A Technique to Predict Indian Premier League Winner using Artificial Intelligence

Abdullah Umar Nasib
Department of Computer Science and Engineering
BRAC University
Dhaka, Bangladesh
umarnasib13@gmail.com

Ajmain Inqiad Alam
Department of Computer Science and Engineering
BRAC University
Dhaka, Bangladesh
ajmaininqiadalam@gmail.com

Mahfuzur Rahman
Department of Computer Science and Engineering
BRAC University
Dhaka, Bangladesh
mrasif30@gmail.com

Quazi Ahnaf Alif
Department of Computer Science and Engineering
BRAC University
Dhaka, Bangladesh
ahanafalif@gmail.com

***Abstract*—Many researches has been made to predict the first innings run in an ODI cricket match as prediction is an influential think for the team management and their economic outcome. However, in premier leagues like IPL, PSL, BIG BASH this little calculation is tremendously useful in helping the franchise and their owners in terms of their business. In this paper, we have predict the winner team of every match in Indian Premier League(IPL) based on the previous performance of players, toss winner, match venue and the city of the venue. We used decision tree model to predict the winner using the available dataset of previous eight seasons of IPL. The used dataset covers all the match by match and ball by ball details starting from 2008 to 2017. The prediction accuracy of this model was 89.844% where using other algorithms, we achieved poor accuracy rate. We came up with the conclusion that significant facts to decide the winner prediction of an IPL match are 'team1', 'team2', 'city', 'toss_winner', 'venue', 'toss_decision', 'result', 'dl_applied', 'win_by_runs', 'win_by_wickets', 'id'. Based on these analyses, our proposed model will determine the winner of every single match.**

Keywords—IPL; winner prediction; performance analysis; the venue, classifier analysis.

I. INTRODUCTION

Cricket is one of the most played and popular game today in the world. The game has billion of supporters and number of people has been involved with it [5]. Indian subcontinent including India, Pakistan, Srilanka, Bangladesh, Australian arena, and Africa has a tremendous craze for the game. This outdoor game has been played in open field where ball and bat is involved. The game has formats like ODI (One Day International), Test, T20, 6 a side. Apart from Test matches, in all other format, each team gets to bat once and so does ball. However, in test matches, per team get to bat and ball twice.

The game of Cricket is a play of 11 players' team where bat and ball is involved in 22 yard pitch. Batsman tries to hit the ball and score as much as possible where the bowler tries not to give away marks and turning the batsman out following the rules. Each batsman gets to play balls unless gets out. Therefore, if all of them are not out, the team gets to play the full overs which is varies from match format.

T20 is the youngest and most interesting format of cricket. Per team gets to play bat 20 overs where every six legal deliveries count as an over. This format of cricket is less time consuming comparing to ODIs and Test matches. Normally this game ends within 6 hours where each inning takes around 3 hours to finish. Being a shorter duration game, people of diverse sectors can enjoy this game. This version of cricket has a tremendous impact on some countries economy today. India, Australia, Pakistan and many more countries organize their premier league T20s and add a huge amount to their government revenue.

Premier leagues like IPL, Big Bash, PSL and CPL impose a serious impact on the spectators of cricket as well as contribute a healthy amount of revenue in the government fund of the organizing countries. On that note, Indian Premier League (IPL) is the most-attended cricket leagues in the world. The brand value of IPL in 2017 was US$5.3 billion where according to BCCI, the contribution to the GDP of Indian economy in 2015 was US$182 million by the season of 2015[8]. The prize money of IPL is UB$2.3 million where half of the winning prize money is distributed among the payer by rules and rest of the amount goes to the owner of the team. Therefore, winning a season in IPL matter a lot for the players and the franchisees.

In this paper, a method has been demonstrated to predict the winner of a match right after the toss has taken place [1]. We have used a rich dataset containing ball by ball data and match by match data to predict the winner [2]. Before every match we just need to know the humidity, city the match taking place, which team won the toss, decision they have taken after winning the toss. Hence, having the information we analyze our dataset with the proposed method and the method

will let me show the predicting winner of the match. In the dataset, we have got all the details of the ten previous seasons starting from 2008 to 2017.

In this paper, we have followed the following format: In the immediate section after this, we have briefly discussed on the previous researches in the very field to predict the winner. Where in section III, we focused on the proposed model and used algorithms with complete workflow. Hence, under IV which is Training and Testing the model has been talked about where section V is to focus on the comparison and analysis of our proposed model with existing researches. Following that, the conclusion section is been shown in IV no section and the paper concludes with the references we have used in the paper and research work.

II. RELATED WORKS

Only a very few researches has been made predicting the winner of a match in cricket. However, some of the works are well known to predict the total score of an innings. One of them is "Winning And Score Predicting (WASP)", which is a PhD research project by Scott Brooker and Seamus Hogan at University of Canterbury[6]. The work evaluates about how well the normal batting team will do against the normal bowling alley group under given conditions and the present condition of the amusement. In the 1st innings it gauges the extra runs that can be scored with the given number of balls and wickets remaining. In the second innings it predicts the triumphant likelihood with the given number of balls and wickets remaining, runs scored at the given circumstance. The assessments have been produced using a dynamic programming another work from Thapar University, India by Tejinder Singh, Vishal Singla and Parteek Bhatia to predict the winner and the score using data mining entitled "Score and Winning Prediction in Cricket through Data Mining" is a mentionable research in the very field. In that paper they proposed a has two methods, where first predicts the score of first innings not only on the basis of current run rate but also considers number of wickets fallen, venue of the match and batting team [7]. Whereas, second method predicts the outcome of the match in the second innings considering the same attributes as of the former method along with the target given to the batting team. These two methods have been implemented using Linear Regression Classifier and Naïve Bayes Classifier for first innings and second innings respectively.

III. PROPOSED MODEL

This section presents the detail descriptions of proposed model which contains several phases. Fig. 1 represents the whole working procedure of our proposed model.

In the proposed technique we have divided the whole system in two separate part called train and test where the first one to train the code and the second one to check the accuracy of the designed model. Firstly, we had to choose the suitable algorithm for our technique and we choose decision tree method, is one of the most successful techniques for supervised classification learning. A decision tree or a classification tree is a tree in which each internal (non-leaf) node is labeled with an input feature. The arcs coming from a node labeled with a feature are labeled with each of the possible values of the feature. Each leaf of the tree is labeled with a class or a probability distribution over the classes.

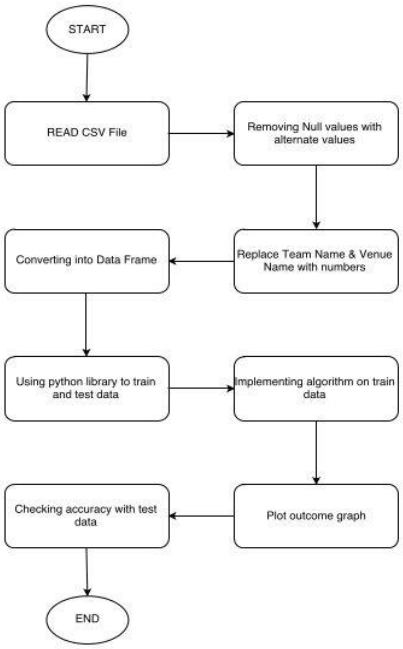
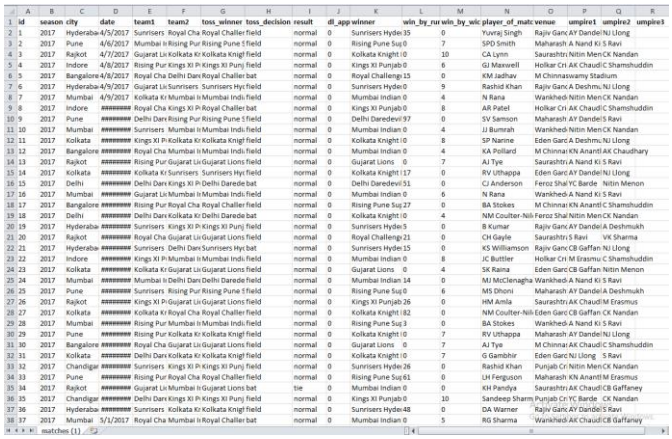


Fig. 1. System workflow

We have taken our dataset shown in Fig. 2 from kaggle.com, a well-known online resource for dataset. Our experimental dataset consists of two files. One file have the ball by ball details of previous 10 seasons IPL matches. Where the second one consists of the detail data of match by match information.

To train our system, we have written dataset readable python code. Before using the dataset, the null values has checked and replaced by identical numerical numbers. Otherwise, it would decrease the accuracy rate of the prediction when testing. After removing the null values, the data has modified and the team names were also replaced by few more identical numbers as followings: 'Mumbai Indians': 1 ,'Kolkata Knight Riders': 2 ,'Royal Challengers Bangalore': 3, 'Deccan Chargers': 4, 'Chennai Super Kings': 5, 'Rajasthan Royals': 6, 'Delhi Daredevils': 7, 'Gujarat Lions': 8, 'Kings XI Punjab': 9, 'Sunrises Hyderabad': 10, 'Rising Pune Supergiant': 11, 'Kochi Tuskers Kerala': 12, 'Pune Warriors': 13. Whereas the venue names are also replaced in the following identical numerical numbers to simplify the process: 'Barabati Stadium': 1, 'Brabourne Stadium':2, 'De Beers Diamond Oval':3, 'Buffalo Park':4, 'Dr DY Patil Sports Academy':5, 'Dr. Y.S. Rajasekhara Reddy ACA-VDCA Cricket Stadium':6, 'Dubai International Cricket Stadium':7, 'Eden Gardens':8, 'Feroz Shah Kotla':9, 'Green Park':10, 'Himachal Pradesh Cricket Association Stadium':11, 'Holkar Cricket Stadium':12, 'JSCA International Stadium Complex':13, 'Kingsmead':14, 'M

Chinnaswamy Stadium':15, 'MA Chidambaram Stadium, Chepauk':16, 'Maharashtra Cricket Association Stadium':17, 'Nehru Stadium':18, 'New Wanderers Stadium':19, 'Newlands':20, 'OUTsurance Oval':21, 'Punjab Cricket Association IS Bindra Stadium, Mohali':22, 'Punjab Cricket Association Stadium, Mohali':23, 'Rajiv Gandhi International Stadium, Uppal':24, 'Sardar Patel Stadium, Motera':25, 'Saurashtra Cricket Association Stadium':26, 'Sawai Mansingh Stadium':27, 'Shaheed Veer Narayan Singh International Stadium':28, 'Sharjah Cricket Stadium':29, 'Sheikh Zayed Stadium':30, "St George's Park": 31, 'Subrata Roy Sahara Stadium':32, 'SuperSport Park':33, 'Vidarbha Cricket Association Stadium, Jamtha':34, 'Wankhede Stadium':35. City names has also replaced with numbers: 'Abu Dhabi':1, 'Ahmedabad':2, 'Bangalore':3, 'Bloemfontein':4, 'Cape Town':5, 'Centurion':6, 'Chandigarh':7, 'Chennai':8, 'Cuttack':9, 'Delhi':10, 'Dharamsala':11, 'Dubai':0, 'Durban':12, 'East London':13, 'Hyderabad':14, 'Indore':15, 'Jaipur':16, 'Johannesburg':17, 'Kanpur':18, 'Kimberley':19, 'Kochi':20, 'Kolkata':21, 'Mumbai':22, 'Nagpur':23, 'Port Elizabeth':24, 'Pune':25, 'Raipur':26, 'Rajkot':27, 'Ranchi':28, 'Sharjah':29, as well as 'Visakhapatnam':30. We also have replaced 'toss_decision' column's value: 'bat':1, 'field':2. The values of 'result is also replaced with numbers: 'no result':1, 'normal':2 'tie':3.



id	season	city	date	team1	team2	toss_winner	toss_decision	result	dl_applied	winner	win_by_runs	win_by_wickets	player_of_match	venue	umpire1	umpire2	umpire3
1	2017	Hyderabad	4/5/2017	Sunrisers Hyderabad	Royal Challengers Bangalore	Royal Challengers Bangalore	field	normal	0	Sunrisers Hyderabad	35	0	Yuvraj Singh	Rajiv Gandhi International Stadium, Uppal	AY Dandekar	NJ Llong	
2	2017	Pune	4/6/2017	Mumbai Indians	Rising Pune Supergiant	Rising Pune Supergiant	field	normal	0	Rising Pune Supergiant	7	0	SPD Smith	Maharashtra Cricket Association Stadium	A Nand Kishore	S Ravi	
3	2017	Rajkot	4/7/2017	Gujarat Lions	Kolkata Knight Riders	Kolkata Knight Riders	field	normal	0	Kolkata Knight Riders	0	10	CA Lynn	Saurashtra Cricket Association Stadium	Nitin Menon	CK Nandan	
4	2017	Indore	4/8/2017	Rising Pune Supergiant	Kings XI Punjab	Kings XI Punjab	field	normal	0	Kings XI Punjab	0	6	GJ Maxwell	Holkar Cricket Stadium	AK Chaudhary	C Shamshuddin	
5	2017	Bangalore	4/8/2017	Royal Challengers Bangalore	Delhi Daredevils	Royal Challengers Bangalore	bat	normal	0	Royal Challengers Bangalore	15	0	KM Jadhav	M Chinnaswamy Stadium			
6	2017	Hyderabad	4/9/2017	Gujarat Lions	Sunrisers Hyderabad	Sunrisers Hyderabad	field	normal	0	Sunrisers Hyderabad	0	9	Rashid Khan	Rajiv Gandhi International Stadium, Uppal	AY Dandekar	NJ Llong	
7	2017	Mumbai	4/9/2017	Kolkata Knight Riders	Mumbai Indians	Mumbai Indians	field	normal	0	Mumbai Indians	0	4	N Rana	Wankhede Stadium	Nitin Menon	CK Nandan	
8	2017	Indore	########	Royal Challengers Bangalore	Kings XI Punjab	Royal Challengers Bangalore	bat	normal	0	Kings XI Punjab	0	8	AR Patel	Holkar Cricket Stadium	AK Chaudhary	C Shamshuddin	
9	2017	Pune	########	Delhi Daredevils	Rising Pune Supergiant	Delhi Daredevils	bat	normal	0	Delhi Daredevils	97	0	SV Samson	Maharashtra Cricket Association Stadium	AY Dandekar	S Ravi	
10	2017	Mumbai	########	Sunrisers Hyderabad	Mumbai Indians	Mumbai Indians	field	normal	0	Mumbai Indians	0	4	JJ Bumrah	Wankhede Stadium	Nitin Menon	CK Nandan	
11	2017	Kolkata	########	Kings XI Punjab	Kolkata Knight Riders	Kolkata Knight Riders	field	normal	0	Kolkata Knight Riders	0	8	SP Narine	Eden Gardens	AY Dandekar	NJ Llong	
12	2017	Bangalore	########	Royal Challengers Bangalore	Mumbai Indians	Mumbai Indians	field	normal	0	Mumbai Indians	0	4	KA Pollard	M Chinnaswamy Stadium	KN Ananthapadmanabhan	AK Chaudhary	
13	2017	Rajkot	########	Rising Pune Supergiant	Gujarat Lions	Gujarat Lions	field	normal	0	Gujarat Lions	0	7	AJ Tye	Saurashtra Cricket Association Stadium	A Nand Kishore	S Ravi	
14	2017	Kolkata	########	Kolkata Knight Riders	Sunrisers Hyderabad	Sunrisers Hyderabad	field	normal	0	Kolkata Knight Riders	17	0	RV Uthappa	Eden Gardens	AY Dandekar	NJ Llong	
15	2017	Delhi	########	Delhi Daredevils	Kings XI Punjab	Delhi Daredevils	bat	normal	0	Delhi Daredevils	51	0	CJ Anderson	Feroz Shah Kotla	YC Barde	Nitin Menon	
16	2017	Mumbai	########	Gujarat Lions	Mumbai Indians	Mumbai Indians	field	normal	0	Mumbai Indians	0	6	N Rana	Wankhede Stadium	A Nand Kishore	S Ravi	
17	2017	Bangalore	########	Rising Pune Supergiant	Royal Challengers Bangalore	Rising Pune Supergiant	field	normal	0	Rising Pune Supergiant	27	0	BA Stokes	M Chinnaswamy Stadium	KN Ananthapadmanabhan	C Shamshuddin	
18	2017	Delhi	########	Delhi Daredevils	Kolkata Knight Riders	Delhi Daredevils	bat	normal	0	Kolkata Knight Riders	0	4	NM Coulter-Nile	Feroz Shah Kotla	Nitin Menon	CK Nandan	
19	2017	Hyderabad	########	Sunrisers Hyderabad	Kings XI Punjab	Sunrisers Hyderabad	field	normal	0	Sunrisers Hyderabad	5	0	B Kumar	Rajiv Gandhi International Stadium, Uppal	AY Dandekar	A Deshmukh	
20	2017	Rajkot	########	Royal Challengers Bangalore	Gujarat Lions	Royal Challengers Bangalore	field	normal	0	Royal Challengers Bangalore	21	0	CH Gayle	Saurashtra Cricket Association Stadium	S Ravi	VK Sharma	
21	2017	Hyderabad	########	Sunrisers Hyderabad	Delhi Daredevils	Sunrisers Hyderabad	bat	normal	0	Sunrisers Hyderabad	15	0	KS Williamson	Rajiv Gandhi International Stadium, Uppal	CB Gaffaney	NJ Llong	
22	2017	Indore	########	Kings XI Punjab	Mumbai Indians	Mumbai Indians	field	normal	0	Mumbai Indians	0	8	JC Buttler	Holkar Cricket Stadium	M Erasmus	C Shamshuddin	
23	2017	Kolkata	########	Kolkata Knight Riders	Gujarat Lions	Gujarat Lions	field	normal	0	Gujarat Lions	0	4	SK Raina	Eden Gardens	CB Gaffaney	Nitin Menon	
24	2017	Mumbai	########	Mumbai Indians	Delhi Daredevils	Delhi Daredevils	field	normal	0	Mumbai Indians	14	0	MJ McClenaghan	Wankhede Stadium	A Nand Kishore	S Ravi	
25	2017	Pune	########	Sunrisers Hyderabad	Rising Pune Supergiant	Rising Pune Supergiant	field	normal	0	Rising Pune Supergiant	0	6	MS Dhoni	Maharashtra Cricket Association Stadium	AY Dandekar	A Deshmukh	
26	2017	Rajkot	########	Kings XI Punjab	Gujarat Lions	Gujarat Lions	field	normal	0	Kings XI Punjab	26	0	HM Amla	Saurashtra Cricket Association Stadium	AK Chaudhary	M Erasmus	
27	2017	Kolkata	########	Kolkata Knight Riders	Royal Challengers Bangalore	Royal Challengers Bangalore	field	normal	0	Kolkata Knight Riders	82	0	NM Coulter-Nile	Eden Gardens	CB Gaffaney	CK Nandan	
28	2017	Mumbai	########	Rising Pune Supergiant	Mumbai Indians	Rising Pune Supergiant	bat	normal	0	Rising Pune Supergiant	3	0	BA Stokes	Wankhede Stadium	A Nand Kishore	S Ravi	
29	2017	Pune	########	Rising Pune Supergiant	Kolkata Knight Riders	Kolkata Knight Riders	field	normal	0	Kolkata Knight Riders	0	7	RV Uthappa	Maharashtra Cricket Association Stadium	AY Dandekar	NJ Llong	
30	2017	Bangalore	########	Royal Challengers Bangalore	Gujarat Lions	Gujarat Lions	field	normal	0	Gujarat Lions	0	7	AJ Tye	M Chinnaswamy Stadium	AK Chaudhary	C Shamshuddin	
31	2017	Kolkata	########	Delhi Daredevils	Kolkata Knight Riders	Kolkata Knight Riders	field	normal	0	Kolkata Knight Riders	0	7	G Gambhir	Eden Gardens	NJ Llong	S Ravi	
32	2017	Chandigarh	########	Sunrisers Hyderabad	Kings XI Punjab	Sunrisers Hyderabad	field	normal	0	Sunrisers Hyderabad	26	0	Rashid Khan	Punjab Cricket Association Stadium, Mohali	Nitin Menon	CK Nandan	
33	2017	Pune	########	Rising Pune Supergiant	Royal Challengers Bangalore	Rising Pune Supergiant	field	normal	0	Rising Pune Supergiant	61	0	LH Ferguson	Maharashtra Cricket Association Stadium	KN Ananthapadmanabhan	M Erasmus	
34	2017	Rajkot	########	Gujarat Lions	Mumbai Indians	Gujarat Lions	bat	tie	0	Mumbai Indians	0	0	KH Pandya	Saurashtra Cricket Association Stadium	AK Chaudhary	CB Gaffaney	
35	2017	Chandigarh	########	Delhi Daredevils	Kings XI Punjab	Kings XI Punjab	field	normal	0	Kings XI Punjab	0	10	Sandeep Sharma	Punjab Cricket Association Stadium, Mohali	YC Barde	CK Nandan	
36	2017	Hyderabad	########	Sunrisers Hyderabad	Kolkata Knight Riders	Kolkata Knight Riders	field	normal	0	Sunrisers Hyderabad	48	0	DA Warner	Rajiv Gandhi International Stadium, Uppal	AY Dandekar	S Ravi	
37	2017	Mumbai	5/1/2017	Royal Challengers Bangalore	Mumbai Indians	Royal Challengers Bangalore	bat	normal	0	Mumbai Indians	0	5	RG Sharma	Wankhede Stadium	AK Chaudhary	CB Gaffaney	

Fig. 2. Dataset contains match by match details

The data has been converted in data frame in this segment to access in details using our python programme. As already shown in Fig. 1, the dataset was divided into two portion called train and test, we did the separation after the data frame was prepared. From the dataset, data of first 9 seasons, means from 2008 to 2016, was used to train the system [12]. We implemented decision tree learning algorithm on the separated dataset.

Decision tree learning algorithm works through recursive partitioning of the training set in order to obtain subsets that are as pure as possible to a given target class [9]. Each node of the tree is associated to a particular set of records T that is splatted by a specific test on a feature.

A. Mathematical Formulation

Given training vectors x_i \in R^n, i=1,…, l and a label vector y \in R^l, a decision tree recursively partitions the space such that the samples with the same labels are grouped together.

Let the data at node m be represented by Q. For each candidate split \theta = (j, t_m) consisting of a feature j and threshold t_m, partition the data into

```
Q_{left}(\theta) and Q_{right}(\theta) subsets
Q_{left}(\theta) = {(x, y) | x_j <= t_m}
Q_{right}(\theta) = Q \setminus Q_{left}(\theta)
```

The impurity at m is computed using an impurity function H (), the choice of which depends on the task being solved (classification or regression).

```
G (Q, \theta) = \frac{n_{left}}{N_m} H (Q_ {left} (\theta)) + \frac{n_{right}}{N_m} H (Q_ {right} (\theta))
```

Select the parameters that minimises the impurity

```
\theta^* = \operatorname{argmin}_\theta G (Q, \theta)
```

Recourse for subsets

```
Q_ {left} (\theta^*) and Q_ {right} (\theta^*)
```

Until the maximum allowable depth is reached,

```
N_m < \min_{samples} or N_m = 1.
```

B. Classification Criteria

If a target is a classification outcome taking on values 0, 1,…, K-1, for node m, representing a region R_m with N_m observations [9], let

$$p_{mk} = 1/N_m \sum_{x_i \in R_m} I(y_i = k) \quad (1)$$

Be the proportion of class k observations in node m
Common measures of impurity are Gini [10]

$$H(X_m) = \sum_k p_{mk}(1 - p_{mk}) \quad (2)$$

Cross-Entropy [11]

$$H(X_m) = -\sum_k p_{mk} \log(p_{mk}) \quad (3)$$

And Misclassification [11]

$$H(X_m) = 1 - \max(p_{mk}) \quad (4)$$

Where X_m is the training data in node m

This is how the decision tree learning algorithm mathematically performs in the library of Python Programming.
Support vector machine (SVM) supports both dense and sparse sample vectors as input. SVM is capable of perform multi classification on a dataset. Given training vectors $x_i \in \mathbb{R}^p$, i=1… n, in two classes, and a vector $y \in \{1, -1\}^n$, SVC solves the following primal problem [12]:

$$\min_{w,b,\zeta} \frac{1}{2} w^T w + C \sum_{i=1}^{n} \zeta_i$$
$$\text{subject to } y_i(w^T \phi(x_i) + b) \geq 1 - \zeta_i,$$
$$\zeta_i \geq 0, i = 1, ..., n \quad (5)$$

Its dual is [12]

$$\min_{\alpha} \frac{1}{2} \alpha^T Q \alpha - e^T \alpha$$
$$\text{subject to } y^T \alpha = 0$$
$$0 \leq \alpha_i \leq C, i = 1, ..., n \quad (6)$$

where e is the vector of all ones, C > 0 is the upper bound, Q is an n by n positive semi definite matrix, $Q_{ij} \equiv y_i y_j K(x_i, x_j)$, where $K(x_i, x_j) = \phi(x_i)^T \phi(x_j)$ is the kernel. Here training vectors are implicitly mapped into a higher (maybe infinite) dimensional space by the function ϕ [13].

The decision function is [13]:

$$\text{sgn}(\sum_{i=1}^{n} y_i \alpha_i K(x_i, x) + \rho) \quad (7)$$

In random forests each tree in the ensemble is built from a sample drawn with replacement from the training set. In addition, when splitting a node during the construction of the tree, the split that is chosen is no longer the best split among all features [14]. Instead, the split that is picked is the best split among a random subset of the features. As a result of this randomness, the bias of the forest usually slightly increases but, due to averaging, its variance also decreases, usually more than compensating for the increase in bias, hence yielding an overall better model [14].

MLP Classifier trains iteratively since at each time step the partial derivatives of the loss function with respect to the model parameters are computed to update the parameters [15].

Given a set of training examples $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ where $x_i \in \mathbf{R}^n$ and $y_i \in \{0, 1\}$, a one hidden layer one hidden neuron MLP learns the function $f(x) = W_2 g(W_1^T x + b_1) + b_2$ where $W_1 \in \mathbf{R}^m$ and $W_2, b_1, b_2 \in \mathbf{R}$ are model parameters W_1, W_2 represent the weights of the input layer and hidden layer, respectively; and b_1, b_2 represent the bias added to the hidden layer and the output layer, respectively. $g(\cdot) : R \to R$ is the activation function, set by default as the hyperbolic tan [16]. It is given as,

$$g(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}} \quad (8)$$

For binary classification, f(x) passes through the logistic function $g(z) = 1/(1 + e^{-z})$ to obtain output values between zero and one [16]. A threshold, set to 0.5, would assign samples of outputs larger or equal 0.5 to the positive class, and the rest to the negative class.

IV. Experimental Setup and Result Analysis

After completing the preprocessing and the training part we did the experimental setup and result analysis. In this very part, we followed the following stages which are testing the system and following that, representing the result graphically and finally analyzing the achieved result:

A. *Testing the system*

We completed the testing segment of our model in this part. With the rest of the data of 2017 IPL matches, we ran our project and tested our proposed model. Based on the pre-set parameters, our proposed model took decision on the above mentioned algorithm and gave us the output of the winner of every single match. We have experimented taking two teams as opponents and provided by the information of the parameters, the system, have shown us the output in graphical representation format.

B. *Result in graphical Representation*

The decision tree learning algorithm gives us the output in form of winner of a match in graphical representation as displayed in Fig. 3. In the shown figure, in y axis: teams name converted into numbers as mentioned earlier 'Mumbai Indians': 1 ,'Kolkata Knight Riders': 2 ,'Royal Challengers Bangalore': 3, 'Deccan Chargers': 4, 'Chennai Super Kings': 5, 'Rajasthan Royals': 6, 'Delhi Daredevils': 7, 'Gujarat Lions': 8, 'Kings XI Punjab': 9, 'Sunrises Hyderabad': 10, 'Rising Pune Supergiant': 11, 'Kochi Tuskers Kerala': 12, 'Pune Warriors': 13. This means Mumbai Indians is team no 1. Therefore, if prediction shows winner team 1 this means Mumbai Indians will win the match. However, in x axis: match number is given.

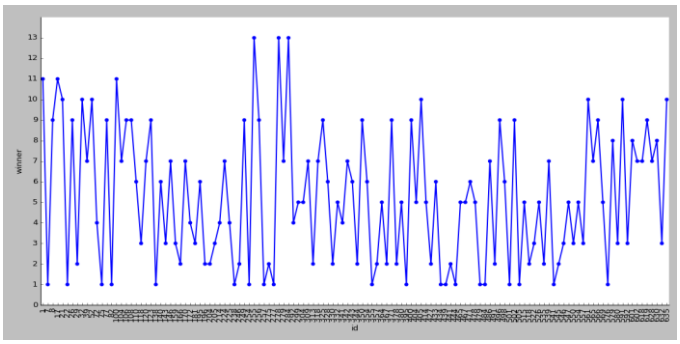


Fig. 3. Graphical representation of prediction using Decision Tree Classifier

In Fig. 3, id 3 means, match no 3. This is how our program gives the output of the predicted winner of every single IPL match.

Random forest classifier algorithm gives name of every match winner which is shown in Fig. 4.

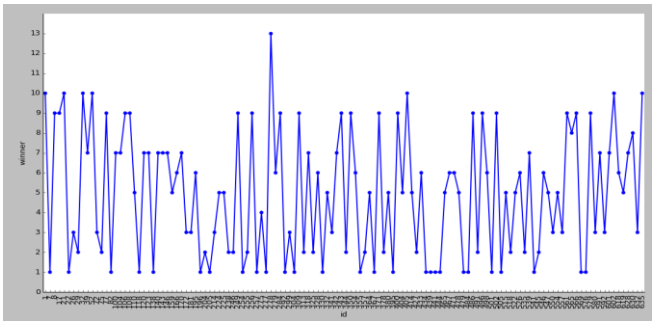


Fig. 4. Graphical representation of prediction using Random Forest Classifier

In this Fig. 4, in y axis, we have team names converted into number as previously discussed and in x axis, we have match id which indicated the match no.

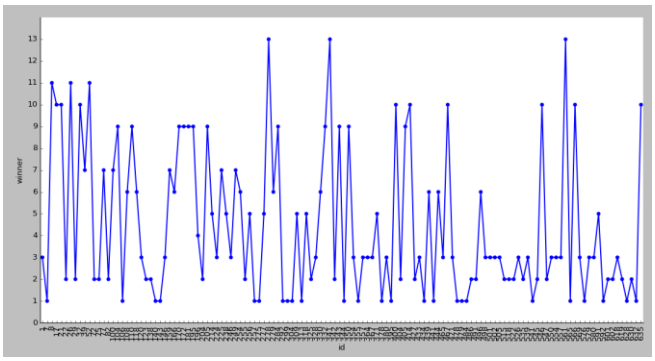


Fig. 5. Graphical representation of prediction using MLP Classifier

In Fig. 5, team names in form of numbers has shown in y axis and match no is shown in x axis. This figure tells that on that match which team is going to win.

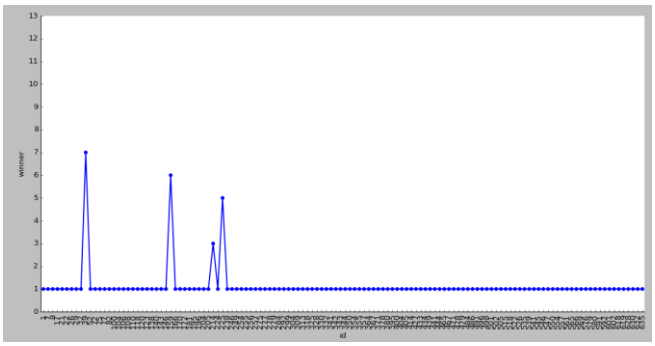


Fig. 6. Graphical representation of prediction using SVM Classifier

In Fig. 6, team names in form of numbers has shown in y axis and match no is shown in x axis. This figure tells that on that match which team is going to win.

C. Result Analysis

The proposed model gives us the accuracy of the match winner of 89.884% using decision tree learning algorithm. However, we also have run few more approach with different algorithms but didn't get an expected or better than this accuracy rate. The experimented algorithms are being noted below:

TABLE I. ALGORITHMS WITH ACCURACY

Serial	Name of Algorithm	Accuracy
01	SVM classifier	17.969%
02	Decision Tree Classifier	89.844%
03	Random Forest Classifier	65.625%
04	MLP Classifier	20.312%

With this experimental result we came to the conclusion that for our proposed model, decision tree approach is the most suitable and applicable method to be chooses.

CONCLUSION AND FUTURE SCOPE

The main purpose of this paper is to demonstrate our proposed model to predict the winner of an Indian Premier League match analyzing the previous ten season match dataset. In this model, decision tree classify method has been used to analysis the data and after the derivation, predict the winner of every single match. This model has some pre-set parameters to decision making criteria fulfilling. Not likely the existing models that can predict the scores of an innings or guess the result of an ODI match, out model can predict the winner of Indian Premier League T20 match winner with an accuracy of 89.844 percent. Since, with the available dataset of IPL; this analysis prediction was made, it's more likely to be possible to predict the winner of all format of games in any tournament on the condition of having a rich dataset [12]. Furthermore, there are other parameters like ranking of the players, performance of bowlers against left or right arm batsman, performance of batman against the off or leg break bowlers which is not available till today. With that dataset this prediction rate accuracy would be more prestige and trustworthy.

REFERENCES

[1] dzone.com. (2018). *Predicting the Outcome of Cricket Matches Using AI - DZone AI*. [online] Available at: https://dzone.com/articles/ipl-cricket-analytics-and-predictive-model [Accessed 26 Mar. 2018].

[2] Spin, D. (2018). *How I Used Machine Learning To Predict Soccer Games For 24 Months Straight*. [online] Doctor Spin. Available at: https://doctorspin.me/digital-strategy/machine-learning/ [Accessed 26 Mar. 2018].

[3] Medium. (2018). *Betting: Football Chat now with AI predictor – Football Score Chat – Medium*. [online] Available at: https://medium.com/@ScoreChat/betting-football-chat-now-with-ai-predictor-f1f4b922d4d0 [Accessed 26 Mar. 2018].

[4] Nyquist, R. and Pettersson, D. (2017). *Football match prediction using deep learning*. [online] Chalmers studentarbeten. Available at: http://studentarbeten.chalmers.se/publication/250411-football-match-prediction-using-deep-learning [Accessed 26 Mar. 20.

[5] Khabir Uddin Mughal. Top 10 Most Popular Sports In The World. http://sporteology.com/top-10-popular-sports-world/[Accessed 2 Feb 2018.].

[6] I. Bhandari, E. Colet, and J. Parker. Advanced Scout:Data mining and knowledge discovery in NBA data. Data Mining and Knowledge Discovery, 1(1):121{125,1997.

[7] S. Tejinder, S. Vishal, B. Parteek. Score and Winning Prediction in Cricket through Data Mining. 2015 International Conference on Soft Computing Techniques and Implementations- (ICSCTI) Department of ECE, FET, MRIU, Faridabad, India, Oct 8-10, 2015.

[8] Thenewsminute.com. (2018). [online] Available at: https://www.thenewsminute.com/article/kerala-beat-west-bengal-dramatic-shootout-win-santosh-trophy-6th-time-78842 [Accessed 1 Apr. 2018].

[9] L. Breiman, J. Friedman, R. Olshen, and C. Stone. Classification and Regression Trees. Wadsworth, Belmont, CA, 1984.

[10] J.R. Quinlan. C4. 5: programs for machine learning. Morgan Kaufmann, 1993

[11] T. Hastie, R. Tibshirani and J. Friedman. Elements of Statistical Learning, Springer, 2009.

[12] I. Guyon, B. Boser, and V. Vapnik, "Automatic Capacity Tuning of Very Large VC-dimension Classifiers," *Advances in Neural Information Processing Systems*, pp. 147–155, 1993.

[13] C. Cortes and V. Vapnik, "Support-vector networks," Machine Learning, vol. 20, no. 3, pp. 273–297, 1995.

[14] L. Breiman, "Random Forests", Machine Learning, 45(1), 5-32, 2001.

[15] G. E. Hinton, "Connectionist learning procedures," *Artificial Intelligence*, vol. 40, no. 1-3, pp. 185–234, 1989.

[16] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Nature*, vol. 323, no. 6088, pp. 533–536, 1986.