A MACHINE LEARNING APPROACH TO PREDICTING OUTCOMES OF ONE-DAY INTERNATIONAL CRICKET MATCHES

S. Kanapathipillai*[1] and P.S. Palliyaguruge[1]

[1]*Department of Computing, Faculty of Applied Sciences, Rajarata University of Sri Lanka, Mihintale, Sri Lanka.*
**sivasorubyk@gmail.com*

After football, cricket is the second most popular sport in the world. It is played in more than 19 countries, and its popularity grows day by day. Cricket generates an extensive volume of structured and semi-structured data during each match, including ball-by-ball records such as runs scored, wickets taken, bowling economy, batting strike rates, player performance metrics, partnerships, and match events (boundaries, extras). Machine Learning techniques are fantastic at spotting patterns in large and complex datasets, making them a perfect fit for analyzing cricket data and predicting outcomes based on historical trends. Yet, there's still a gap when it comes to reliable tools that can accurately forecast the results of ODI matches, especially those that take into account intricate factors like batting partnerships, all while delivering top-notch performance. The factors chosen for predicting ODI match outcomes such as team, opponent, ground, toss result, batting and bowling decisions, runs scored, wickets lost, number of fours and sixes, extras conceded, run rate, average batting strike rate, wickets taken, best individual bowling performance, and batting partnerships were selected for their proven impact on match results, as highlighted in previous studies and cricket expertise. The Kaggle data source (Asia Cup Cricket Dataset 1984 – 2022) was chosen. Data was processed to several data preprocessing techniques, such as Null value handling and categorical encoding, to input data into the model, which is suitable for model training. For this research, we opted for the Extreme Gradient Boosting (XGBoost) algorithm to predict final match results. XGBoost stood out because of its excellent predictive capabilities, its knack for handling missing values, built-in regularization, and its resilience against overfitting, making it especially effective for structured, tabular cricket match data. It produced an accuracy of 98% as well. Furthermore, a regression-based ranking system was developed to assess player performance, as regression models are well-suited for predicting continuous-valued outcomes such as player contribution scores. For the toss decision, Support Vector Machine (SVM) was selected due to its strong performance in binary classification tasks, effectively handling high-dimensional data and small sample sizes. We integrated the predictive models into a user-friendly web application built with the Flask framework, using Pickle for model serialization. This application serves captains, coaches, analysts, and cricket enthusiasts by offering actionable insights for strategic planning, thus contributing to the expanding field of sports analytics and empowering data-driven decision-making in cricket.

Keywords: *Machine learning, ODI cricket prediction, XGBoost, SVM, Feature selection, Sports analytics, Ensemble learning.*

Acknowledgements: I extend my heartfelt gratitude to all the academic and non-academic staff at Rajarata University for their contributions to my academic journey.