

# Comparative Analysis of Machine Learning Models for Cricket Score and Win Prediction: A Case Study of Linear Regression, Random Forest, Neural Network, and Elastic Net

Omar Faruq<sup>1</sup>, Arifa Sultana Mily<sup>2</sup>, Ashraful Islam<sup>3</sup>, Shahriar Sadman Dihan<sup>4</sup>, Mohammad Asif Khan<sup>5</sup>

<sup>12345</sup>UG-Students, Department of Computer Science Engineering, East West University, A, 2 Jahurul Islam Ave, Dhaka

\*\*\*

**Abstract** - Cricket is the 2<sup>nd</sup> most popular sport in the world behind football. Especially, in Bangladesh the popularity of cricket and the passion that people show for it, can't be described in words. Cricket is played all around the year. So, we have built a system that predicts a team's final score at any point of the match. The things that are taken into consideration are the teams that are playing, where the match is being played, what's the current runs and wickets, how many runs have been scored, and how many wickets have fallen in the last 5 overs. We are using 3 different machine algorithms; Linear Regression, Random Forest, Elastic Net, and Neural Network to predict the score. There is also a live win prediction system that will predict which team is going to win at any moment of the match. The Random Forest algorithm is used for the win prediction system. Random Forest has the highest R<sup>2</sup> value. It also has the lowest mean squared error and mean absolute error.

**Key Words:** Machine Learning, Linear Regression, Random Forest, Elastic Net, Neural Network, ODI Cricket Winning Prediction, ODI Cricket Score Prediction, Cricket Score Prediction, Cricket Win Prediction

## 1. INTRODUCTION

Cricket is a global phenomenon, loved by millions around the world. In the Sixteenth Century in England, cricket originated. International cricket matches are almost always being played. Fans eagerly wait for big events like the World Cup, and Champions Trophy, all of which are being played by a lot of teams representing their nations.

The main formats in international cricket are Test, ODI, and T20. Among these ODI can easily be considered the most popular. There is massive media coverage, and financial incentives surrounding cricket. Before and during the match fans and cricket experts make their predictions.

So, there is a big demand for score prediction systems. Modern-day broadcasters want to show the viewers as many detailed statistics as possible. So, score Prediction and win prediction is a big deal for them. Our prediction system will do exactly that. It will do score prediction and

win prediction for any international ODI match. The task will be done by multiple efficient machine-learning algorithms.

### 1.1 Objectives

1. To Predict the cricket score of international ODI matches.
2. To Predict the winner of the match.
3. Being more efficient than existing systems.
4. A more complex system that will take a lot of things into consideration when doing prediction.
5. Using different models so that the best-performing model can be found and used for prediction.

### 1.2 Scope

Our project aims to change the perception of cricket fans predicting cricket scores by relying on statistical data rather than some fan-made theories. This scope makes predicting cricket outcomes, such as live score or win predictor easier with more reliability and accuracy.

## 2. PROPOSED SYSTEM

Cricket score and winning prediction It's a system that can predict the live cricket scores and win predictions. In this system, we use 4 types of models. There are two segments in this model. In the initial portion we will predict the live score of a live cricket match considering various parameters and, in the second portion, the model will predict the winner of the match by considering different parameters and efficient algorithms.

## 3. METHODOLOGY

### 3.1 Algorithms

#### 1. Linear Regression:

Linear regression is a type of supervised machine learning algorithm that can calculate a linear relationship between a dependent variable and one or more independent variables. There are two types of linear regression; univariate linear regression and multivariate linear regression. The

algorithm tries to find the best possible value of the dependent variable by finding the equation of the best-fit line for a set of paired data. The equation provides a straight line that represents the relationship between the dependent and independent variables. Linear regression is used for both classification and regression problems. We used multivariate linear regression.

### 2. Random Forest:

Random forest is a classifier which is a collection of multiple decision trees. Instead of hoping for a single decision tree, Random Forest takes the prediction of each tree and using the bagging method predicts the final result. The more decision trees there are the greater the accuracy.

### 3. Neural network:

A neural network is a multi-layered machine learning process. There are three types of layers in neural networks. The first layer is the input layer. Then there are hidden layers. The accuracy will increase if the number of hidden layers increases. All the calculation is done in the hidden layer. The final layer is the output layer that shows the output. Our neural network only has one hidden layer.

### 4. Elastic Net:

Elastic Net linear regression with regularization concept that uses both Lasso and Ridge Regulation techniques. Regulation is used so that the model doesn't overfit. Lasso is used to identify the most influential features it can discard some features when training the model. On the other hand, Ridge doesn't discard any features rather it only improves the overall fitness of the model. The elastic net that we used is a 50-50 mix of the Lasso and Ridge regularization techniques.

### 3.2 System Description

To fit the model, we collected the dataset from GitHub and the dataset provided us ball by ball-by-ball information about the match. We divided the dataset into two segments, training and testing. This model predicts both the winner and the live score. For live score prediction or prediction of winner, we considered parameters are overs bowled, runs scored, number of wickets fallen, runs scored in the previous 5 overs, and wickets fell in the previous 5 overs. We applied more than one classifier algorithm to these parameters and to build up the model we used Google Colaboratory execution.

### 4. DATASET

Many websites have lots of datasets but every dataset is useful for cricket score prediction we collected datasets for cricket score prediction from GitHub. the prediction has 350899 rows and 15 columns. which had 9 types of numerical datasets and 5 types of categorical datasets over

which different techniques are applied. In this paper, we use three models for introducing a score that does not entirely support the current run rate, although this paper introduces a model that includes three methods and takes into account the number of wickets, the match's location, and the batting team. The second approach focuses on the result of the game between the second innings, with the batting team being given a target while taking into account the same factors as the first inning. These techniques were used to classify the worst-case linear regression for the first shift and second inning.

mid	date	venue	bat_team	bowler	runs	wickets	overs	runs_last_5_overs	wickets_last_5_overs	total	
1	#####	Civil Serv	England	Ireland	ME Tresco DT Johnstc	0	0	0.1	0	0	301
1	#####	Civil Serv	England	Ireland	ME Tresco DT Johnstc	0	0	0.2	0	0	301
1	#####	Civil Serv	England	Ireland	ME Tresco DT Johnstc	4	0	0.3	4	0	301
1	#####	Civil Serv	England	Ireland	ME Tresco DT Johnstc	6	0	0.4	6	0	301
1	#####	Civil Serv	England	Ireland	ME Tresco DT Johnstc	6	0	0.5	6	0	301
1	#####	Civil Serv	England	Ireland	ME Tresco DT Johnstc	6	0	0.6	6	0	301
1	#####	Civil Serv	England	Ireland	EC Joyce D Langfor	6	0	1.1	6	0	301
1	#####	Civil Serv	England	Ireland	EC Joyce D Langfor	6	0	1.2	6	0	301
1	#####	Civil Serv	England	Ireland	EC Joyce D Langfor	6	0	1.3	6	0	301
1	#####	Civil Serv	England	Ireland	EC Joyce D Langfor	7	0	1.3	7	0	301
1	#####	Civil Serv	England	Ireland	EC Joyce D Langfor	8	0	1.4	8	0	301
1	#####	Civil Serv	England	Ireland	ME Tresco D Langfor	8	0	1.5	8	0	301
1	#####	Civil Serv	England	Ireland	ME Tresco D Langfor	9	0	1.6	9	0	301
1	#####	Civil Serv	England	Ireland	ME Tresco DT Johnstc	10	0	2	10	0	301
1	#####	Civil Serv	England	Ireland	ME Tresco DT Johnstc	10	0	2.1	10	0	301
1	#####	Civil Serv	England	Ireland	ME Tresco DT Johnstc	11	0	2.2	11	0	301
1	#####	Civil Serv	England	Ireland	EC Joyce DT Johnstc	11	0	2.3	11	0	301
1	#####	Civil Serv	England	Ireland	EC Joyce DT Johnstc	12	0	2.4	12	0	301
1	#####	Civil Serv	England	Ireland	ME Tresco DT Johnstc	12	0	2.5	12	0	301
1	#####	Civil Serv	England	Ireland	ME Tresco DT Johnstc	16	0	2.6	16	0	301
1	#####	Civil Serv	England	Ireland	EC Joyce D Langfor	17	0	3	17	0	301
1	#####	Civil Serv	England	Ireland	EC Joyce D Langfor	17	0	3.1	17	0	301
1	#####	Civil Serv	England	Ireland	EC Joyce D Langfor	21	0	3.2	21	0	301
1	#####	Civil Serv	England	Ireland	EC Joyce D Langfor	21	0	3.3	21	0	301
1	#####	Civil Serv	England	Ireland	EC Joyce D Langfor	21	0	3.4	21	0	301
1	#####	Civil Serv	England	Ireland	EC Joyce D Langfor	21	0	3.5	21	0	301
1	#####	Civil Serv	England	Ireland	EC Joyce D Langfor	21	0	3.6	21	0	301
1	#####	Civil Serv	England	Ireland	ME Tresco DT Johnstc	25	0	4.1	25	0	301
1	#####	Civil Serv	England	Ireland	ME Tresco DT Johnstc	29	0	4.2	29	0	301
1	#####	Civil Serv	England	Ireland	ME Tresco DT Johnstc	32	0	4.3	32	0	301
1	#####	Civil Serv	England	Ireland	EC Joyce DT Johnstc	32	0	4.4	32	0	301
1	#####	Civil Serv	England	Ireland	EC Joyce DT Johnstc	32	0	4.5	32	0	301

Figure 1. Dataset

### 5. RESULT

Here are some of our project's (Live Cricket Score and Winning Prediction) outcomes.

```

2194/2194 [=====] - 3s 2ms/step
Linear Regression:
MSE: 1482.097478318081
MAE: 29.13189526550931
R2: 0.6184790222410205

Random Forest:
MSE: 74.73676518869883
MAE: 3.467539279175704
R2: 0.9807612898972788

Neural Network:
MSE: 499.0276517264322
MAE: 16.09303880731208
R2: 0.8715404888000386

Elastic Net:
MSE: 1836.821680792653
MAE: 31.78005775067694
R2: 0.5271660508995848

Random Forest has the lowest MSE.
Random Forest has the highest R2.
    
```

Figure 2. Comparison of models Using MSE, MAE & R<sup>2</sup>

Here we can see four types of models among them Random Forest is the best model in all aspects. It has the lowest MSE value, MAE value, and best R<sup>2</sup> value. Mean Squared Error is the average of the squared difference between the data set's original and forecasted values. It computes the residual variance. The mean absolute error is the average of the

absolute difference between the actual and expected values in the dataset. It computes the mean of the dataset's residuals. R-squared is a statistical measure that reflects how much of a dependent variable's fluctuation can be explained by an independent variable in a regression model.

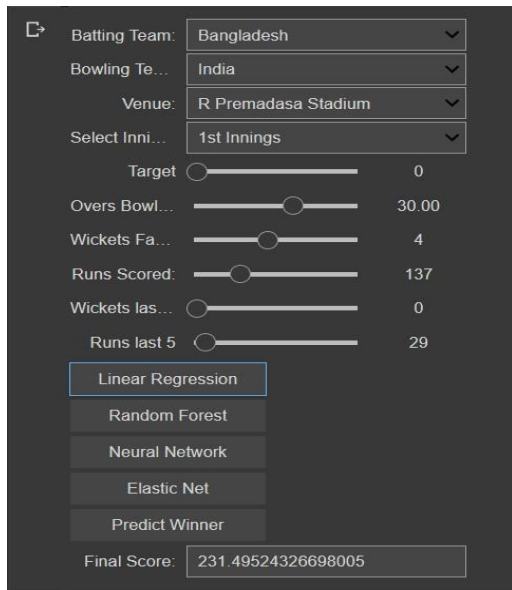


Figure 3. Score prediction using Linear Regression

For Linear Regression, after selecting all the required info if we press the linear regression button the model will show us the predicted final score of the batting team.

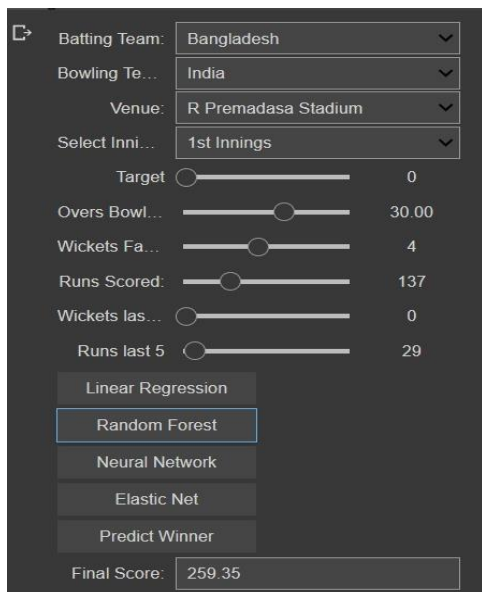


Figure 4. Score prediction using Random Forest

For Random Forest, after selecting all the required info if we press the Random Forest button the model will show us the predicted final score of the batting team.

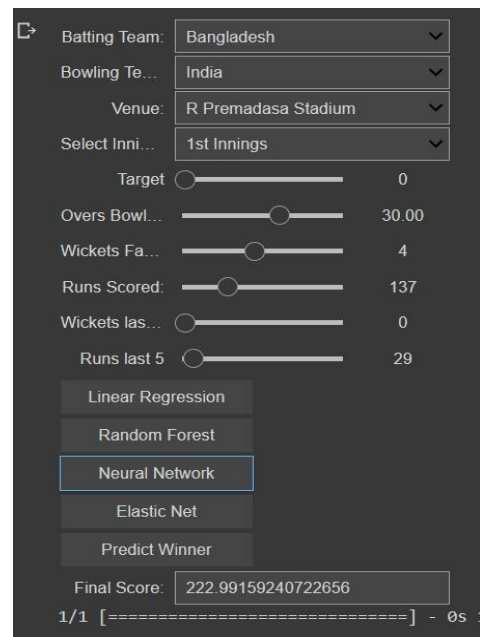


Figure 5. Score prediction using Neural Network

For Neural Network, after selecting all the required info if we press the Neural Network button the model will show us the predicted final score of the batting team.

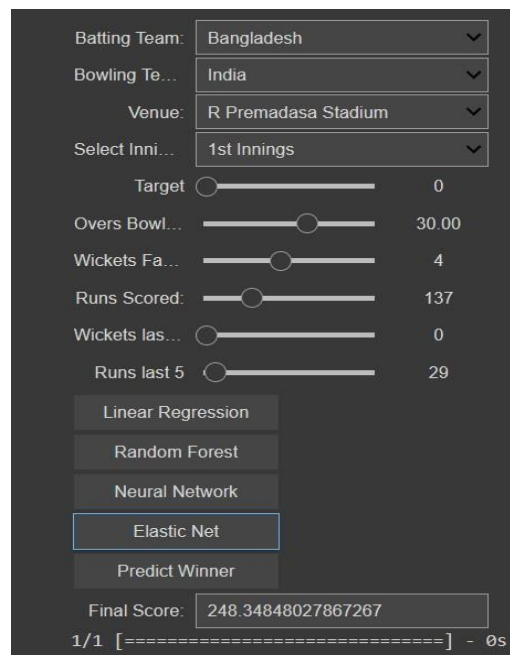


Figure 6. Score prediction using Elastic Net

For Elastic Net, after selecting all the required info if we press the Elastic Net button the model will show us the predicted final score of the batting team.

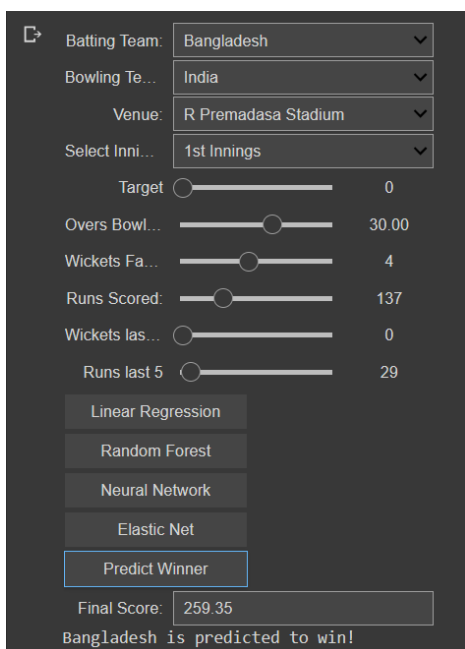


Figure 7. Winner prediction using Random Forest

We can predict the winner for all four phases by selecting all of the required information also we can select if it's 1<sup>st</sup> inning or 2<sup>nd</sup> inning. If we press the Predict Winner button the model will display the anticipated winner team name below the final score.

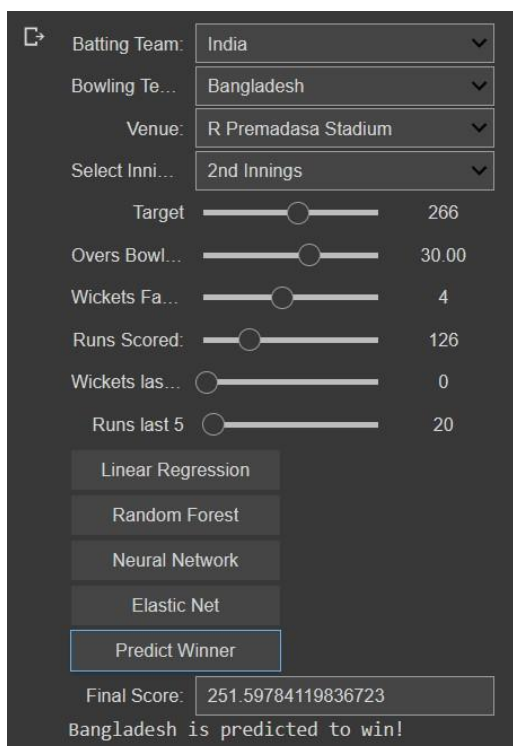
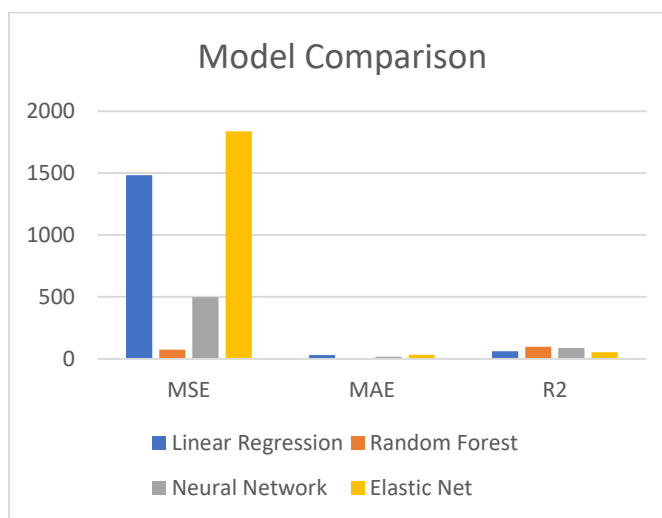


Figure 8. Winner prediction using Random Forest

### 6. Model Analysis

Performance Measure	Linear Regression	Random Forest	Neural Network	Elastic Net
MSE	1482.09	74.73	499.03	1836.82
MAE	29.13	3.46	16.09	31.78
R <sup>2</sup>	0.61	0.98	0.87	0.52

Table 1. Model Comparison



Graph 1. Model Comparison

### 7. CONCLUSION

This paper shows the statistical models for predicting the participation of every team member in every match. Different types of datasets can conduct the analysis and predict the score of the match. Also, Data Science will be converged including pre-processing of data, Visualization of data, preparation of data, feature selection, and implementation of different machine learning models for the predictions. The main goal of this paper is to use past data to forecast the final score and match winner. To accurately forecast the score of innings and obtain the desired outcome, several machine learning models will be applied to specified data for predicting the winning team.

### 8. REFERENCES

[1] An Ultimate Tutorial to Neural Networks (2023), AI&MachineLearning.[https://www.simplilearn.com/tutorials/deep-learning-tutorial/neural-network?fbclid=IwAR3uHPZ-ju0e0gXagNBSDkUGUg3DVG66w1s9Za76UB05xYQGIRX1vg\\_pKGg#:~:text=A%20neural%20network%20is%20usually,more%20than%20one%20hidden%20layer](https://www.simplilearn.com/tutorials/deep-learning-tutorial/neural-network?fbclid=IwAR3uHPZ-ju0e0gXagNBSDkUGUg3DVG66w1s9Za76UB05xYQGIRX1vg_pKGg#:~:text=A%20neural%20network%20is%20usually,more%20than%20one%20hidden%20layer)

**[2]** Cricket Score Prediction Using Machine Learning (2023),  
[https://ijirt.org/master/publishedpaper/IJIRT157821\\_PAPER.pdf](https://ijirt.org/master/publishedpaper/IJIRT157821_PAPER.pdf)

**[3]** Cricket Score Prediction Using Machine Learning (2021), RSCOE, Assistant Professor, Pune, Maharashtra, India. 2,3,4,5RSCOE, Pune, Maharashtra,India,<https://turcomat.org/index.php/turkbilmat/article/view/1546/1299>

**[4]** Cricket Score and Winning Prediction (2022), Omkar Mozar, Soham More, Shubham Nagare, Prof. Nileema Pathak, <https://www.irjet.net/archives/V9/i4/IRJET-V9I4181.pdf>

**[5]** Cricket Score Predictor, Shivam Mitra (2019).<https://github.com/codophobia/CricketScorePredictor/blob/master/data/odi.csv>

**[6]** CRICKET TEAM PREDICTION USING MACHINE LEARNING TECHNIQUES (2021), Nilesh M. Patil, Bevan H. Sequeira, Neil N. Gonsalves, Abhishek A. Singh,[https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=3572740](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3572740)

**[7]** Cricket Team Prediction with Hadoop(2017), Shubham Agarwal, lavish yadav, shikha mehta,India,[https://www.researchgate.net/publication/321749942\\_Cricket\\_Team\\_Prediction\\_with\\_Hadoop\\_Statistical\\_Modeling\\_Approach](https://www.researchgate.net/publication/321749942_Cricket_Team_Prediction_with_Hadoop_Statistical_Modeling_Approach)