



IPL WINNER PREDICTION USING ML ALGORITHMS

¹Ruchitha M, ²Dr. Preeti Savant,

¹Student, ²Assistant Professor,

¹MCA Department,

¹Jain (Deemed-to-be-University), Bangalore, India

Abstract: Cricket is a widely-followed sport in India and around the world. In latest years, the T-20 version of this sport has grown in reputation. The Indian Premier League (IPL), a match primarily based totally on this system, has grown in reputation in the latest years. Cricket, on the alternative hand, is visible as a sport of chance. Predicting the winner of a match or healthy is likewise a fear amongst fanatics and followers. Meanwhile, technology is advancing at an astonishing pace. Following the schooling of a model, devices getting to know algorithms are used to expect something. As a result, in this study, we appoint a lot of supervised mastering strategies to are expecting the winners of Indian Premier League matches. Team names, match site, toss winner, toss decision, match winner, gained through what number of runs, and umpires' gift for the match are the various attributes of this system. Logistic regression, selection tree, random forest, SVM, naive Bayes, gradient boost, and KNN are one of the tasks intently monitored approaches.

I. INTRODUCTION

Machine Learning is becoming increasingly popular in sports analytics. Because of the availability of both live and historical data. Sports analytics is the process of gathering and analyzing data from previous matches in order to extract the most important information, with the goal of assisting in effective decision-making. Decision-making can involve anything from deciding which player to buy during an auction to deciding who to put on the field for tomorrow's match, or it can involve more strategic work such as developing tactics for upcoming matches based on recent performances of players.

Machine learning techniques have become an important part of the technology industry, helping to solve a wide range of complex problems in computer science, operations research, and software engineering. Because of increases in computer power, massive volumes of data, and theoretical knowledge, machine learning techniques have seen a comeback in the twenty-first century.

The fundamental intention is to pick out the vital traits that impact the in-shape end result and to choose the suitable machine learning version for the statistics and the best results. In the field of anticipating the outcome of a cricket match, some research has already been published. Some articles use very few basic criteria, which makes predictions less accurate. On the alternative hand, a few research recommend that the machine learning version is wrong. As a result, it is important to remember all the important elements that would impact the match's outcome, in addition to the right statistics schooling and evaluation version. This will boost prediction accuracy greatly.

Many studies papers were posted and loads of works have been achieved withinside the previous few years to forecast the final results of a cricket fit the use of supervised gadget gaining knowledge of algorithms which includes linear regression, help vector machines, logistic regression, choice tree, Bayes network, and random forest, in addition to factors that have an effect on the fit final results. Cricket is a widely finished pastime everywhere in the world. The Twenty20 format is extremely well-known since it is a fast-paced variety of the game that appeals to both on-subject and at-domestic viewers. Preferable intelligence is one of the sphere's long-term desires. Statistical approaches, computing-based intelligence, and symbolic device Machine Learning are only a few of the methodologies. A variety of tools are used in Machine Learning, including search and mathematical optimization, synthetic neural networks, statistics, probability, and economics tactics, to name a few.

II. OBJECTIVE

The objectives of this project are as follows: -

This project is to create a machine learning model. We are applying various techniques of machine learning to predict the match-winner.

For those who want to check the results, the project includes a single user interface.

The advantage of our system, as we use most of the parameters to predict the result will be quite accurate.

People having craze about IPL matches can use this system to get an overview of the team they support.

III. LITERATURE SURVEY

This article was put together by Rabindra Lamsal and Ayesha Choudhary [1] using data from the Indian Premier League's official website. Because the records protected loads of capabilities, they tested it and picked out many of the maximum vital ones. They pre-processed the records and implemented some fashions the use of the Scikit-research system studying package. Low variance, univariate, and recursive traits have been deleted. They determined five essential capabilities out of 15 through using those characteristic choice fashions. Home team, away team, venue, toss winner, toss decision, and winner are the capabilities. The Random Forests and Multiple Linear Regression fashions have been educated the use of records from the primary ten seasons of the IPL, observed through records from the 11th season. Their version became capable of appropriately categorize forty one out of 60 matches. As a result, their accuracy is 68.33 percent, which isn't always great. This version has obstacles in that it most effective makes use of five capabilities and system studying fashions.

Sankaranarayanan [2] proposes a method for predicting whether forthcoming matches will be won or lost based on historical data. To demonstrate mathematical results and the algorithm's performance in forecasting model results, they used Linear Regression, Nearest Neighboring, and Clustering techniques.

Elnaz Davoodi and Ali Reza Khanteymoori [3] provided a technical paper titled "Horse Racing Prediction Using Machine Learning Predictions." have given their thesis on ML predictions on horse racing in which they used numerous fashions of ML which has given pictorial representations of the information the use of graphs, bar charts and used integrated characteristic of information pre-processing which made the paintings simpler assessment of class fashions.

Technical paper that's given through K. Hiba Sadia, Aditya Sharma, Adarsh Paul, Sarmistha Padhi, Saurav Sanyal [4] at the topic "Stock Market Prediction Using Machine Learning predictions" in which the essential intention of this studies is to set up the nice version for predicting inventory marketplace value. Considering many techniques and variables that ought to be taken into consideration for the duration of the procedure, I've determined that techniques along with random forest, help vector machines have been now no longer exploited fully. They evolved and reviewed a extra achievable method for predicting dataset of inventory marketplace values and it is primary object where they checked out from the preceding 12 months and anticipated that it is more accurate from those models.

Using the ODI dataset, A predictive version is created using ODI dataset for expecting cricket rating and participant performance. SVM and Naive Bayes which are supervised techniques are been used. KNN and MLP which are clustering strategies are used to appropriately categorize. Expecting the winner of IPL matches [1] ML techniques are used. There is a lot of speculation about who may win the Indian Premier League's famed title every year. IPL is a sport wherein the final results can be modified in a depend of seconds the use of Machine Learning strategies inclusive of SVM, Naive Bayes, KNN, Decision Tree, and Logistic Regression to expecting the winner.

Basketball effects are anticipated the use of system gaining knowledge of techniques [5]. Some Machine Learning knowledge of tactics applied on this examine to assemble a version for forecasting NBA sport outcomes consist of Simple Logistics Classifier, Artificial Neural Networks, SVM and Nave Bayes, and Random Forest. Data from five NBA seasons become accrued for version training everyday using the data from five, NBA everyday season become used as a rating dataset, ensuing in a compelling end result using whilst statistics from 1. A statistics mart comprising NBA statistical statistics is generated using whilst automatic statistics and total cloud-primarily based statistics-controlled operations are completed. The above-stated system gaining knowledge of fashions are then educated and examined via way of means of ingesting statistics from the statistics mart. After using the scoring dataset to assess the version accuracy, Simple Logistics Classifier in the end gives you better end result with an accuracy of 69.67%.

IV. DETAILED DESIGN

HIGH LEVEL DESIGN

The architecture that will be utilised to produce software products is described in high level design. The architecture diagram depicts the overall system, defining the main components and their interfaces that will be produced for the product.

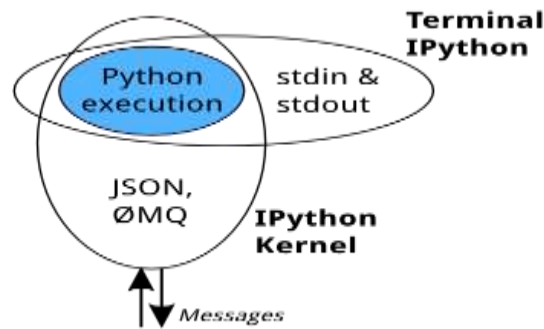
ARCHITECTURE OF IPYTHON NOTEBOOK

The notebook takes interactive computing in a qualitatively new direction by providing a web-based application for capturing the entire computation process, including designing, documenting, and executing code, as well as conveying the results. The Jupyter notebook is formed using two parts:

- A web application is a browser-based tool for interactive document writing that combines explanatory text, mathematics, computations, and the output rich media.
- Notebook documents are a representation of all visible material in the web application, including calculation inputs and outputs, explanatory language, mathematics, graphics, and rich media object representations.

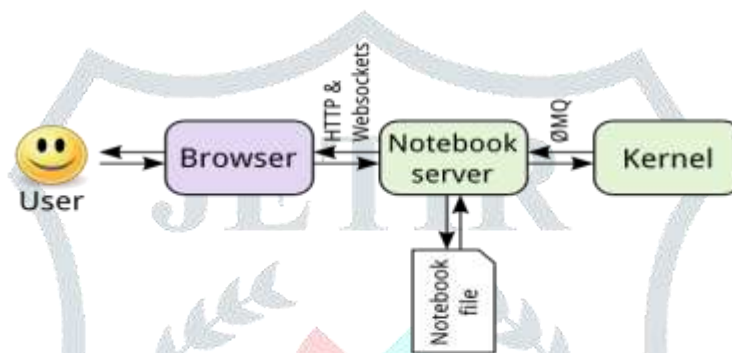
THE PYTHON KERNEL

The Notebook, Qt console, IPython console in the terminal interfaces use the IPython Kernel, and third-party interfaces. The IPython Kernel is a different process that runs user code and performs computing probable completions tasks. JSON messages transmitted across ZeroMQ sockets are used by frontends like the notebook and the Qt console to interface with the IPython Kernel. Figure 5.1 depicts the core execution.



THE NOTEBOOK

The Notebook frontend performs a unique function. It not only runs your code, but it also saves it, together with any markdown notes, in an editable document called a notebook. When it is saved, the browser sends it to the notebook server as a JSON file with the .ipynb extension on disc. The figure depicts its functions briefly.



THE NOTEBOOK DESIGN

For a designer or engineer to keep track of project progress from beginning to end, a design notebook is a good technique. During the design process, it's a location to keep track of research, observations, ideas, drawings, comments, and queries.

SUPERVISED MACHINE LEARNING

Supervised learning is used within the sizable majority of real device gaining knowledge of applications. When you've got enter variables (x) and an output variable (Y), supervised learning takes place whilst you follow a set of rules to examine the mapping feature from the enter to the output ($Y = f(X)$). The purpose is to estimate the mapping feature to the factor that you may forecast the output variables (Y) for brand spanking new enter data (x).

Some of Examples of supervised machine learning techniques are Linear and logistic regression, multi-class classification, Decision Trees, and support vector machines. The data needed to train the algorithm for supervised learning must already be labeled with correct responses. For example, after being trained on a dataset of photos that are appropriately labelled with the animal's species and certain identifying traits, a classification algorithm will learn to identify animals.

Regression and category problems are kind of supervised learning obligations. The motive of each demanding situations is to create an easy version which could expect the fee of the structured characteristic the usage of handiest the characteristic variables. The handiest distinction among the 2 obligations is that the structured function in regression is numerical, while in category it's miles categorical.

The types of classification algorithms used in this project are:

- Logistic regression.
- Naive Bayes Classifier.
- K Nearest Neighbor.
- Support Vector Machines.
- Decision Trees.
- Boosted Trees.
- Random Forest.

V. PROPOSED SYSTEM

a. DATA COLLECTION AND EXPLORATION

```
matches=pd.read_csv("matchesipl.csv")
matches.info()
matches.head()
```

Reading the dataset

```
matches.info()
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 756 entries, 0 to 755
Data columns (total 18 columns):
#   Column                Non-Null Count  Dtype
---  ---
0   id                    756 non-null    int64
1   season               756 non-null    int64
2   city                 749 non-null    object
3   date                 756 non-null    object
4   team1                756 non-null    object
5   team2                756 non-null    object
6   toss_winner          756 non-null    object
7   toss_decision        756 non-null    object
8   result               756 non-null    object
9   dl_applied           756 non-null    int64
10  winner                752 non-null    object
11  win_by_runs          756 non-null    int64
12  win_by_wickets       756 non-null    int64
13  player_of_match      752 non-null    object
14  venue                756 non-null    object
15  umpire1              754 non-null    object
16  umpire2              754 non-null    object
17  umpire3              119 non-null    object
dtypes: int64(5), object(13)
memory usage: 106.4+ KB
```

Summary of Dataset

matches.head()

id	season	city	date	team1	team2	toss_winner	toss_decision	result	dl_applied	winner	win_by_runs	win_by_wickets	player_of_m
0	1	2017	Hyderabad	05-04-17 Sunrisers Hyderabad	Royal Challengers Bangalore	Royal Challengers Bangalore	field	normal	0	Sunrisers Hyderabad	35	0	Yuvraj S
1	2	2017	Pune	06-04-17 Mumbai Indians	Rising Pune Supergiant	Rising Pune Supergiant	field	normal	0	Rising Pune Supergiant	0	7	SPD S
2	3	2017	Rajkot	07-04-17 Gujarat Lions	Kolkata Knight Riders	Kolkata Knight Riders	field	normal	0	Kolkata Knight Riders	0	10	CAI
3	4	2017	Indore	08-04-17 Rising Pune Supergiant	Kings XI Punjab	Kings XI Punjab	field	normal	0	Kings XI Punjab	0	6	GJ Ma
4	5	2017	Bangalore	08-04-17 Royal Challengers Bangalore	Delhi Daredevils	Royal Challengers Bangalore	bat	normal	0	Royal Challengers Bangalore	15	0	KM Jar

First four columns of the Dataset

b. DATA CLEANING

```
print(len(matches[pd.isnull(matches['winner'])]))
matches[pd.isnull(matches['winner'])]
```

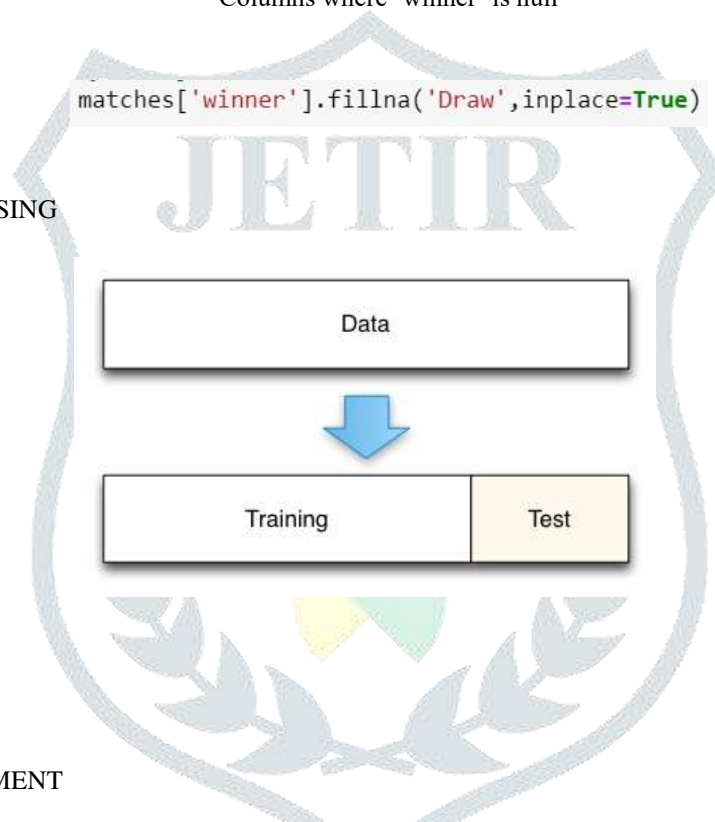
4

	id	season	city	date	team1	team2	toss_winner	toss_decision	result	dl_applied	winner	win_by_runs	win_by_wickets	player_of_m
300	301	2011	Delhi	21-05-11	Delhi Daredevils	Pune Warriors	Delhi Daredevils	bat	no result	0	NaN	0	0	
545	546	2015	Bangalore	29-04-15	Royal Challengers Bangalore	Rajasthan Royals	Rajasthan Royals	field	no result	0	NaN	0	0	
570	571	2015	Bangalore	17-05-15	Delhi Daredevils	Royal Challengers Bangalore	Royal Challengers Bangalore	field	no result	0	NaN	0	0	
744	11340	2019	Bangaluru	30-04-19	Royal Challengers Bangalore	Rajasthan Royals	Rajasthan Royals	field	no result	0	NaN	0	0	

Columns where 'winner' is null

```
matches['winner'].fillna('Draw',inplace=True)
```

c. DATA PREPROCESSING



d. MODEL DEVELOPMENT

```
#generic function for making a classification model and accessing performance
def classification_model(model,data,predictors,outcome):
    model.fit(data[predictors],data[outcome])
    predictions=model.predict(data[predictors])
    #print(predictions)
    print("Confusion Matrix:")
    print(metrics.confusion_matrix(predictions,data[outcome]))
    print()
    accuracy=metrics.accuracy_score(predictions,data[outcome])
    print('Accuracy : %s'%{0:.3%}'.format(accuracy))
```

Generic function for Classification model

```
outcome_var = ['winner']
predictor_var = ['team1', 'team2', 'venue', 'toss_winner','city','toss_decision']
```

The predicting and outcome variables in our project

e. MODEL EVOLUTION

```
#Random Forest Classifier
model=RandomForestClassifier(n_estimators=100)
print("RANDOM FOREST CLASSIFIER:\n")
classification_model(model,df,predictor_var,outcome_var)
import warnings
warnings.simplefilter("ignore")
```

RANDOM FOREST CLASSIFIER:

Confusion Matrix:

```
[[102  4  4  1  4  2  4  0  2  0  0  0  0]
 [  0 79  1  0  1  2  1  0  3  2  0  0  0]
 [  0  4 67  0  1  1  1  0  2  2  1  0  0]
 [  0  0  1 28  0  0  1  0  0  0  0  0  0]
 [  1  1  2  0 90  1  1  0  0  0  0  0  0]
 [  3  1  2  0  1 66  3  0  1  1  0  0  0]
 [  0  2  3  0  2  1 62  1  1  0  0  0  1]
 [  0  0  1  0  0  0  0 12  0  0  0  0  0]
 [  2  1  3  0  0  1  3  0 71  0  0  0  0]
 [  1  0  0  0  0  0  0  0  1 53  0  0  0]
 [  0  0  0  0  0  0  0  0  0  0 14  0  0]
 [  0  0  0  0  0  0  0  0  0  0  0  6  0]
 [  0  0  0  0  1  0  1  0  1  0  0  0 12  0]
 [  0  0  0  0  0  1  0  0  0  0  0  0  0 3]]
```

Accuracy : 87.963%

Predicting the match winner using Random Forest Classifier

```
rf = RandomForestClassifier(n_estimators=100)#no of trees
rf.fit(df[predictor_var],df[outcome_var])
rf_pred = rf.predict(df[predictor_var])

dicVal = encode['winner']
#print(dicVal['MI']) #key value
#print(list(dicVal.keys())[list(dicVal.values()).index(1)]) #find key by value search

team_venue={'MI':'22','KKR':'21','RCB':'2','DC':'14','CSK':'7','RR':'16','DD':'9','GL':'1','KXIP':'6','SRH':'14','RPS':'25','KTK':
#team1, 'team2', 'venue', 'toss_winner', 'city', 'toss_decision'
team1=input("Enter the team1:")
if team1 in dicVal:
    team2=input("Enter the team2:")
    if team2 in dicVal:
        if team1!=team2:
            toss_winner=input("Enter the toss winner:")
            if toss_winner in [team1,team2]:
                input1=[dicVal[team1],dicVal[team2],team_venue[team1],dicVal[toss_winner], '2', '1']
                input1 = np.array(input1).reshape((1, -1))
                output=rf.predict(input1)
                pred_winner=list(dicVal.keys())[list(dicVal.values()).index(output)] #find key by value search output
                if pred_winner in [team1,team2]:
                    print("The predicted winner is:",pred_winner)
                else:
                    print("The predicted winner is:",toss_winner)
            else:
                print("Toss winner must either be Team1 or Team2")
        else:
            print("Team1 and Team2 cannot be same")
    else:
        print("Enter valid team name")
else:
    print("Enter valid team name")

Enter the team1:MI
Enter the team2:RCB
Enter the toss winner:RCB
The predicted winner is: MI
```

It outperforms other algorithms in case of large datasets and multiclass classification tasks. The random forest code snippet is shown below, and it has the highest accuracy of all the algorithms at 89.168 percent.

VI. CONCLUSION

This project has really aided us in grasping the fundamentals of machine learning prediction systems and associated ideas. The IPL prediction mini-project also helped in learning a new Python data science process and its built-in libraries.

By measuring the accuracy of the exclusive algorithms, it's far determined that the maximum appropriate set of rules for predicting the IPL fit winner primarily based totally on numerous facts factors from the historic facts is the random wooded area set of rules. The set of rules may be outstanding for sports activities information evaluation for the reason that it's far skilled on a big series of historic facts and has been selected after being examined on a pattern fact. The mission demonstrates the gadget getting to know version to expect the IPL fit winner with greater accuracy in comparison to formerly carried out gadget getting to know models.

VII. REFERENCE

- [1] Rabindra Lamsal and Ayesha Choudhary, "Predicting Outcome of Indian Premier League (IPL) Matches Using Machine Learning", arXiv:1809.09813 [stat.AP] (September 2018)
- [2] Abhishek Naik, Shivane Pawar, Minakshee Naik, Sahil Mulani, "Winning Prediction Analysis in One-Day-International (ODI) Cricket Using Machine Learning Techniques", International Journal of Emerging Technology and Computer Science, Volume: 3 Issue: 2 (April 2018)
- [3] Arjun Singhvi, Ashish Shenoy, Shruthi Racha, and Srinivas Tunuguntla. "Prediction of the outcome of a Twenty-20 Cricket Match." (2015).
- [4] Swetha, Saravanan.KN, "Analysis on Attributes Deciding Cricket Winning", International Research Journal of Engineering and Technology (IRJET), Volume: 04 Issue: 03 | (March 2017)
- [5] Goddam Jaishankar Harshit, Rajkumar S, "A Review Paper on Cricket Predictions Using Various Machine Learning Algorithms and Comparisons among Them", International Journal for Research in Applied Science & Engineering Technology (IJRASET), IJRASET17099 (April 2018)
- [6] Akhil Nimmagadda, Nidamanuri Venkata Kalyan, Manigandla Venkatesh, Nuthi Naga Sai Teja, Chavali Gopi Raju,
- [7] Research which is given by K. Hiba Sadia, Aditya Sharma, Adarsh Paul, Sarmistha Padhi, Saurav Sanyal on the topic "Stock Market Prediction Using Machine Learning predictions"
- [8] Research paper by Elnaz Davoodi, Ali Reza Khanteymoori on "Horse Racing Prediction Using Machine Learning predictions".
- [9] Vignesh Veppur Sankaranarayanan and Junaid Sattar, "A Data Mining Approach to ODI Cricket Simulation and Prediction", Proceedings of SIAM Conference on Data Mining, pp. 1- 7, 2014.

