■ Cricket Machine Learning Project: Data■Driven Playing XI Selection

Problem Statement

In team cricket, selecting the optimal playing XI for a match is a key strategic decision. Traditional selection often relies on intuition, recent form, or past experience, rather than holistic data■driven insights. The goal of this project is to develop a machine■learning based system that recommends an optimal team of 11 players for an upcoming match by analysing historical player performance, opposition data, venue and match conditions.

Dataset & Source

Dataset Sources: - Kaggle: Cricket data – ICC Cricket - Cricsheet: Cricsheet.org (ball-by-ball match data)

Exploratory Data Analysis (EDA) Questions

1 Distribution of players by role (batsman, bowler, all■rounder, wicketkeeper). 2 Typical batting averages / strike rates for batsmen in different formats. 3 Comparison between recent form and career averages.
4 Impact of venue/location on player performance.
5 Effect of opponent strength on player output.
6 Correlation between different metrics (economy vs wickets, strike■rate vs runs). 7 Performance differences by age or experience.
8 Role■balance of players in successful vs unsuccessful teams.
9 Effect of fatigue or rest days on performance.
10 Venue or pitch trends affecting player selection.

Visualization Questions

1 Histogram of batting averages and strike rates by format.
2 Box■plots comparing recent form vs career averages.
3 Heatmap of correlation matrix between performance metrics.
4 Bar chart of average performance at home vs away.
5 Scatter plot of age vs performance by role.
6 Line chart showing performance trends over time.
7 Stacked bar chart of role distribution in winning vs losing teams.
8 Venue performance map or bar chart.
9 Opponent strength vs runs/wickets violin plot.
10 Feature importance plot after model training.

ML Model Training & Selection

Steps for Model Training: 1. Data Preprocessing and Feature Engineering. 2. Baseline Modelling: Linear/Logistic Regression, Decision Trees. 3. Advanced Modelling: Random Forest, XGBoost, or Neural Networks. 4. Team Selection Optimization: Integer programming or greedy algorithms to pick 11 players under constraints. 5. Model Evaluation: Use accuracy, F1-score, RMSE, or MAE based on prediction type. 6. Deployment: Build a function or dashboard for real-time team recommendations.

Expected Outcome & Impact

Expected Outcomes: - Data-driven insights into player performance trends. - Predictive model for upcoming match player contributions. - Selection of an optimized Playing XI. - Evaluation of data-based decision support for team selection.