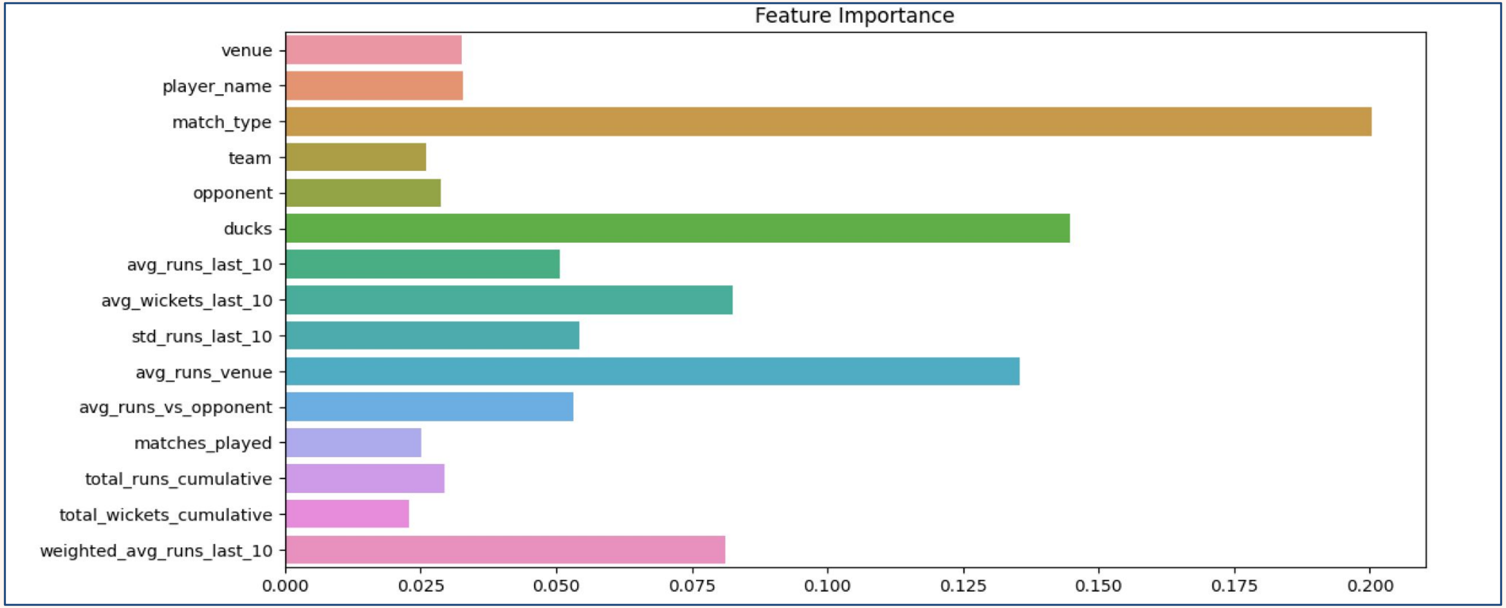
Dream 11 Team Prediction

By Team 84

Problem Statement

Develop a predictive ML model to help Dream11 users create winning cricket fantasy teams by forecasting player performance using historical data and match-specific factors. Provide a simple and user-friendly interface to make insights easy to understand and use.

Feature Importance



Match type, average runs in the last 10 matches, and average wickets in the last 10 matches were identified as the most important features based on observed trends in the data.

Models Explored for Fantasy Points Prediction

To predict fantasy points for players, we experimented with various machine learning models, each contributing unique strengths to our approach:

- **XGBoost** and **CatBoost**: Utilized for their ability to handle complex data and provide accurate, scalable predictions.
- **Random Forest**: Effective for capturing diverse performance metrics and reducing overfitting.
- **K-Nearest Neighbors (KNN)**: Used as a baseline for its simplicity and interpretability.

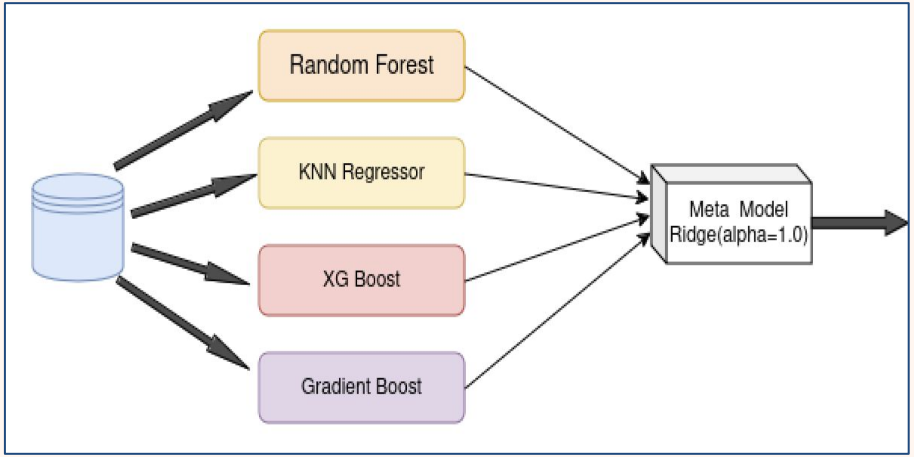
These models helped analyze player performance, identify key trends, and refine the pipeline for optimal prediction accuracy.

Model Stacking

1. **XGBoost, CatBoost, Random Forest, and KNN** were used as base models to predict fantasy points.

2. **Individual Model Limitations:**
    a. Overfitting
    b. Bias towards certain feature distribution

3. **Advantages of Model Stacking:**
    a. Model stacking combines the outputs of multiple base models using a meta-model
    b. Eg. Logistic regression or Weighted Averaging
    c. Leverages their collective strengths, captures complex relationships in the data, ensures minimal error, and achieves maximum predictive performance.



Probability Model

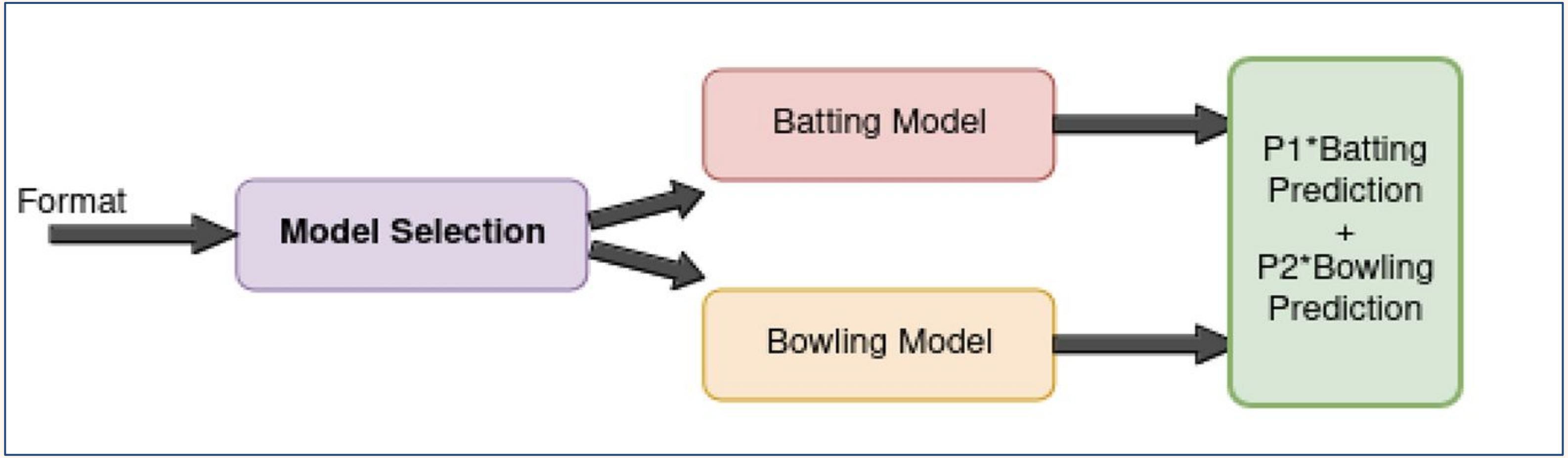
1. This method recognizes that players may not always bat or bowl in a match
2. Assigning probabilities to these actions provides a more realistic measure of their contribution.
3. Trained separate models for each format—**T20, ODI, and Test**

FP = *P_batting* * Predicted Points from Batting Model + *P_bowling* * Predicted Points from Bowling Model

FP → Predicted Fantasy Points based on Players Performance
P_batting → Probability of a player batting in a match
P_bowling → Probability of a player bowling in a match

Probability Model

1. **Batsman -** *P_batting* is close to 1, while *P_bowling* is near 0.
2. **Bowler -** *P_bowling* will be close to 1, and *P_batting* near 0.
3. **All-Rounders:** Both probabilities will have significant values, resulting in a combination of batting and bowling points contributing to their total fantasy points



Understanding Importance Of H2H:

1. Previous models provided robust predictions by analyzing historical performances of players
2. **Limitation -** They do not account for matchups between players in a live match scenario.
3. **Example -** Consider Virat Kohli, whose overall batting statistics are exemplary.
4. If in a match, he faces a bowling attack comprising players like Mitchell Starc, Jasprit Bumrah, The probability of him scoring high fantasy points decreases.
5. Batter of opposition team with a weaker matchup is a better option because his likelihood of scoring more runs and fantasy points is more.

Key Observation: A batter with favorable matchups against a weaker bowling attack is likely to contribute more significantly to a fantasy team than one with strong historical stats but a challenging opposition.

Knowledge Graph for Player Performance Prediction

Graph Structure

1. **Main Node**: Represents the player (Batsman or Bowler) for prediction.
2. **Neighbor Nodes**: Opposing team's players (Bowlers for Batsman, Batsmen for Bowler).

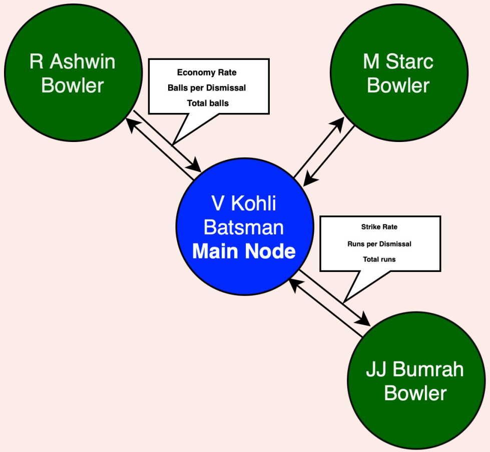
Edge Types and Metrics

Batter-to-Bowler Edge:

- Runs per Dismissal
- Strike Rate
- Total Runs
- Total Balls

Bowler-to-Batter Edge:

- Dismissal Rate
- Economy Rate
- Balls per Dismissal
- Total Dismissals



Advantages of Graph Utilization

- Structured representation of player matchups.
- Detailed insights into interactions and dynamics.
- Critical for nuanced and accurate predictions.

Balancing Head-to-Head Stats:

1. H2H matchups provide valuable insights into how players are likely to perform against specific opposition
2. But it is necessary to realize that the predicted fantasy points based on these stats may not fully reflect a player's current form.
3. For instance **Virat Kohli:**
   a. We considered strong overall record of Virat Kohli
   b. We have also considered that despite his strong batting record, he may not be able to perform at his best against most formidable bowlers.
   c. Did not account for Kohli's ability to perform well even in challenging situations and his recent performances.
4. Included Player's recent performances
5. Calculated the **mean** and **standard deviation of his fantasy points** over these recent matches to provide a more dynamic representation of his current form.

Performance and Knowledge Fusion Pipeline

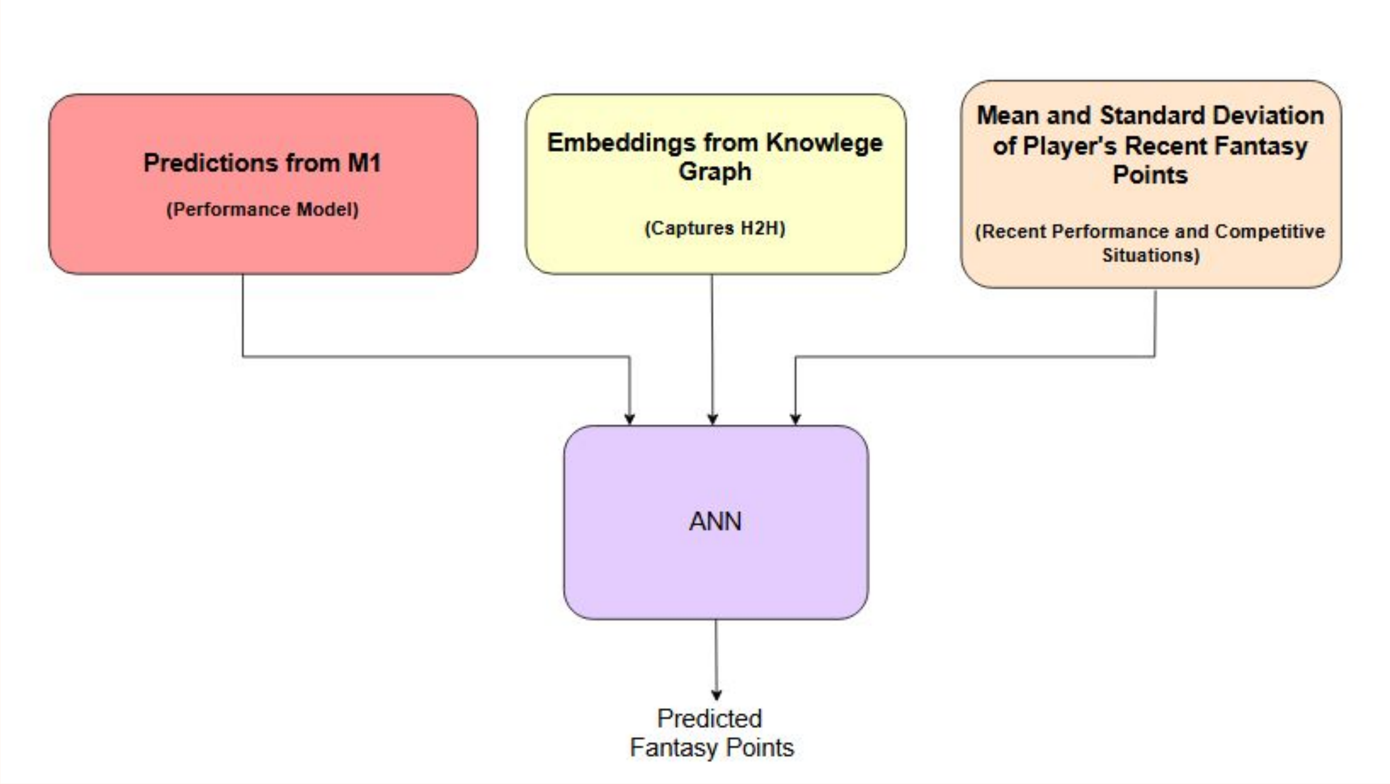
Input Features

1. **Predictions from M1 model (Performance Based Model) -** Captures the players performance record (Format related, Venue Related, Recent Runs etc.)
2. **Embeddings from Knowledge Graph -** Accounts for H2H statistics and captures the relative performance relationship of players
3. **Mean and Standard Deviation of Recent Fantasy Points -** Balances the H2H statistics in order to account for players recent performance and ability to face challenging situations.

ANN

1. These Input Features are fed into the ANN for Final Prediction
2. ANN is capable of learning complex non-linear relationships between the features and the target variable

Architecture Diagram:



Results

Best Performer - <u>Performance and Knowledge Fusion Pipeline</u>

1. **MAE**: 120.48, **MPAE**: 20.13%
2. Combines **probability-based regression**, **knowledge graph embeddings**, and **recent stats** via an ANN for robust predictions.

Second Best - <u>Model Stacking</u>

- Combines predictions from multiple base models (**XGBoost, CatBoost, Random Forest**).
- **MAE**: 143.50, but lacks advanced features like knowledge graph embeddings.

Standalone Models

- **<u>XGBoost & CatBoost</u>**: Effective for structured data, but less contextual.
- **<u>Random Forest</u>**: Higher MAE and MPAE, struggles with non-linear relationships.
- **<u>Linear Regression</u>**: Highest MAE (192.91), weakest for complex feature interactions.

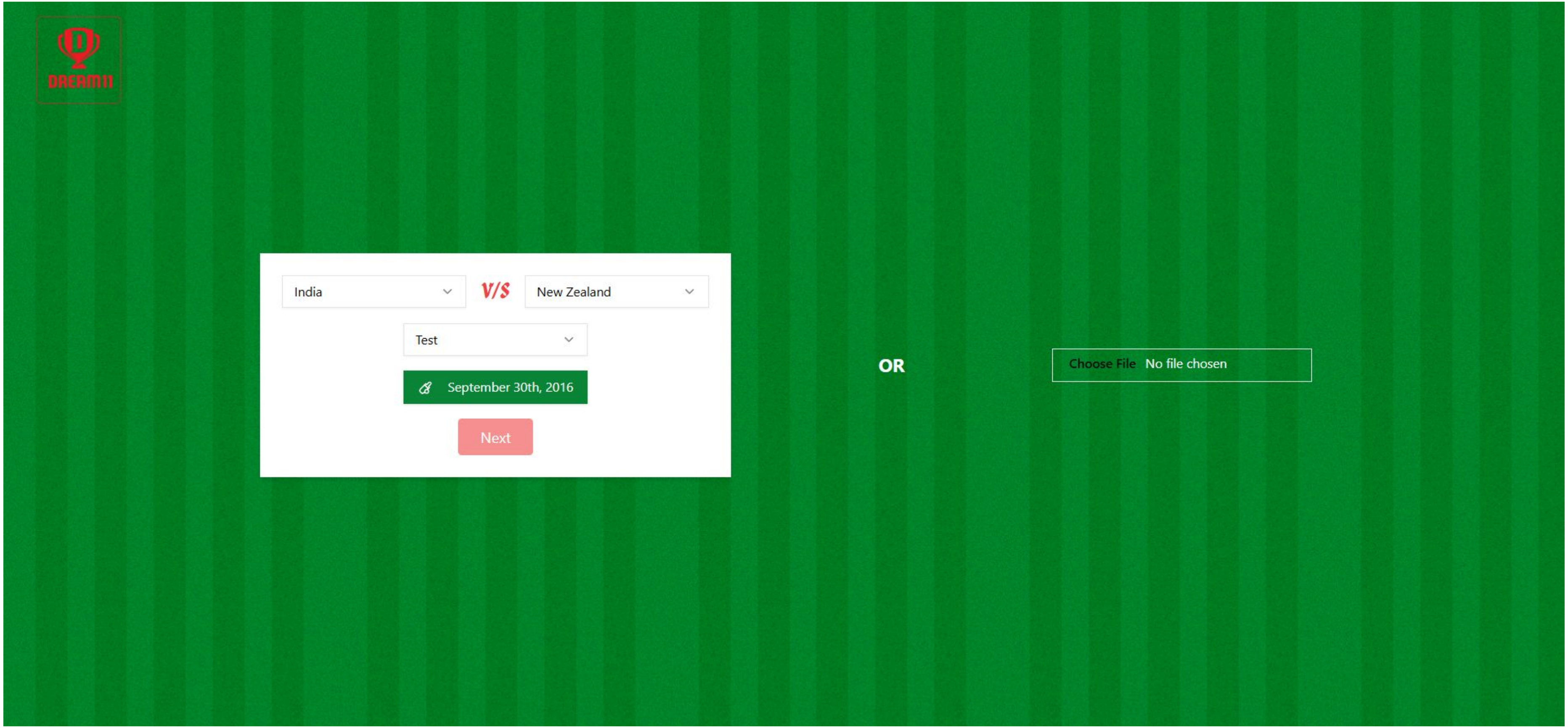
Model	MAE	MPAE(%)
Complete Pipeline	120.48	20.13
Model Stacking	143.50	11.47
XGBoost	164.44	8.21
CatBoost	175.98	6.89
Linear Regression	192.91	3.34
Random Forest	178.78	6.72

Future Enhancements:

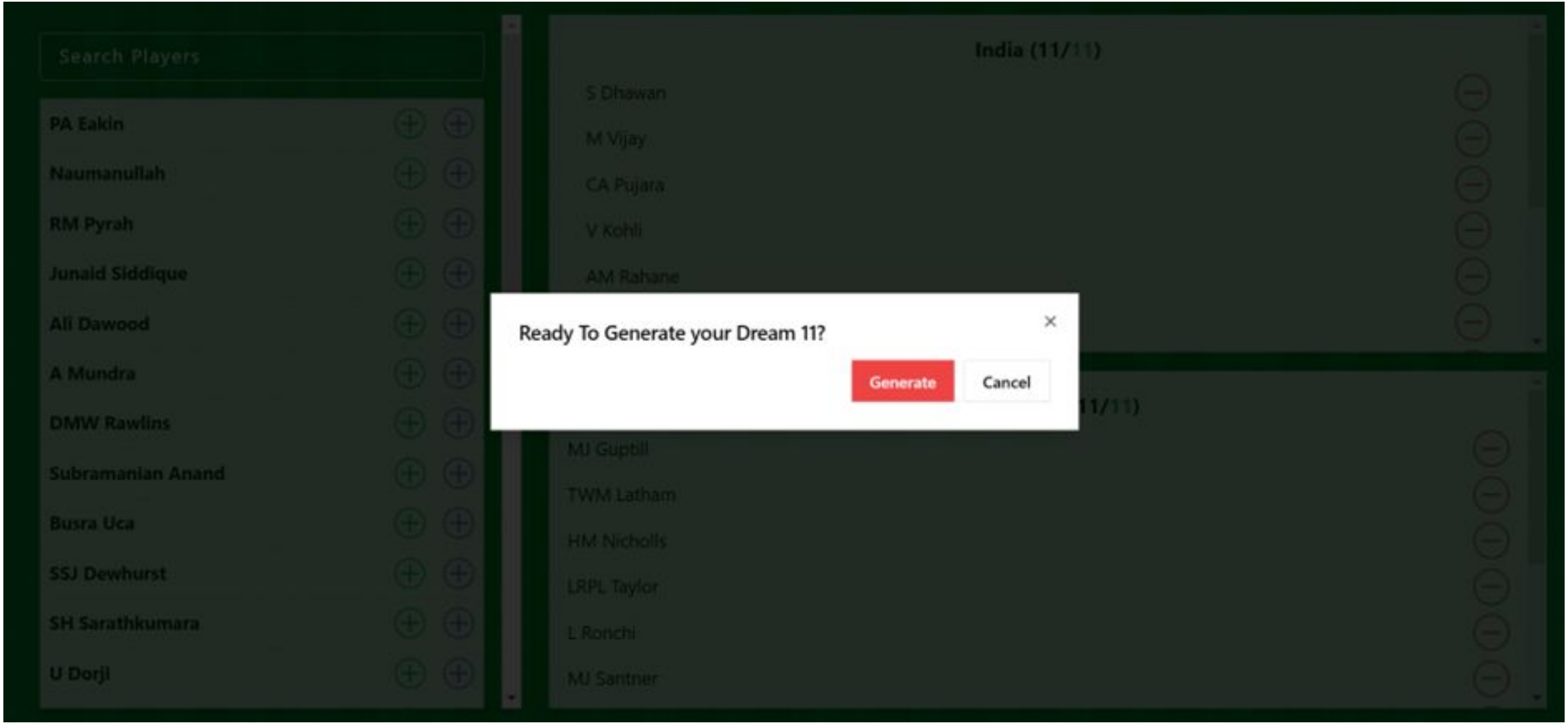
1. **Application of Transformers -**
    a. Experimented with Transformers with basic encoding of Input. Yielded poor results
    b. Incorporated embeddings—a standard technique for transformers—we achieved a significant reduction in error
    c. Dropped it because increased model size resulted in extremely long training times.
2. **Incorporating Pitch Type -** Accounting for pitch characteristics could significantly enhance the model's ability to capture match conditions.
    a. Divide the pitch into 3 or 4 types eg: Green and Grassy, Dry and Dusty etc.
    b. Pitch conditions at a venue change over time
    c. Apply clustering algorithm to find the type of pitch in a match at a particular time.
3. **Semantic Analysis -**
    a. Combine Insights from web scraping and natural language processing
    b. News about player form, injuries, team strategies can help model analyze real time data.

Demonstration:

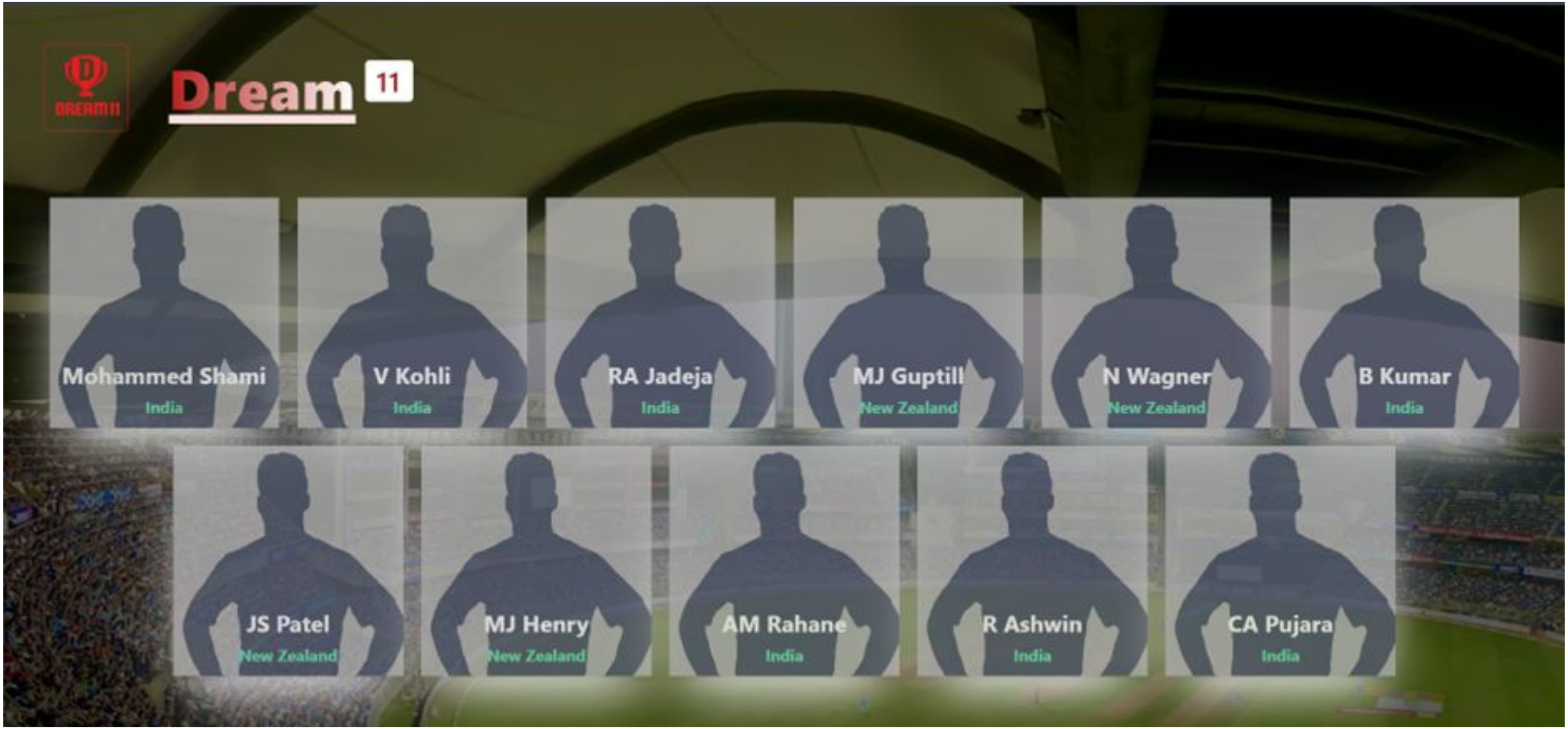
PRODUCT UI


DREAM11
India
V/S
New Zealand
Test
September 30th, 2016
Next
OR
Choose File No file chosen

Team Selection



Dream Team


Dream 11
Mohammed Shami
India
V Kohli
India
RA Jadeja
India
MJ Guptill
New Zealand
N Wagner
New Zealand
B Kumar
India
JS Patel
New Zealand
MJ Henry
New Zealand
AM Rahane
India
R Ashwin
India
CA Pujara
India

Player Stats

Runs
Runs
Average
100's
50's
Runs
2011 2012 2013 2014 2015 2016
Average
Bowler Overall Metrics
Average
2011 2012 2013 2014 2015
Runs
Batting Last 5 Metrics
Runs
2015-12-03 2016-07-21 2016-07-30 2016-08-09 2016-09-22
Economy
Bowler Last 5 Metrics
Economy
2012-01-24 2013-12-18 2014-02-06 2014-02-14 2015-12-03

Player Detailed Analysis

V Kohli India

Batsman

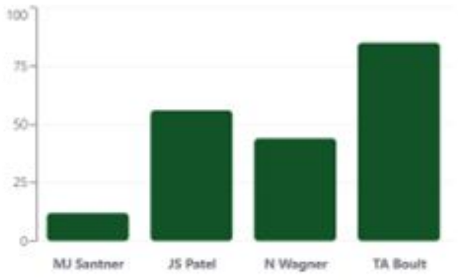
Runs
3272

Strike Rate
53

Batting Average
44

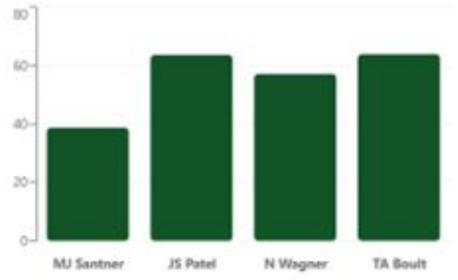
A true master of the game, V Kohli has amassed 3,272 runs with a strike rate of 53.00, showcasing their capability to score efficiently. With an average of 44.00, they bring both stability and explosiveness to the crease. V Kohli has consistently appeared in the top 11 in 33.33% of the last ten matches, reinforcing their pivotal role in the squad.

Head to Head

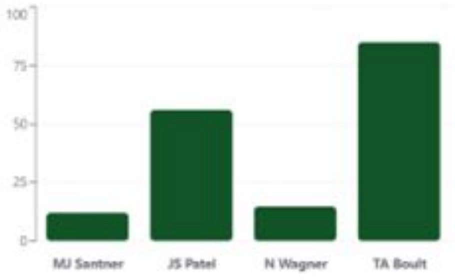


100
75
50
25
0
MJ Santner
JS Patel
N Wagner
TA Boult

Runs Scored
Against other players
197
Total Runs Scored

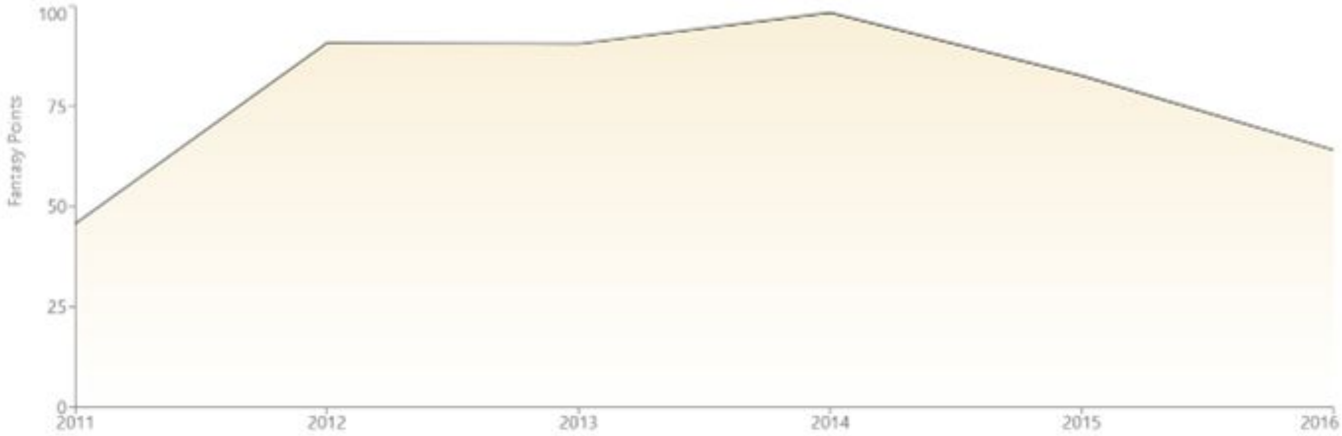


80
60
40
20
0
MJ Santner
JS Patel
N Wagner
TA Boult

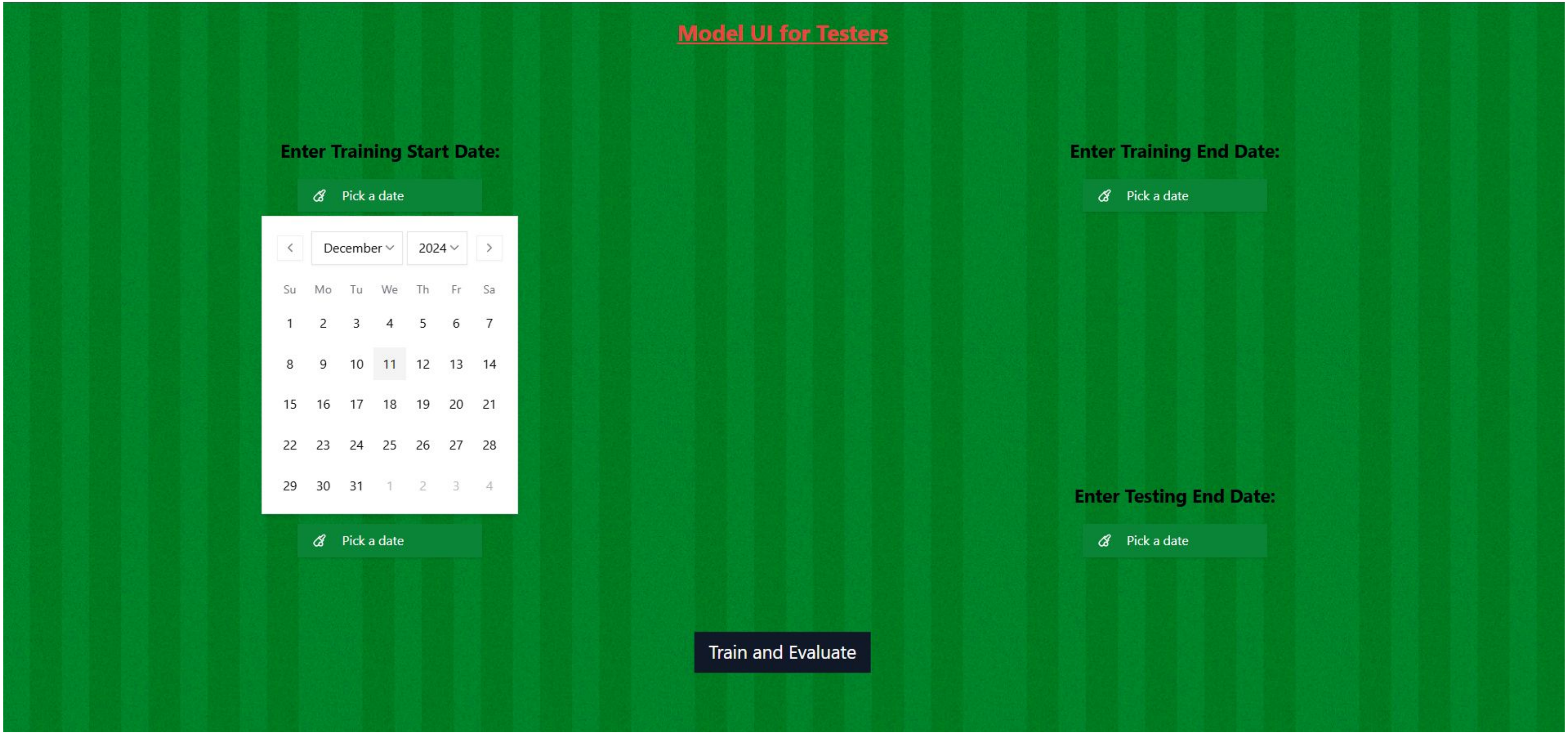


100
75
50
25
0
MJ Santner
JS Patel
N Wagner
TA Boult

Fantasy Points Over the years



MODEL UI


Model UI for Testers
Enter Training Start Date:
Pick a date
December 2024
Su Mo Tu We Th Fr Sa
1 2 3 4 5 6 7
8 9 10 11 12 13 14
15 16 17 18 19 20 21
22 23 24 25 26 27 28
29 30 31 1 2 3 4
Pick a date
Enter Training End Date:
Pick a date
Enter Testing End Date:
Pick a date
Train and Evaluate

THANK YOU SO MUCH!