e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:05/Issue:10/October-2023 Impact Factor- 7.868 www.irjmets.com

CRICKET TOURNAMENT WINNER PREDICTION USING REGRESSION-BASED ML ALGORITHMS

Omkar Gurav[*1]

[*1]UG Student, B. K Birla College (Empowered Autonomous Status) Kalyan, India.

ABSTRACT

Cricket, a sport cherished by millions worldwide, captivates enthusiasts with its inherent unpredictability. The allure of predicting match outcomes adds an intriguing dimension to the game. In recent years, the integration of data analytics and machine learning in sports has revolutionized the understanding of cricket, transcending traditional perceptions and fostering innovative strategies. This research delves into the realm of predictive modeling using regression techniques, specifically linear regression, to Cat Boost Regressor tournament winners.

Keywords: Regression-Models, Cricket, Cat boost regressor, Linear Regression, Prediction.

I. INTRODUCTION

Cricket, a sport adored by millions, is a game filled with excitement and surprises. Predicting the results of an interesting task. The unpredictable nature is one of the main reasons that people enjoy sport. The rise of data analytics and machine learning in matches in sports is challenging and sports has paved the way for a deeper understanding of the game, transcending traditional perceptions and paving the way for innovation strategies.

The significance of this research extends beyond the boundaries of the cricket field. Accurate predictions not only enhance the spectator experience, but also provide valuable insights to team management, bookmarkers, and sports analysts. This project paper showcases the performance of various regression models like linear regression The study is more focused on the less popular regression models than the most popular ones like SVM and Random Forest

II. RELATED WORK

In [1] paper the authors proposed a study of predicting the match results and player performance. In the [2] paper author proposed a study that predicted the winner of the ODI matches. In [3] paper author showcases the cricket score forecasting using neural networks.

III. METHODOLOGY

For formulating the research, we used the "IPL Test Dataset" and performed the data preprocessing on the dataset, and then we performed the data visualization on the dataset in different ways, like Season Feature, Toss Feature, Venue Feature, Runs Across Seasons, etc.We split the 20 % data for model testing and 80% data for model training purposes. After data splitting, we train the different models for predicting the tournament winner and find out the model with the best accuracy.

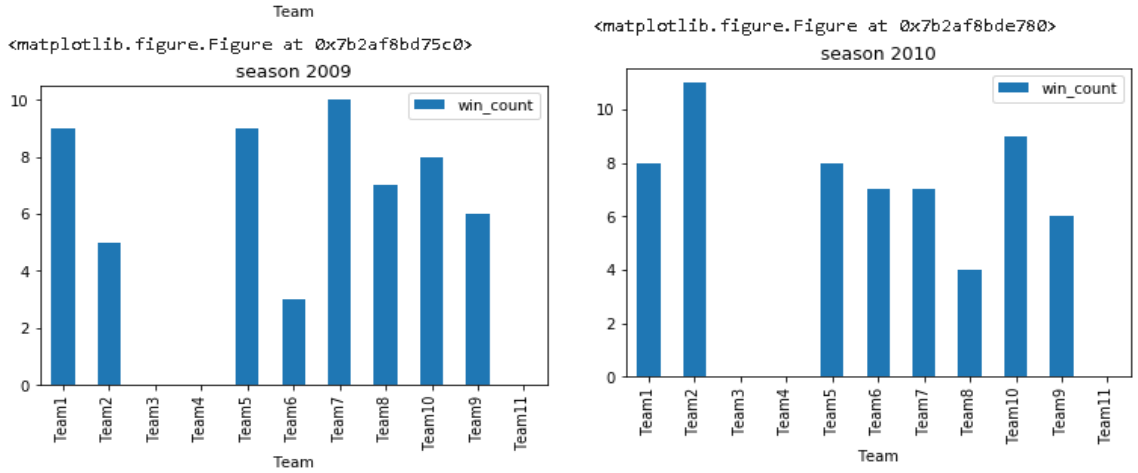

Figure 1: Shows the number of matches won by teams (Season Feature).

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:05/Issue:10/October-2023 Impact Factor- 7.868 www.irjmets.com

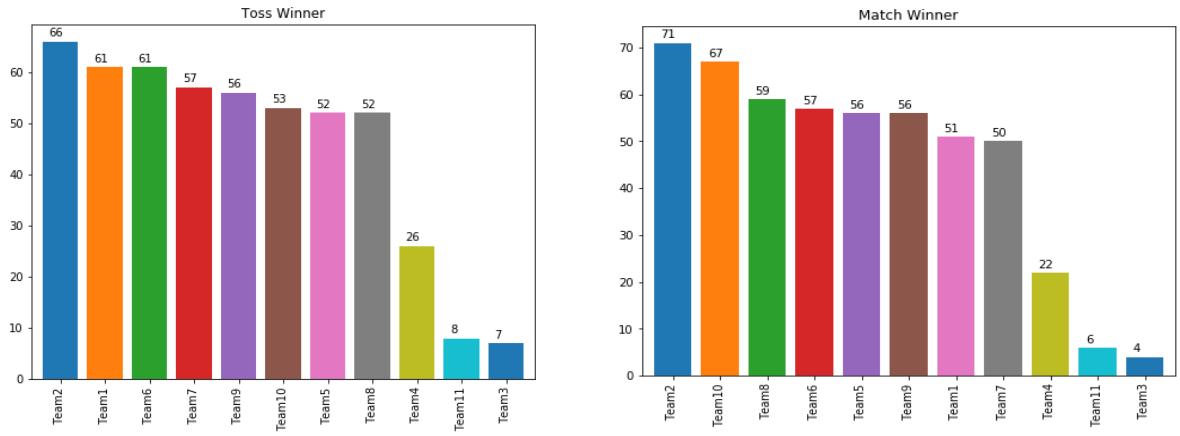

Figure 2: Visualizes the relation between the Toss Winner and Match Winner (Toss Feature).

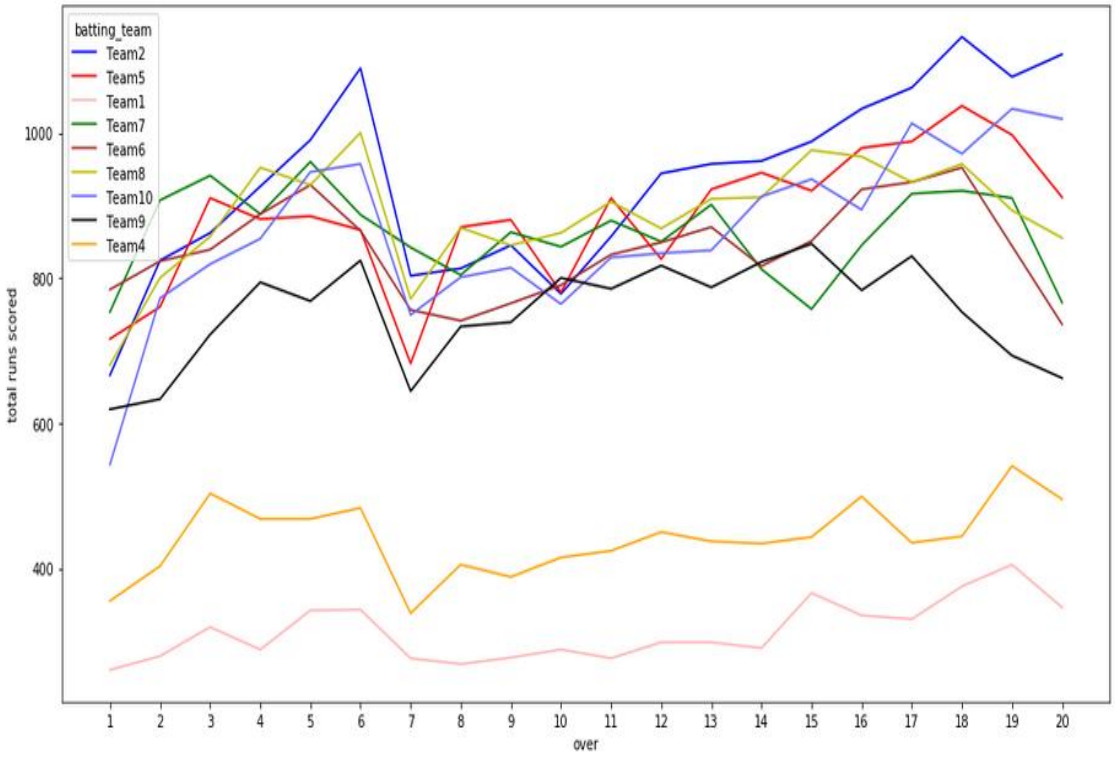

Figure 3: Shows the Runs Scored Across Season by each team (Runs Across Season)

IV. RESULTS AND DISCUSSION

As mentioned before after setting up the experiments and training the models on the dataset the following are the results.

Accuracy

Machine learning model accuracy is the measurement used to determine which model is best at identifying relationships and patterns between variables in a dataset based on the input, or training, data.

As mentioned before after setting up the experiment and training the models on the datasets the following are the results.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:05/Issue:10/October-2023 Impact Factor- 7.868 www.irjmets.com

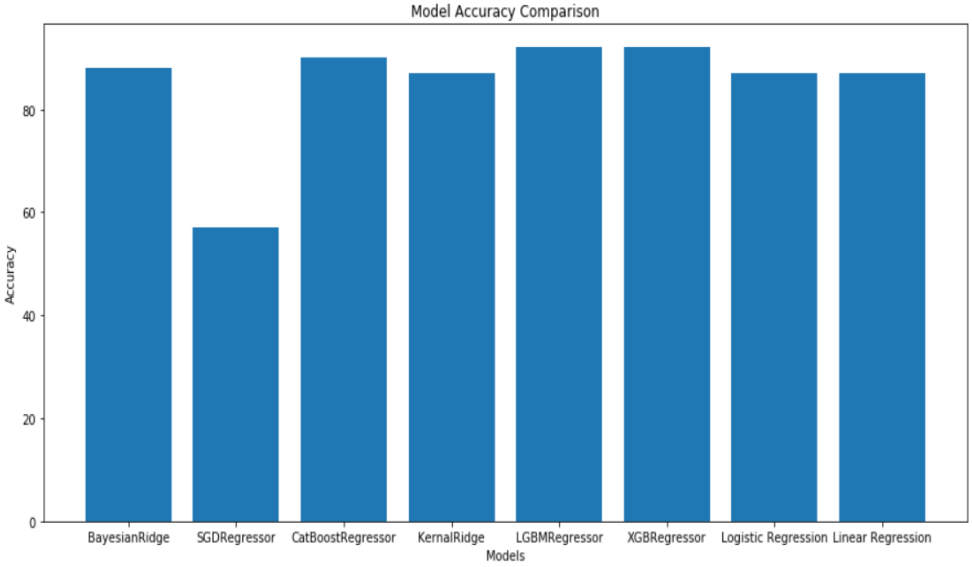

Figure 4: Model Accuracy Comparison

Confusion Matrix

A confusion matrix is a matrix that summarizes the performance of a machine learning model on a set of test data. It is often used to measure the performance of classification models, which aim to predict a categorical label for each input instance. The matrix displays the number of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) produced by the model on the test data.

Some confusion matrices of high-accuracy models are shown below:

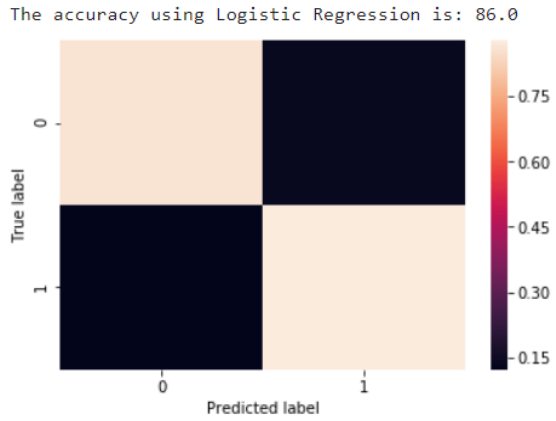

Figure 5: Confusion Matrix for Logistic Regression

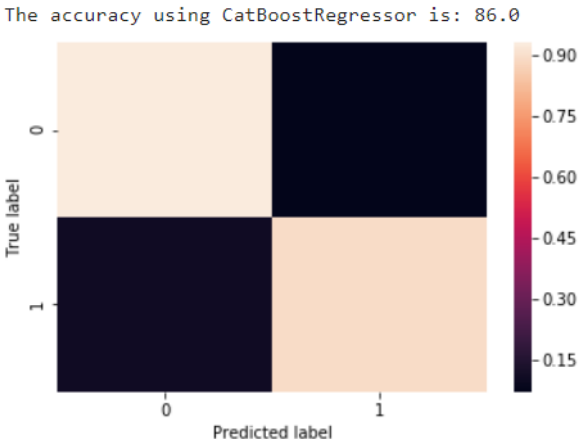

Figure 6: Confusion Matrix for Cat Boost Regressor

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:05/Issue:10/October-2023 Impact Factor- 7.868 www.irjmets.com

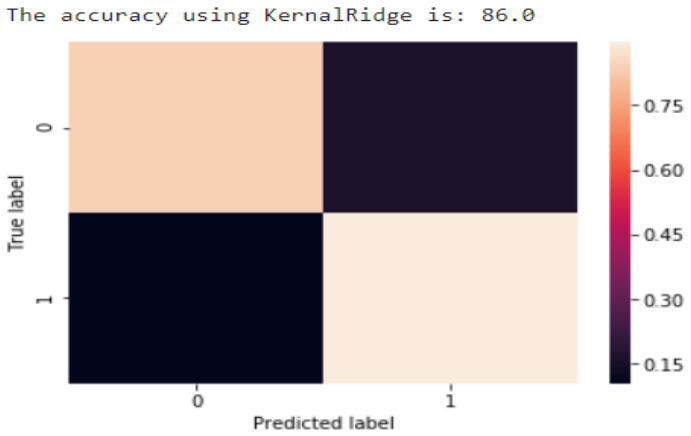

Figure 7: Confusion Matrix for Kernal Ridge

V. CONCLUSION

In conclusion, this study demonstrated the efficacy of regression-based machine learning algorithms in predicting cricket tournament winners. The findings emphasize the importance of considering various factors beyond player statistics to enhance the accuracy of predictions. Future research could explore deep learning algorithms for even more precise predictions.

VI. REFERENCES

[1] A. Bandulasiri, "Predicting the Winner in One Day International Cricket," J. Math. Sci., vol. 3, no. 1.

[2] H. Mittal, D. Rikhari, J. Kumar, and A. K. Singh, "A study on Machine Learning Approaches for Player Performance and Match Results Prediction." arXiv, Aug. 23, 2021. Accessed: Oct. 19, 2023. [Online]. Available: http://arxiv.org/abs/2108.10125

[3] R. Bunker and T. Susnjak, "The Application of Machine Learning Techniques for Predicting Match Results in Team Sport: A Review," J. Artif. Intell. Res., vol. 73, pp. 1285–1322, Apr. 2022, doi: 10.1613/jair.1.13509.