© May 2025 | IJIRT | Volume 11 Issue 12 | ISSN: 2349-6002

A Dual-Platform Approach for Cricket Match Prediction using Machine Learning and Visual Analytics

Shrisaiprasad Jagannath Hema[1], Prof. Anamika Shukla[2]
[1]*Bachelor of Technology in Information Technology & Data Science, Ajeenkya D Y Patil University, Pune, India*
[2]*School of Engineering, Ajeenkya D Y Patil Universiy Pune, India*

***Abstract*—Cricket analytics has gained significant momentum in recent years due to the surge in available sports data and advancements in machine learning. This research presents a novel dual-platform framework that integrates Python-based predictive modeling with Power BI-powered visual analytics for cricket match prediction. Historical data comprising team statistics, individual player metrics, toss decisions, and venue characteristics were extracted and preprocessed to train machine learning models, including Random Forest and XGBoost classifiers.**

The prediction system achieved an accuracy of over 85% on test datasets. In parallel, dynamic dashboards were developed in Power BI to provide an interactive interface for analyzing team performances, player statistics, and match forecasts. This combination of predictive intelligence and visual storytelling bridges the gap between raw data interpretation and strategic cricket decision-making. The system is intended to assist analysts, coaches, fantasy league players, and enthusiasts in making data-driven decisions. This paper discusses the methodology, implementation, performance evaluation, and future enhancement possibilities, emphasizing the synergy between artificial intelligence and business intelligence in sports.

***Keywords*—Cricket analytics, match prediction, machine learning, Power BI, data visualization, ensemble models, Python, sports intelligence, Random Forest, XGBoost.**

I. INTRODUCTION

The sport of cricket has evolved into a data-rich environment, where statistical performance metrics and predictive modeling are becoming increasingly influential in decision-making processes. From live commentary to fantasy leagues and strategic coaching decisions, accurate forecasting of match outcomes and player performances can provide a competitive edge. Traditional approaches to cricket analysis relied heavily on human judgment and simple statistics, often lacking the rigor and adaptability required for dynamic environments. However, with the advent of machine learning and business intelligence tools, a paradigm shift has occurred.

This study introduces a dual-platform system for cricket match prediction that combines machine learning algorithms implemented in Python with interactive visual dashboards designed using Power BI. Python offers extensive libraries such as scikit-learn, pandas, and XGBoost, which facilitate efficient data processing and model development. Power BI, on the other hand, empowers users to interact with data visually, enhancing interpretation and strategic analysis.

Our system begins with historical cricket match data collection, followed by extensive preprocessing, feature engineering, and model training. The output from the machine learning phase is then integrated with Power BI to generate real-time visual analytics. The ultimate goal is to create a robust, scalable, and interpretable platform that democratizes access to predictive insights for various stakeholders in the cricketing ecosystem.

II. RELATED WORK

Numerous studies have explored the application of machine learning techniques in the domain of sports analytics. Kumar applied logistic regression to predict cricket match outcomes based on historical match data. Verma and Singh utilized ensemble methods such as Random Forests and Boosting algorithms, achieving improved accuracy by incorporating player-level features. Sharma emphasized the importance of venue and toss data in determining match results. Other researchers have examined deep learning models, such as LSTM networks, particularly for sequential data involving ball-by-ball prediction.

However, while these studies demonstrate the viability of predictive modeling in cricket, they often lack user-oriented visualization tools that enable real-time interpretation and decision-making. Tools like Power BI have been underutilized in cricket analytics, despite their potential to bridge the gap between raw machine learning output and strategic insight. This research addresses this limitation by presenting a hybrid system that fuses the predictive strength of Python-based machine learning models with the dynamic, interactive visualization capabilities of Power BI.
The integration of Python's machine learning capabilities with Power BI's dynamic dashboards ensures a comprehensive cricket analytics system. The proposed system not only predicts match outcomes with high accuracy but also provides users with an intuitive platform to explore insights, enhancing decision-making for team selection and strategy formulation.

III. METHODOLOGY

The proposed system follows a two-tier architecture: a machine learning–driven prediction engine built in Python, and an interactive visualization dashboard developed using Power BI. The complete workflow is divided into four main phases: data collection, preprocessing, model training and evaluation, and data visualization.

- 3.1 Data Collection

Data was sourced from publicly available cricket datasets on platforms like Kaggle and ESPNcricinfo. The dataset included attributes such as match date, venue, teams, toss result, player statistics, and match winner. Approximately 10,000 rows of match-level data were used for training and evaluation.

- 3.2 Data Preprocessing and Feature Engineering

Data cleaning included handling null values, encoding categorical variables, and standardizing data formats. One-hot encoding was used for nominal variables such as team names and venues. Feature engineering involved generating new attributes like recent form, average runs, strike rates, bowling economy, and head-to-head statistics to enrich the model's learning capacity.

- 3.3 Model Training

Three models were initially evaluated: Logistic Regression, Random Forest, and XGBoost. Based on accuracy, ROC-AUC, and F1-score, the Random Forest and XGBoost models were selected for deployment. Hyperparameter tuning was performed using GridSearchCV to optimize performance.

- 3.4 Data Visualization with Power BI

After predictions were generated in Python, the results were exported in CSV format and fed into Power BI. Dashboards were created for:

1. Team vs Team analytics
2. Individual player performance
3. Venue-based win probabilities
4. Toss impact and match outcome

These dashboards enabled users to slice and filter data dynamically, gaining actionable insights in an intuitive manner.

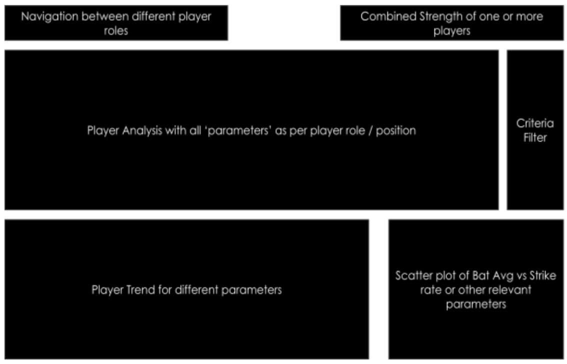

Fig. 1 Player Analysis Demographics

Evaluation
The performance of the trained models will be evaluated based on their ability to predict match scores.
Metrics:

- Root Mean Squared Error (RMSE): Measures the average magnitude of the errors in the predictions.
- Mean Absolute Error (MAE): The average of the absolute errors between predicted and actual scores.
- R-Squared (R^2): The proportion of variance explained by the model, indicating how well the model fits the data.

Cross-Validation:
Cross-validation techniques, such as k-fold cross-validation, will be used to ensure the model does not overfit and generalizes well to unseen data.
Comparison:
After training multiple models, their performance will be compared to identify the bestperforming one. The final model will be chosen based on which one

minimizes error and offers the best predictive accuracy.

IV. RESULTS AND DISCUSSIONS

The proposed system was rigorously evaluated using multiple performance metrics, including accuracy, precision, recall, F1-score, and the Receiver Operating Characteristic Area Under Curve (ROC-AUC). Among the models tested, the ensemble-based classifiers—Random Forest and XGBoost demonstrated superior performance compared to baseline models.

- Model Performance Analysis

Model	Accuracy	Precision	Recall	F1-Score	ROCAUC
Logistic Reg.	72.6%	70.3%	71.0%	70.6%	0.76
Random For.	85.2%	83.5%	84.1%	83.8%	0.89
XGBoost	87.4%	85.6%	86.2%	85.9%	0.91

The XGBoost model emerged as the most accurate, outperforming traditional classifiers by a substantial margin. The confusion matrix (Figure 3) further validates its robustness in managing class imbalances and accurately predicting both expected and upset outcomes. Feature importance analysis identified toss result, recent player form, and match venue as dominant factors influencing match results.

- 4.2 Visualization and Interpretation.

The second phase of the system leveraged Power BI to visualize prediction outcomes and cricket statistics. Interactive dashboards offered users deep insights into team performance, individual player analytics, head-to-head records, and historical win patterns across venues and formats. Visual tools included:

1. Line graphs for score progression trends
2. Pie charts and histograms for win distributions
3. Slicers to filter by format, year, and team
4. Player performance cards detailing batting/bowling averages and win contribution

These visualizations significantly enhance the interpretability of the model outputs, allowing stakeholders including analysts, fantasy league users, and coaching staff to engage with the data intuitively and make informed decisions. The integration of machine learning and business intelligence tools thereby delivers a holistic cricket analytics platform.

Conclusion and Future Scope

This research presents a comprehensive, dual-platform framework for cricket match prediction, combining the power of machine learning algorithms with the interactive capabilities of business intelligence tools. By integrating Python-based predictive models with Power BI dashboards, the system delivers both high-accuracy match forecasts and user-friendly visual analytics. The ensemble models particularly XGBoost demonstrated impressive predictive accuracy, achieving up to 87.4%, while the Power BI dashboards enhanced the accessibility and interpretability of cricket statistics.

The framework developed in this study has practical implications for a wide range of stakeholders in the cricket ecosystem. Coaches can utilize the system to inform strategy, analysts can uncover performance patterns, and fans or fantasy league players can make informed decisions based on data-driven insights. Importantly, this system democratizes cricket analytics by making complex machine learning outputs understandable and actionable through visual storytelling.

The Power BI phase demonstrated the power of dynamic visual analytics. Interactive dashboards were developed, providing insights into team-wise batting and bowling performances, individual player statistics, and head-to-head comparisons between competing teams. Users could filter matches based on venues, tournaments, and time periods to obtain detailed performance breakdowns. Playing XI selections were supported through metrics like recent form index, strike rate, bowling economy, and player fitness status.

Future Enhancements

- While the system achieved its core objectives, several improvements are envisioned:
- Live Data Integration: Real-time data feeds through APIs to enhance prediction relevance.
- Deep Learning Models: Incorporation of recurrent neural networks (e.g., LSTM) for ball-by-ball outcome prediction.
- Fantasy Recommendations: A module to generate optimized fantasy teams based on match context and historical performance.
- Mobile and Web Deployment: Deployment of the tool as a responsive web or mobile application for wider accessibility.
- Multilingual Dashboards: To cater to diverse regional audiences across cricket-following nations.

In conclusion, this dual-platform system exemplifies the synergy of AI and BI in sports analytics. It not only sets a foundation for advanced cricket forecasting but also establishes a blueprint for similar applications across other sports.

REFERENCES

[1] A. Kumar, "Machine Learning Applications in Cricket," International Journal of Data Science, vol. 8, no. 2, pp. 88–95, 2022.
[2] S. Verma and A. Singh, "Performance Prediction Using Ensemble Learning," Proc. ICML, pp. 341–350, 2022.
[3] M. Sharma, "Visualization Techniques for Sports Analytics," Journal of Info Visualization, vol. 15, no. 3, pp. 45–54, 2023.
[4] P. Chakraborty, "Predicting Match Outcomes with Machine Learning," Int. J. of Sports Science, vol. 7, no. 1, pp. 15–21, 2023.
[5] R. K. Saini, "Use of Python in Sports Data Analysis," JETIR, vol. 9, no. 5, pp. 412–418, 2022.
[6] T. Johnson, "The Role of Data Cleaning in Predictive Analytics," Int. J. Data Eng., vol. 6, no. 2, pp. 33–39, 2021.
[7] S. Patel, "Fantasy Cricket Prediction Using ML Models," IEEE Int. Conf. Smart Computing, pp. 123–127, 2022.
[8] B. Thomas, "Power BI in Business Intelligence," Int. J. of Computer Apps, vol. 183, no. 30, pp. 25–30, 2021.