ff

Shaikh Abdul Rafay

September 2024

1 Literature Review

Cricket score prediction has emerged as a significant area of research, driven by advancements in machine learning and data analytics. Various models have been developed to predict match outcomes, player performance, and team scores, using both traditional machine learning algorithms and big data frameworks.

In a study by Mundhe et al. (2021), a web application was developed for live T20 match prediction. The system used Multivariate Polynomial Regression for predicting the final score and a Random Forest Classifier to predict the match winner. The model relied on real-time data scraping, considering factors like the current score, overs bowled, and wickets lost. While moderately accurate, the study highlighted the challenges of predicting outcomes in T20 cricket due to its volatile nature. Their approach emphasized the need for real-time data integration and improved algorithms to handle the unpredictable aspects of cricket [3].

Hatharasinghe and Poravi (2021) explored the application of data mining and machine learning in cricket, focusing on player performance, match simulation, and team selection. They reviewed several models, including those based on historical cricket data and social media analysis, to predict match outcomes. Despite the potential of these models, they noted limitations in predicting international matches due to insufficient training data, particularly for new players. The study suggested that a combination of approaches, integrating both historical and real-time data, could significantly improve accuracy [4].

Awan et al. (2021) applied a big data approach using the Spark ML framework to predict team scores and match outcomes. Their study demonstrated that Spark ML, with its ability to handle large datasets, outperformed traditional frameworks like Scikit-learn. Using linear regression, the model achieved test accuracy rates of 95%, with improved performance in terms of mean absolute error (MAE) and root mean square error (RMSE). The authors highlighted the scalability of big data approaches in cricket analytics and their potential for application across other sports [1].

Bharadwaj et al. (2024) emphasizes the importance of analyzing cricket player performance using machine learning techniques. The study focuses on various parameters such as player consistency, form, performance against spe-

cific opponents, and venue impact. It applies algorithms like Naïve Bayes, Decision Tree, Random Forest, and Support Vector Machine (SVM) to predict batting and bowling outcomes. Among these, Random Forest and Decision Tree were found to be the most effective in capturing player performance dynamics, offering reliable predictions for both batting scores and wickets taken. [2]

Each of these studies contributes to the growing body of work on cricket score prediction, illustrating the evolution from traditional machine learning models to big data frameworks. The integration of real-time data and ensemble methods, such as Random Forest and XGBoost, has led to improved predictive accuracy, but challenges remain, particularly in fast-paced formats like T20 cricket.

References

[1] Mazhar Javed Awan, Syed Arbaz Haider Gilani, Hamza Ramzan, Haitham Nobanee, Awais Yasin, Azlan Mohd Zain, and Rabia Javed. Cricket match analytics using the big data approach. *Electronics*, 10(19), 2021.

[2] F. Bharadwaj, A. Saxena, R. Kumar, R. Kumar, S. Kumar, and Ž. Stević. Player performance predictive analysis in cricket using machine learning. *Revue d'Intelligence Artificielle*, 38(2):449–457, 2024.

[3] Manuka Maduranga Hatharasinghe and Guhanathan Poravi. Data mining and machine learning in cricket match outcome prediction: Missing links. In *2019 IEEE 5th International Conference for Convergence in Technology (I2CT)*, pages 1–4, 2019.

[4] Eeshan Mundhe, Ishan Jain, and Sanskar Shah. Live cricket score prediction web application using machine learning. In *2021 International Conference on Smart Generation Computing, Communication and Networking (SMART GENCON)*, pages 1–6, 2021.