present significant challenges in predicting accurate results of a game. Moreover; the accuracy of a prediction depends on the size of data used for the same.

1.2 Purpose of Study:

Betting is an illegal activity some county like India, but most of the country, it’s a profitable business because Betting is not only skill of gambling, it is also a game of statistical skills. Not only betting , there are many fantasy match platform like Dream11,My cricket circle where millions of users invest money to get handsome amount of profits where the statistical skills and data is important. 04 Application of Machine Learning for Predictive Analytics: Indian Premier League (IPL) T-20 Cricket Matches 2 Increased prediction accuracy in the game of cricketusing machine learning

Kalpdrum Passi, Niravkumar Pandey 2018 No. of Innings, Batting Average,Strike Rate,Highes Score,Overs,Bowling Average,Bowling Strike Rate,Four/Five Wicket Haul,Venue,Centuries,Fifties,Batting,MatchTime,Hand,MatchType,BattingPosition,Bowli ng Hand Random Forest builds the most accurate prediction models for both batting and bowling in all the cases. Also, the accuracy of the models increases as increase the size of the training dataset for all algorithms except in case of Naïve Bayes for batting where the accuracydecreases as we increase the size of the training set. Selection of the right players for each match plays a significant role in a team's victory. An accurate prediction of how many runs a batsman is likely to score and how many wickets a bowler is likely to take in a match will help the team management select best players for each match. 3 Prediction of Live Cricket Score and Winning Pramila M. Chawan 2018 Pitch,Toss,Team strength,Home Ground Advantage a predictive model, a user makes a prediction on every game, and ends up watching that game to check if his prediction is going right Thus the project will not only improve the existing system of Fantasy Cricket, but will also augment the reach of Cricket in India Mr. Suyash Mahajan, Ms. Gunjan Kandhari, 4 Cricket Analytics and Predictor Ms. Salma Shaikh, Ms. Rutuja Pawar,Mr. Jash Vora,Ms. A. R. Deshpande 2019 City, Venue, Toss Result, Home Team, Away Team From the previous data, it is beneficial to the owner to get the details of the IPL match played and the users who predict the winning percentage of the team and get the statistics of the player. Figure 1. Comparison of bird image Fig 5 .Detecting Emotion via Azure Figure 2: Working of Blockchain This study primarily aims to find out different statistical measures from IPL historical data and predict outcome of a match based on important factors to help users of betting sites and fantasy cricket league with scientific proof to support in their decision making process

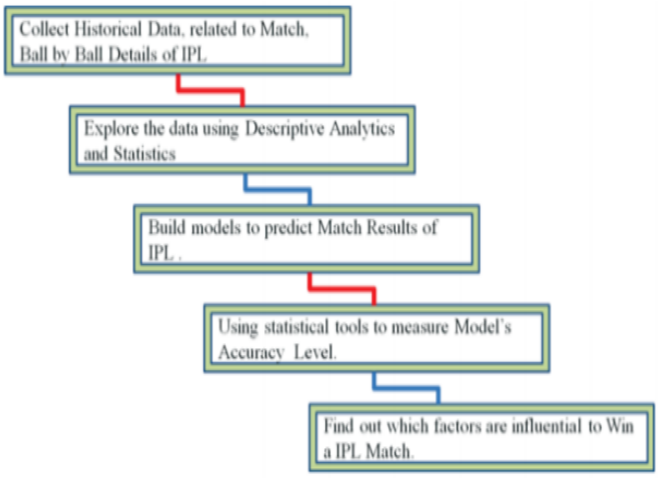
2 LITERATURE SURVEY

The Predicting outcome of Indian Premier league using machine learning by Rabindra Lamsal, Ayesha Choudhary in 2018 in which there are variables utilized like home team, away team, toss winner, venue, umpires, home team score, away team score, power play score, playing 11 players, Number of wickets taken, Number of dot balls given, Number

of fours, Number of sixes, Number of catches, Number of stampings. The Multilayer perception classifier outperformed other classifiers with correctly predicting 43 out of 60, 2018 Indian Premier League matches. The Twenty 20 format of cricket carries a lot of unpredictable, because a single over can change the continued pace of the match. Increased prediction accuracy in the game of cricket using machine learning by Kalpdrum Passi, Nirav kumar Pandey in 2018 in which variables was No. of Innings, Batting Average, Strike Rate, Highest Score, Overs, Bowling Average, Bowling Strike Rate, Four/Five Wicket Haul, Venue ,Centuries, Fifties, Batting, Match Time, Hand, Match Type, Batting Position, Bowling Hand in which analysis was Random Forest builds the most accurate prediction models for both batting and bowling in all the cases. Also, the accuracy of the models increases as increase the size of the training data set for all increase the size of the training data set for all algorithms except in case of Naïve Bayes for batting where the accuracy decreases as we increase the dimensions of the training set. Selection of the proper players for every match plays a big role in a team's victory. An accurate prediction of what percentage runs a batsman is probably going to attain and the way many wickets a bowler is probably going to require match which will help the team to select the players for a specific match. Prediction of Live Cricket Score and Winning by Pramila M. Chawan in 2018 in which variables used was Pitch, Toss, Team strength, Home Ground Advantage in which result was a predictive model, a user makes a prediction on every game, and ends up watching that game to check if his prediction is going right. Thus the project will not only improve the existing system of Fantasy Cricket, but will also augment the reach of Cricket in India. Cricket Analytics and Predictor by Mr. Suyash Mahajan, Ms. Gunjan Kandhari, Ms. Salma Shaikh, Ms. Rutuja Pawar, Mr. Jash Vora, Ms. A. R. Deshpande in 2019 in which variable used was City, Venue, Toss Result, Home Team, Away Team in which the previous data, it is beneficial to the owner to get the details of the IPL match played and the users who predict the winning percentage of the team and get the statistics of the player

3 THEORITICAL ANALYSIS

Methodology :Overall Project Plan (Context Diagram)



Sample Design:

Secondary data is what is collected by someone other than . Some Common sources of secondary data include government public services department's Repository, libraries, internet searches and censuses. For this project, I have used Secondary data source to collect data. Data Source: For this project work, data has been taken from Kaggle.com. Kaggle is subsidiary of Google LLC. It is an online community of data scientists and machine learning aspirant. It is also a repository of open source data. Analytical Methodology: This Project work focus on following two

Analytical Methods *– I. Descriptive Analytics II. Predictive Analytics*

A. Descriptive Analytics: Descriptive Analytics is a Method use in primary stages of any Analytics project to create a summary of historical data to mine useful knowledge, based on which further analysis can be done. In simple language, Descriptive analytics answered question like "what happened?" . In our Project work, Descriptive model focus on two aspects: i. Describe the data statistically. ii. Describe important factors.

B. Predictive Analytics: Predictive Analytics is a Method use in Advance stages of Analytics Projects to Predict Unknown future events based on different factors. Predictive Analytics use different Algorithms to build predictive models. Some of

popular Algorithms, used in Predictive Modelling, are – Logistic Regression, Linear Regression, KNN, Decision Tree, Random Forest, SVM etc. For our project work, we will use predictive analytics to predict result of any matches.

4 FLOWCHART

Data Flow Diagram (DFD):

Tool Used :

A. IBM Cognos: Cognos is a web-based Business Intelligence platform by IBM. It provides a toolset for analytics, reporting, and monitoring of different metrics. The IBM Cognos consists of different components which is used to meet the different information requirements in any company. IBM Cognos has components such as IBM Cognos Framework Manager, IBM Cognos Cube Designer, IBM Cognos Transformer to help to analyse data easily. Cognos will use in this project work as a Descriptive analytics tool.

B. Python: Python is an interpreted, high-level, general-purpose language. Python is developed by Guido van Rossum in 1991.Its language uses objectoriented approach which help programmers to write clearer and logical code for any type of projects. Because of its extensive libraries, Great Community, memory management ,python is very popular among Machine Learning community. For this Project report , Python will

be used for both descriptive and predictive analytics.

C. IBM Watson Studio (Watson Assistant): IBM Watson Studio is an software to make it easy to develop, train, manage models, and deploy AI applications and it used intent,entities and dialog. It is evolving with lot of new features to build Artificial Intelligence applications. IBM Watson Assistant is a cloud service that allows to develop virtual assistantin the software they are developing and brand the assistant as their own.Watson assistant will help to develop a virtual assistant for this project work

5 EXPERIMENTAL INVESTIGATIONS

Analysis and Interpretations

A. Descriptive Analysis Total Match, Session and Different teams, Venue In the Matches Data set, total 756 matches are there and there are total 12 session were played starting from 2008 and the latest session 2019.The bubble chart in dashboard is showing venues with highest session host. Eden gardens hosted 11 sessions, followed by Feroz Shah Kotla, Wankhede, M. Chinnaswamy with 10 session.

Most Successful Team of IPL In terms of Total match winning

Mumbai Indians is the most successful teams who win 109 matches throughout the 11 sessions (2008- 2019). Chennai Super kings is the second most successful team with 100 match win, followed by Kolkata Night Riders with 92 match win. On the other hand Kochi Tuskers Kerala has the least number of match winwith only 6. But one should consider Kochi Tusker Kerala as the most unsuccessful team because Kochi Tusker Kerala played only one session

In terms of Match winning Percentage

In terms of winning percentage, Delhi Capital is the front runner having 62.5% winning rate, Mumbai Indians stood second with 58.3% ratefollowed by Royal Challengers Bangalore (55.6%)

Most Successful Player of IPL In terms of Man of the Match winner

Column chart is showing Chris Gayle own highest number of Man of the Match award followed by AB de Villers and David Warner. In case of Indian players, MS Dhoni top the

list followed by Rohit Sharma.

In terms of Most Run

n terms of most runs, Virat Kohli is the most successful player with 5434 run across 12 session of IPL. Suresh Raina and Rohit Sharma is in 2nd and 3rd position in this list with 5415 and 4914 runs respectively.

In terms of Most Wickets

In terms of wicket, Lasith Malinga is the best bowler with 170 wickets in his name followed by Amit Mishra and Harbhajan Singh with 156 and 149 wickets respectively

Most Economical Bowler

Economy rate is very important in any T20 games . Economy rate indicates the average runs conceded for each over bowledDale Steyn is the most economical bowler with 6.69 economy followed by R Ashwin (6.75) and Sunil Narine(6.79).

Toss's effect on Match wins: Toss win vs. Match win

From the dashboard, it is clear that toss win or loss has least effect to win a match. after winning the toss, team wins the match is higher than when team win the toss but did not manage to win the match,it is clear that Toss Decision's have major impact on Match win. When team choose field, chance of win matches get increase, naturally choosing batting first increase chance of losing a match

6 Result and Analysis

1. In this it is seen the most IPL matches session took place in Eden Garden and Wankhade stadium which is also a important factor for winning the match.

2. Mumbai Indians is the most successful teams who win 109 matches throughout the 11 sessions (2008-2020).Chennai Super kings is the second most successful team with 100 match wins, followed by Kolkata Night Riders with 92 match win. The Mumbai Indians wins mainly the IPL matches.

3. Mostly the man match of the IPL goes to Chris Gaylewho own highest number followed by AB de Villers and David Warner and in Indian playersMS Dhoni top the list.

4. In this it is clear that toss win or loss has least effect to win a match. After winning the toss, team wins the match is higher than when team win the toss but did not

manage to win the match but however the decision to balling or batting first play a major role

5. In different model of machine learning , we find that Decision Tree, Random Forest and Support Vector Machine have the most accuracy model with around 88% accuracy level.

7 Business / Social Impact:

As I am forecasting the predictive analysis of IPL ,it will help the Sponsers to know whom they can sponser and which players they can buy in auction

8 Conclusion:

Analytics can be used for Cricket match Prediction and it's analysis in very easy way. For IPL game,Teams, Venue, Winning Toss, Venue of the Match and Decision after winning the toss are important influencers to win a match. Different Machine Learning helps to predict outcome of a match. Right selection of Machine Learning Model helps to increase Accuracy of Prediction. From Different Classification Models, Support Vector Machine, Decision Tree and Random forest are best to predict outcome of an IPL games. All of the following gives almost 88% accuracy Level. With this we can predict the IPL match through machine learning models.

9.Scope of future study

I have taken only few factors as a predictor .but in cricket there are many factors which could have impact on a Match result. Player is one of that influencer. Every year IPL teams use to change their players; lots of new cricketer gets chances to play for the teams. So, in future studies Player can be use as a Predictor. Another important factor is pitch, in future pitch can be use as predictor. In this study I used only simple popular Classification models, more complex model can be used to increase accuracy in future. Thus using modelling effective decision making can be accomplished through using tools and techniques in sports analytics.