e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
Volume:03/Issue:06/June-2021 Impact Factor- 5.354 www.irjmets.com

CRICKET ANALYSIS USING MACHINE LEARNING

B V S Sai Praneeth[*1], V Srighan Reddy[*2], P Jayanth[*3], K Jeevan Reddy[*4]

[*1,2,3]Student, Department of Electronics and Communication Engineering, SreeNidhi Institute of Science and Technology, Hyderabad, Telangana, India.

[*4]Associate Professor, Department of Electronics and Communication Engineering, SreeNidhi Institute of Science and Technology, Hyderabad, Telangana, India.

ABSTRACT

Indian Premier League is a franchise cricket based, annual tournament which is organized by Board of Cricket for Control in India (BCCI) every year. IPL is one of the highly rated cricket tournaments around the world and several international cricket players go through the auction process where they are sold to franchise owners at very high prices. So the amount of money spent by each franchise team builds a huge pressure on the owners to search for victories, which depends on the performance of the team. In general, a franchise signs in players based on their previous year's performance. The result of the match is predicted by analyzing their previous encounters, the venues they played under etc. To predict the outcome of the match, we have used Machine Learning where the current problem statement comes under Classification type of problem. Different Machine Learning algorithms like Random Forest and Support Vector Machine (SVM) algorithms are used in the design of the model. We have used K-Fold cross validation process to use the entire data-set both for training and testing it before the user inputs are given. The classifier which has highest accuracy is then used to predict the end result of the cricket match.

Keywords: Machine Learning, Classification, Random Forest Algorithm, Support Vector Machine (SVM) Algorithm, K-Fold cross validation.

I. INTRODUCTION

CRICKET is one of most sought after games in the world. It is the second most played game in the world after football. The original form of the game was first invented in Saxon or Norman times by the children who might be playing in the dense woodlands of Weald and clearings of South-East England back in 1611. Later in the middle of sixteenth century the game had developed enough to be known as village cricket. The first cricket match including country names was played in 1709. The game has only been played by more and more countries ever since.

The game has implemented several modifications since its inception. Several rules have been added and some rules have also been removed. In the millennial era, it is one of the only games with a large viewership and generates large revenue in India. The first ever first-class match was played in India in 1864 between Madras and Calcutta but not many records exist of that period. It's only after the independence that it had started developing revenue for the economy. The first-class cricket was the longer format of the game with no limit on the number of overs (six consecutive legal deliveries) and was played for a minimum of 3 days to a maximum of 5 days. It included 4 innings'. Later, the games was shortened to a limited over format of 50 overs per side. Later, in 2003 the 20 over format of the game was invented. All the international teams started playing this format in 2005.

After the famous Indian victory in the first 20-20 world cup tournament, this format has only gained more popularity. Hence, in 2008 a state level league including eight to ten teams was founded by Board of Control for Cricket in India. It was called Indian Premier League also known as the IPL. Initially not many people were interested but as years passed it has only gained more and more popularity and generated a large amount of revenue through sponsorship and viewership. BCCI was eyeing at at least Rs.4000 crore as sponsorship revenue in the latest 2021 edition of the tournament before it got postponed indefinitely because of the ongoing pandemic. If the tournament is canceled there is an estimated loss of up to Rs.2500 crore.

In a tournament where every match contributes to large amount of revenue, it is very important to predict the outcome of a match based on the facts from the past. This information can include the history of each individual teams, performance in the past tournaments, performance in the present tournaments, matches won by each teams in head to head competitions of the past, the venue, team captains, highest scorer etc. Now, in the 21st century, with the innumerable advancements in technology, collection of data has become comparatively easier.

Now, with the data being available very easily it has become an important task to develop models that predict the outcome of the match.

Hence, we need to develop a model that predicts the outcome of a match played among the two teams with the help of some standard Machine learning algorithms along with some necessary modifications so that the models can make use of all the factors available to obtain a prediction with a good accuracy. The model should be able to predict the outcome dynamically, and not just for certain types of data. That is a very important factor for any model.

So, now after understanding the importance of the league, some basic understanding of mathematics, statistics, and some standard machine learning tools we will begin developing a program to predict the outcome of the match.

II. METHODOLOGY

We have followed general Machine Learning Cycle to build our project which includes the following steps:

1) **Data Collection:** We had collected the dataset from www.kaggle.com

2) **Hypothesis Generation:** Here entire dataset and its features ae analyzed by studying various publications and came to know about he features that will have a impact in predicting the result of the match.

3) **Data-Preprocessing: T**he dataset we had consists of missing values, outliers and duplicate records. So, in order to build a model , we had to first correct these values in the necessary features to improve accuracy and checked these features using univariate and bivariate analysis.

4) **Data-Encoding:** As our project is about a Classification Problem, to analyse some categorical variables we have done encoding of those feature values for better results.

5) **Model Selection and Evaluation:** We had built our model using Random forest and Support Vector machine (SVM) and by using K-Fold cross validation technique entire data-set is used for both training and testing of the model.

a) Random Forest:

Random Forest is basically a tool that builds as many decision trees as possible to the dataset and makes a more accurate and stable prediction. The number of hyperparameters in a Random Forest is similar to that of simple decision tree hence making it very simple to implement and also the Random Forest contributes to more randomness to the model making it more dynamic.

b) **Support Vector Machine:**

Support Vector Machine (SVM) algorithm, is used for both the Classification and Regression tasks. However, it is primarily used for Classification tasks of Machine Learning. SVM algorithm creates the best decision boundary (also known as Hyperplane) which splits the N-dimensional space into distinct classes such that we can easily enter the new data point into the correct category in future.

6) **Real Time Testing**: Once the model is ready, inputs given by the user are taken and analyzed to predict the winner of the match.

III. MODELING AND ANALYSIS

We have used Jupyter Notebook and Anaconda Navigator as tools for building the project. After importing necessary libraries like Scikit learn, Matplotlib, Numpy, Pandas into the notebook and the required dataset after Data preprocessing and Encoding, we have created a function called classification( ) in which we have described our model. The architecture of the project is shown as below:

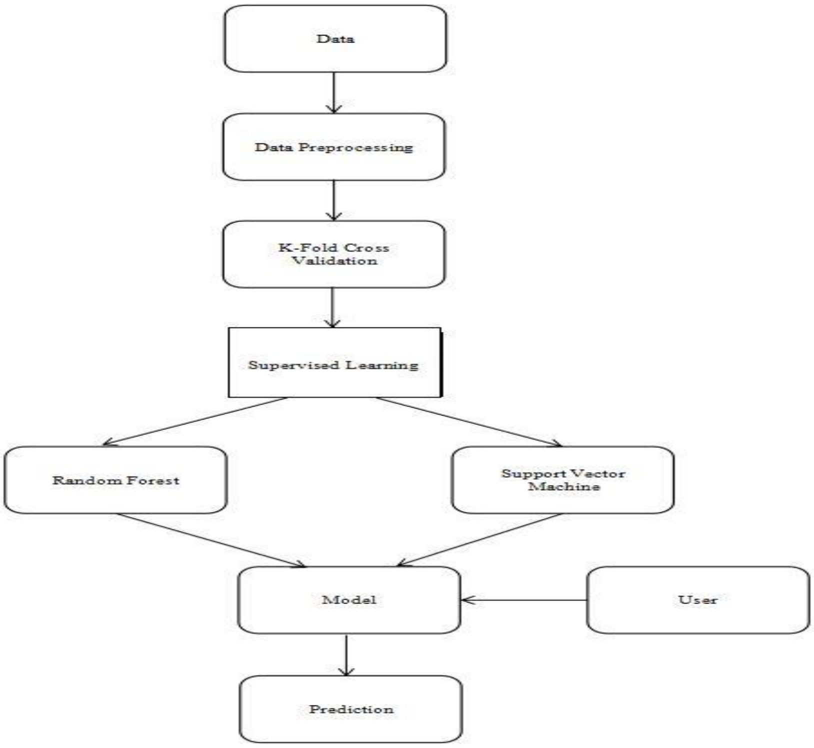
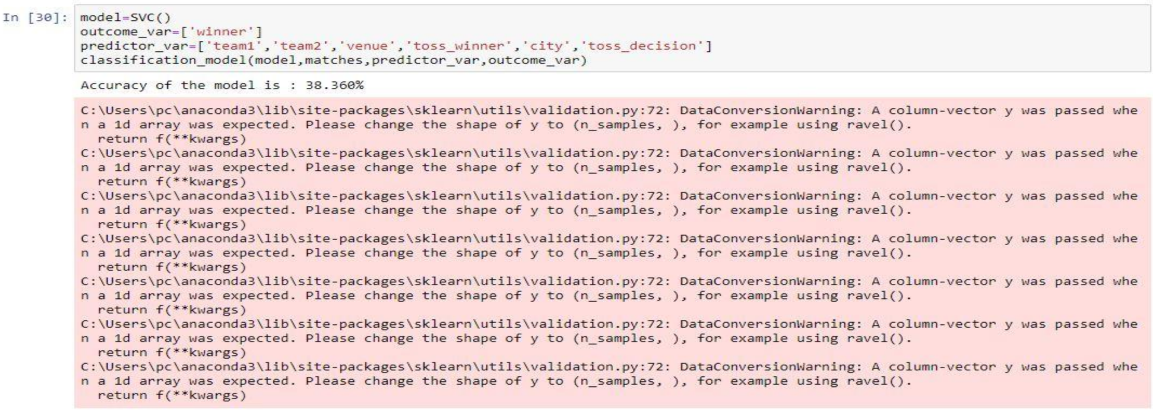


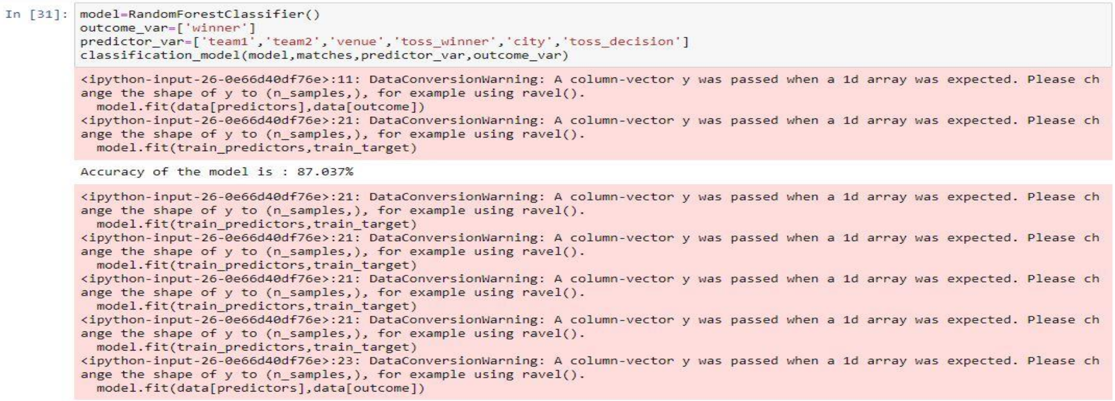
Figure 1: System Architecture

Now that our model is designed, we retrieve the desired results in the necessary form. As we are just predicting the winner of a match played between two teams, we must ascertain the accuracy at which out model predicts the output.



```
In [30]: model=SVC()
         outcome_var=['winner']
         predictor_var=['team1','team2','venue','toss_winner','city','toss_decision']
         classification_model(model,matches,predictor_var,outcome_var)

         Accuracy of the model is : 38.360%

         C:\Users\pc\anaconda3\lib\site-packages\sklearn\utils\validation.py:72: DataConversionWarning: A column-vector y was passed whe
         n a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().
           return f(**kwargs)
         C:\Users\pc\anaconda3\lib\site-packages\sklearn\utils\validation.py:72: DataConversionWarning: A column-vector y was passed whe
         n a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().
           return f(**kwargs)
         C:\Users\pc\anaconda3\lib\site-packages\sklearn\utils\validation.py:72: DataConversionWarning: A column-vector y was passed whe
         n a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().
           return f(**kwargs)
         C:\Users\pc\anaconda3\lib\site-packages\sklearn\utils\validation.py:72: DataConversionWarning: A column-vector y was passed whe
         n a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().
           return f(**kwargs)
         C:\Users\pc\anaconda3\lib\site-packages\sklearn\utils\validation.py:72: DataConversionWarning: A column-vector y was passed whe
         n a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().
           return f(**kwargs)
         C:\Users\pc\anaconda3\lib\site-packages\sklearn\utils\validation.py:72: DataConversionWarning: A column-vector y was passed whe
         n a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().
           return f(**kwargs)
         C:\Users\pc\anaconda3\lib\site-packages\sklearn\utils\validation.py:72: DataConversionWarning: A column-vector y was passed whe
         n a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().
           return f(**kwargs)
```

Figure 2: Testing Using Support Vector Machine (SVM) Algorithm

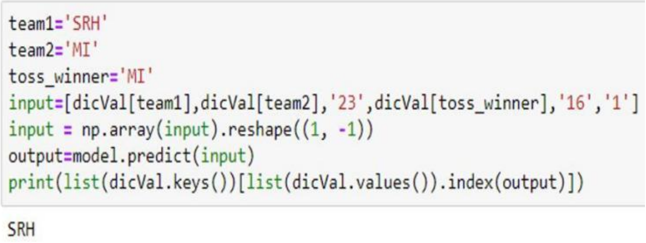


```
In [31]: model=RandomForestClassifier()
         outcome_var=['winner']
         predictor_var=['team1','team2','venue','toss_winner','city','toss_decision']
         classification_model(model,matches,predictor_var,outcome_var)

         <ipython-input-26-0e66d40df76e>:11: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please ch
         ange the shape of y to (n_samples,), for example using ravel().
           model.fit(data[predictors],data[outcome])
         <ipython-input-26-0e66d40df76e>:21: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please ch
         ange the shape of y to (n_samples,), for example using ravel().
           model.fit(train_predictors,train_target)

         Accuracy of the model is : 87.037%

         <ipython-input-26-0e66d40df76e>:21: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please ch
         ange the shape of y to (n_samples,), for example using ravel().
           model.fit(train_predictors,train_target)
         <ipython-input-26-0e66d40df76e>:21: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please ch
         ange the shape of y to (n_samples,), for example using ravel().
           model.fit(train_predictors,train_target)
         <ipython-input-26-0e66d40df76e>:21: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please ch
         ange the shape of y to (n_samples,), for example using ravel().
           model.fit(train_predictors,train_target)
         <ipython-input-26-0e66d40df76e>:21: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please ch
         ange the shape of y to (n_samples,), for example using ravel().
           model.fit(train_predictors,train_target)
         <ipython-input-26-0e66d40df76e>:23: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please ch
         ange the shape of y to (n_samples,), for example using ravel().
           model.fit(data[predictors],data[outcome])
```

Figure 3: Testing Using Random Forest Classifier

IV. RESULTS AND DISCUSSION

As Random Forest Algorithm has an accuracy which is far greater than that achieved by Support Vector Machine (SVM) algorithm, we have used Random Forest in the final model for predicting the winner of the match. Then the inputs from the user are taken after the toss in predicting the outcome of the match.



```
team1='SRH'
team2='MI'
toss_winner='MI'
input=[dicVal[team1],dicVal[team2],'23',dicVal[toss_winner],'16','1']
input = np.array(input).reshape((1, -1))
output=model.predict(input)
print(list(dicVal.keys())[list(dicVal.values()).index(output)])

SRH
```

Figure 4: Prediction of Winner of the match by the user inputs

So, we can simply say that most of these machine learning problems are optimization problems in which we minimize a target label into different numerical constraints.

V. CONCLUSION

The tools used for Machine Learning are very powerful and some strong predictions can be made through efficient use of these tools. Understanding Machine Learning concepts will be very important as the amounts of data generated by the modern advancements in technology will hugely help in the making decisions that might effect the future in a positive way. These tools will also play a very important role in making important decisions in business organizations in which the impact will directly effect the revenue generated by the firm. Random Forest is one of the accurate machine learning classifier since it takes multiple decision trees in account and builds an ensemble model of them. As we got an accuracy of 87.037% with Random Forest which is way more than Support Vector Machine Algorithm, we have used it to build the model using Random Forest in order to predict the result.

ACKNOWLEDGEMENTS

This project is accomplished based on the knowledge gained during pandemic infused lockdown by the authors under the guidance of a Mr. K. Jeevan Reddy who is an Associate Professor at Department of Electronics and Communication Engineering in SreeNidhi Institute of Science and Technology, Hyderabad, Telangana, India.

VI. REFERENCES

[1] Abhishek Naiket. Al, “Winning Prediction Analysis in One-Day-International (ODI) Cricket Using Machine Learning Techniques”, IJETCS, vol. 3, issue 2, ISSN:2455-9954, April 2018.

[2] Akhil Nimmagadda et. Al, “Cricket score and winning prediction using data mining”, IJARnD Vol.3, Issue3.

[3] Bunker, Rory & Thabtah, Fadi. (2017) “A Machine Learning Framework for Sport Result Prediction. Applied Computing and Informatics”, 15. 10.1016/j.aci.2017.09.005.

[4] Esha Goel and Er. Abhilasha, “Random Forest: A Review”, IJARCSSE, Volume 7, Issue 1, DOI: 10.23956/ijarcsse/V7I1/01113, 2017.

[5] I. P. Wickramasinghe et. al, "Predicting the performance of batsmen in test cricket," Journal of Human Spo

[6] Munir, F., Hasan, M.K., Ahmed, S., Md Quraish, S., 2015. Predicting a T20 cricket match result while the match is in progress (Doctoral dissertation, BRAC University).rt & Exercise”, vol. 9, no. 4, pp.744-751, May 2014.

[7] N. Pathak, H. Wadhwa, Applications of modern classification techniques to predict the outcome of ODI cricket, Procedia Comput. Sci. 87 (2016) 55–60.

[8] Passi, Kalpdrum & Pandey, Niravkumar. (2018) “Predicting Players' Performance in One Day International Cricket Matches Using Machine Learning” 111-126. 10.5121/csit.2018.80310.

[9] R. P. Schumaker, O. K. Solieman and H. Chen, "Predictive Modeling for Sports and Gaming” in Sports Data Mining, vol. 26, Boston, Massachusetts: Springer, 2010.

[10] Rabindra Lamsal and AyeshaChoudhary, “Predicting Outcome of Indian Premier League (IPL) Matches Using Machine Learning”

[11] Rameshwari Lokhande and P.M.Chawan, “Live Cricket Score and Winning Prediction”, International Journal of Trend in Research and Development, Volume 5(1), ISSN: 2394-9333

[12] Ujwal U J et. At, “Predictive Analysis of Sports Data using Google Prediction API” International Journal of Applied Engineering Research”, ISSN 0973-4562 Volume 13, Number 5 (2018) pp. 2814-2816.