



**IJIRCCCE**

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

**Volume 10, Issue 6, June 2022**

**ISSN** INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA

**Impact Factor: 8.165**



9940 572 462



6381 907 438



ijircce@gmail.com



www.ijircce.com

# Match Winning Prediction and Cricket Analysis Using Machine Learning

M.Ram Bhupal<sup>1</sup>, G.Likhitendra<sup>2</sup>, B.John Sunil<sup>3</sup>, G.Eshwar Choudary<sup>4</sup>

Assistant Professor, Dept. of Information Technology, Vasireddy Venkatadri Institute of Technology, Guntur,  
Andhra Pradesh, India<sup>1</sup>

B.Tech Student, Dept. of Information Technology, Vasireddy Venkatadri Institute of Technology, Guntur,  
Andhra Pradesh, India<sup>2,3,4</sup>

**ABSTRACT:** Our System will predict the outcome of the matches played in the past years. It also predicts other attributes such as match final score. It is done using various algorithms. The toss outcome was also predicted. We intend to use more features such as venue, toss outcome, toss winning team etc.. This is all evaluated by using algorithms with the highest accuracy.

## I. INTRODUCTION

Cricket is a bat-and-ball game played between two teams of eleven players on a field at the center of which is a 22-yard (20-meter) pitch with a wicket at each end, each comprising two bails balanced on three stumps. The batting side scores runs by striking the ball bowled at the wicket with the bat (and running between the wickets), while the bowling and fielding side tries to prevent this (by preventing the ball from leaving the field, and getting the ball to either wicket) dismiss each batsmen (so they are “out”). Means of dismissal include being bowled, when the ball hits the stumps and dislodges the bails, and by the fielding side either catching the ball after it is hit by the bat and before it hits the ground, or hitting a wicket with the ball before a batter can cross the crease in front of the wicket. When ten batsmen have been dismissed, the innings ends and the teams swap roles. The game is adjudicated by two umpires, aided by a third umpire and the match referee in international matches. A cricket field is a large grass field on which the game of cricket is played.

Though it is a fun machine learning project, it can be extended to corporate level also, where Sporting channels would like to show the metrics which can be used to increase the audience pulse, which shows the factors which can change even a losing game inclined to winning. It is used for creating online polls, Sports Business.

## II. PROPOSED METHODOLOGY

The proposed system, unlike the currently available systems, focuses on predicting the best playing 11 instead of only predicting the performances of a particular player. The system makes use of Random Forest Algorithm which has been found to be the best and most adaptable for this model. The system will consider team1, team2, toss\_winner, toss\_decision, venue, winner as main parameters for IPI, T20. In the same way along with above attributes it will also consider team1\_player1\_bat\_rank, team1\_player1\_bowl\_rank, team1\_player2\_bat\_rank, team1\_player2\_bowl\_rank are also considered from the dataset. While coming to football home\_team, away\_team, home\_score, away\_score, winner are the attributes that are considered. Further, cleaned data is split into training (95%) and testing data (5%). Training data is fed to all machine learning models and accuracy of each model is noted. Model with the highest accuracy is selected for further prediction. This model is used to make predictions over the input data provided, input data may include – home team, opposition team, current weather condition and analysis done on historical data.

## III. MACHINE LEARNING MODEL MODELING

In this paper, we propose a machine learning model scheme that tests every machine learning algorithm and calculates the accuracy score and selects the algorithm with the highest score.



```
In [32]: from sklearn.svm import SVC
model1 = SVC().fit(x_train, y_train)
model1.score(x_test, y_test)
```

Out[32]: 0.519298245614035

```
In [33]: from sklearn.tree import DecisionTreeClassifier
model2 = DecisionTreeClassifier().fit(x_train, y_train)
model2.score(x_test, y_test)
```

Out[33]: 0.5368421052631579

```
In [34]: from sklearn.ensemble import RandomForestClassifier
model3 = RandomForestClassifier(n_estimators=250).fit(x, y)
model3.score(x_test, y_test)
```

Out[34]: 0.7859649122807018

```
In [35]: from sklearn.ensemble import GradientBoostingClassifier
model3=GradientBoostingClassifier(n_estimators=250).fit(x_train,y_train)
model3.score(x_test,y_test)
```

Out[35]: 0.5789473684210527

```
In [36]: from sklearn.linear_model import LogisticRegression
model4=LogisticRegression().fit(x_train,y_train)
model4.score(x_test, y_test)
```

Out[36]: 0.5578947368421052

The test results show that the Random Forest classifier has more accuracy than others.

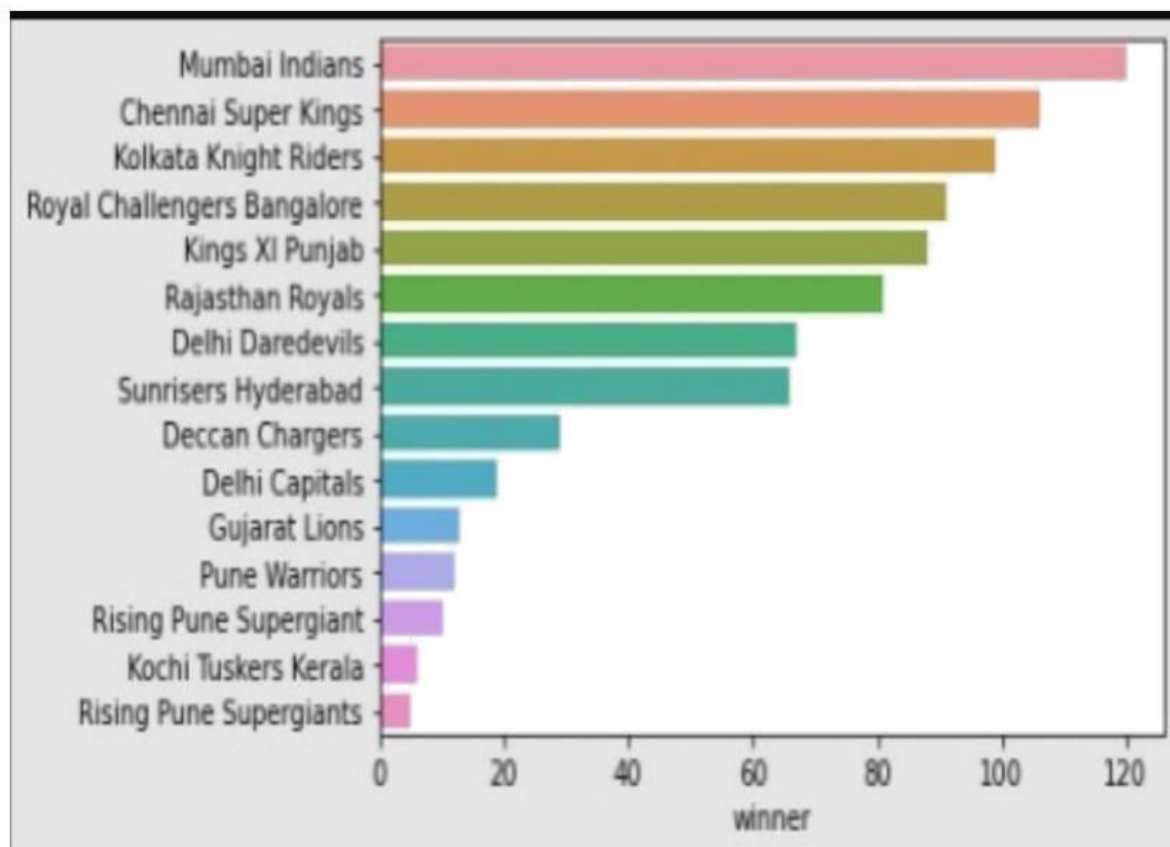
The input to the model is CSV(Comma Separated Value) file that has various attributes such as team1 , team2 , toss\_decision , toss\_winner , venue and others..

id	city	date	player_of_venue	neutral_ve	team1	team2	toss_winn	toss_decis	winner	result	result_ma	eliminator	method	umpire1	umpire2	Season	
1	Bangalore	#####	BB McCullin	M Chinnas	0	RCB	KKR	RCB	field	KKR	runs	140	N	NA	Asad Rauf	RE Koertze	2008
2	Chandigarh	#####	MEK Husse	Punjab Cric	0	KXIP	CSK	CSK	bat	CSK	runs	33	N	NA	MR Benso	SL Shastri	2008
3	Delhi	#####	MF Mahar	Feroz Sha	0	DD	RR	RR	bat	DD	wickets	9	N	NA	Aleem Dar	GA Pratapl	2008
4	Mumbai	#####	MV Bouch	Wankhede	0	MI	RCB	MI	bat	RCB	wickets	5	N	NA	SJ Davis	DJ Harper	2008
5	Kolkata	#####	DJ Hussey	Eden Gard	0	KKR	Dec	Dec	bat	KKR	wickets	5	N	NA	BF Bowder	K Harihara	2008
6	Jaipur	#####	SR Watson	Sawai Mar	0	RR	KXIP	KXIP	bat	RR	wickets	6	N	NA	Aleem Dar	RB Tiffin	2008
7	Hyderabad	#####	V Sehwag	Rajiv Gand	0	Dec	DD	Dec	bat	DD	wickets	9	N	NA	IL Howell	AM Sahebi	2008
8	Chennai	#####	ML Hayden	MA Chidar	0	CSK	MI	MI	field	CSK	runs	6	N	NA	DJ Harper	GA Pratapl	2008
9	Hyderabad	#####	YK Pathan	Rajiv Gand	0	Dec	RR	RR	field	RR	wickets	3	N	NA	Asad Rauf	MR Benso	2008
10	Chandigarh	#####	KC Sangakl	Punjab Cric	0	KXIP	MI	MI	field	KXIP	runs	66	N	NA	Aleem Dar	AM Sahebi	2008
11	Bangalore	#####	SR Watson	M Chinnas	0	RCB	RR	RR	field	RR	wickets	7	N	NA	MR Benso	IL Howell	2008
12	Chennai	#####	JDP Oram	MA Chidar	0	CSK	KKR	KKR	bat	CSK	wickets	9	N	NA	BF Bowder	AV Jayapre	2008
13	Mumbai	#####	AC Gilchris	Dr DY Patil	0	MI	Dec	Dec	field	Dec	wickets	10	N	NA	Asad Rauf	SL Shastri	2008
14	Chandigarh	#####	SM Katic	Punjab Cric	0	KXIP	DD	DD	bat	KXIP	wickets	4	N	NA	RE Koertze	I Shivr	2008
15	Bangalore	#####	MS Dhoni	M Chinnas	0	RCB	CSK	CSK	bat	CSK	runs	13	N	NA	BR Doctro	RB Tiffin	2008
16	Kolkata	#####	ST Jayasur	Eden Gard	0	KKR	MI	KKR	bat	MI	wickets	7	N	NA	BF Bowder	AV Jayapre	2008
17	Delhi	#####	GD McGra	Feroz Sha	0	DD	RCB	RCB	field	DD	runs	10	N	NA	Aleem Dar	I Shivr	2008
18	Hyderabad	#####	SE Marsh	Rajiv Gand	0	Dec	KXIP	KXIP	field	KXIP	wickets	7	N	NA	BR Doctro	RB Tiffin	2008
19	Jaipur	#####	SA Asnodk	Sawai Mar	0	RR	KKR	RR	bat	RR	runs	45	N	NA	RE Koertze	GA Pratapl	2008

The Output is quantitative results and can be represented using seaborn and matplotlib.



### 6.1.2 Most Winning Team



## IV. SYSTEM OVERVIEW

The phase of the project structure was designed to enable the research team to build iteratively of findings from each stage and to use this to guide subsequent stages of the research. Iterative working requires flexibility and, thus, despite the apparently linear structure, there was a degree of overlap between the stages. A brief overview of the stages is described below.



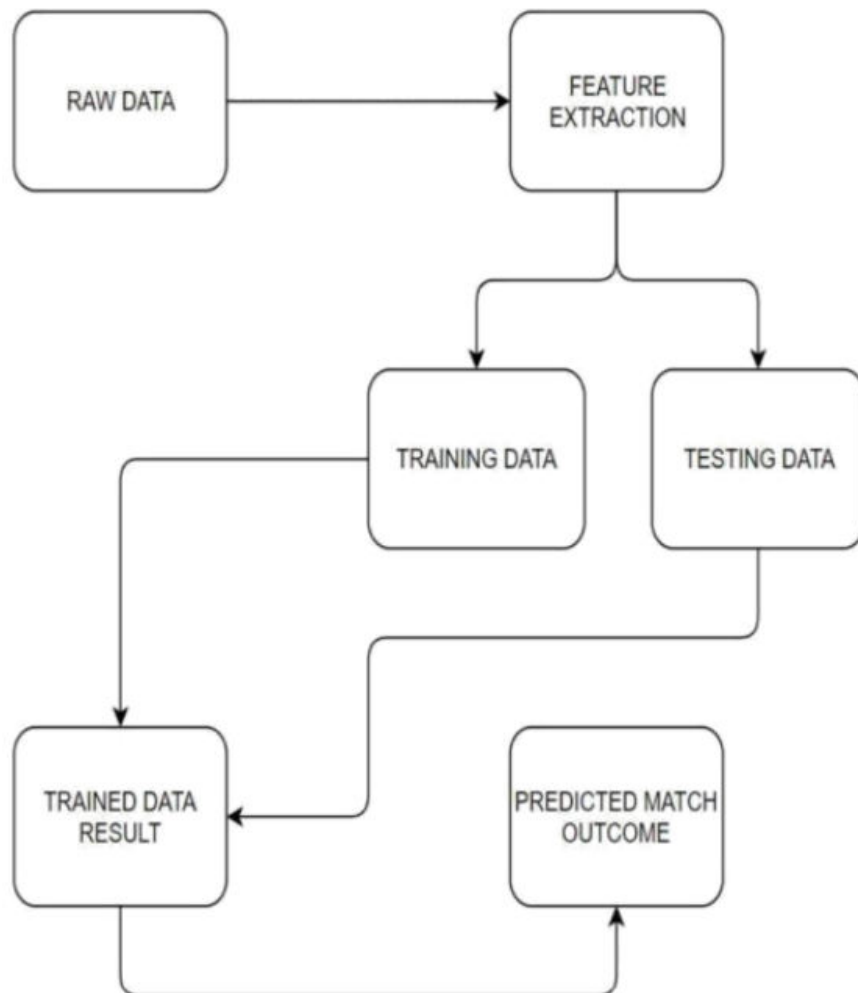


Fig: System design

### V. CONCLUSION

The selection of the right players for any match plays a significant role in the game outcome. A good estimate of the runs a batsman scores, and how many wickets a bowler is likely to take in a match, lets team management pick the best players for each match. Such findings, however, could only be extracted from the data that would come from various sources. We modeled the 11 members team in this project based on the stats and characteristics of the players. Different other factors that influence players' performance could be known as weather, or the design of the ground that could not be included in this study due to data unavailability. Random Forest proved to be the most accurate classifier with optimum precision for the datasets.

### VI. FUTURE SCOPE

Though it is a fun machine learning project, it can be extended to corporate level also, where Sporting channels would like to show the metrics which can be used to increase the audience pulse, which shows the factors which can change



even a losing game inclined to winning. It is used for creating online polls, Sports Business. 34 IPL teams do their own analysis on each and every player as they spend millions on each of them; they use various analyses like Player Vs Ground Battle, before bidding the player. This model can be used for predicting the outcome of other sports also like test cricket, hockey, tennis, baseball, rugby etc.

#### REFERENCES

1. S. Muthuswamy and S. S. Lam, "Bowler Performance Prediction for One-day International Cricket Using Neural Networks," in Industrial Engineering Research Conference, 2008.
2. P. Wickramasinghe, "Predicting the performance of batsmen in test cricket," Journal of Human Sport & Exercise, vol. 9, no. 4, pp. 744-751, May 2014.
3. J G. D. I. Barr and B. S. Kantor, "A Criterion for Comparing and Selecting Batsmen in Limited Overs Cricket," Operational Research Society, vol. 55, no. 12, pp. 1266-1274, December 2004.
4. S. R. Iyer and R. Sharda, "Prediction of athletes performance using neural networks: An application in cricket team selection," Expert Systems with Applications, vol. 36, pp. 5510-5522, April 2009.
5. M. G. Jhanwar 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 (ECML PKDD 2016 2016), 2016.



INNO  SPACE  
SJIF Scientific Journal Impact Factor

Impact Factor: 8.165

 **doi**<sup>®</sup>  
**CROSS** **ref**

**ISSN** INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

 9940 572 462  6381 907 438  [ijircce@gmail.com](mailto:ijircce@gmail.com)



[www.ijircce.com](http://www.ijircce.com)

Scan to save the contact details