IPL Score Prediction

ADITYA KAMAL (2106277) - Slide 1 to Slide 4

CHHAGAN RAM CHOUDHARY (2106110) - Slide 5 to Slide 7

JEET AGARWAL (210120) - Slide 8 to Slide 11

IPL Score Prediction Model: Introduction

Objective: To predict the final score of IPL matches using historical data (2008-2017) and key match factors like teams, scores, and overs.

Approach: We apply Linear Regression, Decision Tree, and Random Forest algorithms to predict outcomes, enhancing the cricket experience for fans and strategists.

IPL Score Prediction Model: Data to Decision

Data Preparation:

- Import and explore IPL dataset.
- Clean data by removing irrelevant columns.
- Retain only consistent teams and data post 5 overs.

Feature Engineering:

- Convert ‘date’ to datetime format.
- Apply one-hot encoding to categorical team variables.

Modeling:

- Split data into training (<=2016) and test (>=2017) sets.
- Train Linear Regression, Decision Tree, and Random Forest models.
- Evaluate and select the best model for score prediction.

Accurate predictions are vital in cricket for formulating winning strategies and guiding betting decisions, offering a competitive edge by anticipating match outcomes.

Key Features and Their Rationale in IPL Score Prediction

Date: Provides temporal context, crucial for capturing trends over seasons.

bat_team & **bowl_team**: Indicate competing teams, whose strengths and weaknesses significantly influence the score.

runs & **wickets**: Reflect the current match situation, directly impacting the projected score.

overs: Shows the match progress, affecting the batting team's strategy.

runs_last_5 & **wickets_last_5**: Offer insights into the recent momentum and pressure, which are predictive of future performance.

total: The final score to predict, serving as the target variable in our model.

Exploratory Data Analysis: Correlation and Patterns

In our IPL score prediction model, statistical analysis plays a pivotal role in understanding the relationships between different features. Here's how we approached it:

- **Numeric Dataframe Creation**: We isolated numeric features to focus on quantifiable variables that influence match outcomes.
- **Correlation Matrix**: By calculating the correlation between these features, we identified how strongly each pair of variables is related.
- **Heatmap Visualization**: Using `seaborn`, we plotted a heatmap to visually represent these correlations, making it easier to spot high correlation coefficients.

Key Insights:

- **Positive Correlations**: Features like 'runs' and 'runs_last_5' showed a strong positive correlation, indicating that a higher score in the last five overs often contributes to a larger total.
- **Negative Correlations**: 'wickets' and 'runs_last_5' might display a negative correlation, suggesting that taking more wickets could slow down scoring.
- **No Correlation**: Some features may have little to no correlation, implying no direct linear relationship.

Feature Engineering in IPL Score Prediction

1

One-Hot Encoding:

- Categorical variables like ‘bat_team’ and ‘bowl_team’ are transformed into a numerical format.
- Each team is represented as a binary vector, ensuring the model interprets these features correctly.

2

Streamlining the Dataset:

- Unnecessary columns such as ‘mid’, ‘venue’, and player names are removed to focus on relevant match data.
- Only matches involving consistent teams are retained to maintain data quality.
- Data from the first 5 overs is excluded to avoid early-game volatility.
- The ‘date’ column is converted to a datetime object for chronological analysis.

Model Training Overview

Linear Regression: Predicts scores based on a linear relationship between input features.

Decision Tree: Maps decisions using a tree-like model for non-linear predictions.

Random Forest: Combines multiple trees for a more accurate and stable prediction.

Data Splitting for Training and Testing:

- Data up to 2016 used for training.
- Data from 2017 onwards for testing.

Result: Trained models to forecast IPL scores, each offering unique insights and accuracy.

Model Prediction Results & Insights

Prediction Range:

- The model estimates a final score of 185 to 200 for Chennai Super Kings batting against Royal Challengers Bangalore.

Key Insight:

- Performance in the last 5 overs is crucial, reflecting the model's adaptability to game dynamics.

Algorithm Comparison:

- Linear Regression offers simplicity, Decision Trees capture complexity, and Random Forest provides the best generalization.

Applications for the IPL Score Prediction Model

Live Match Prediction

Teams can use the model to make real-time decisions, like setting a target score or planning bowling changes.

Comprehensive Analytics

Analyze player performance and match conditions to strategize for upcoming games.

Betting and Gambling

Bettors can use predictions to make informed wagers, though it’s important to gamble responsibly.

Conclusion and Next Steps

Success: The model successfully predicts IPL scores within a reasonable range, enhancing match strategy and betting decisions.

R2 Score: 0.75 for Linear Regression Model.

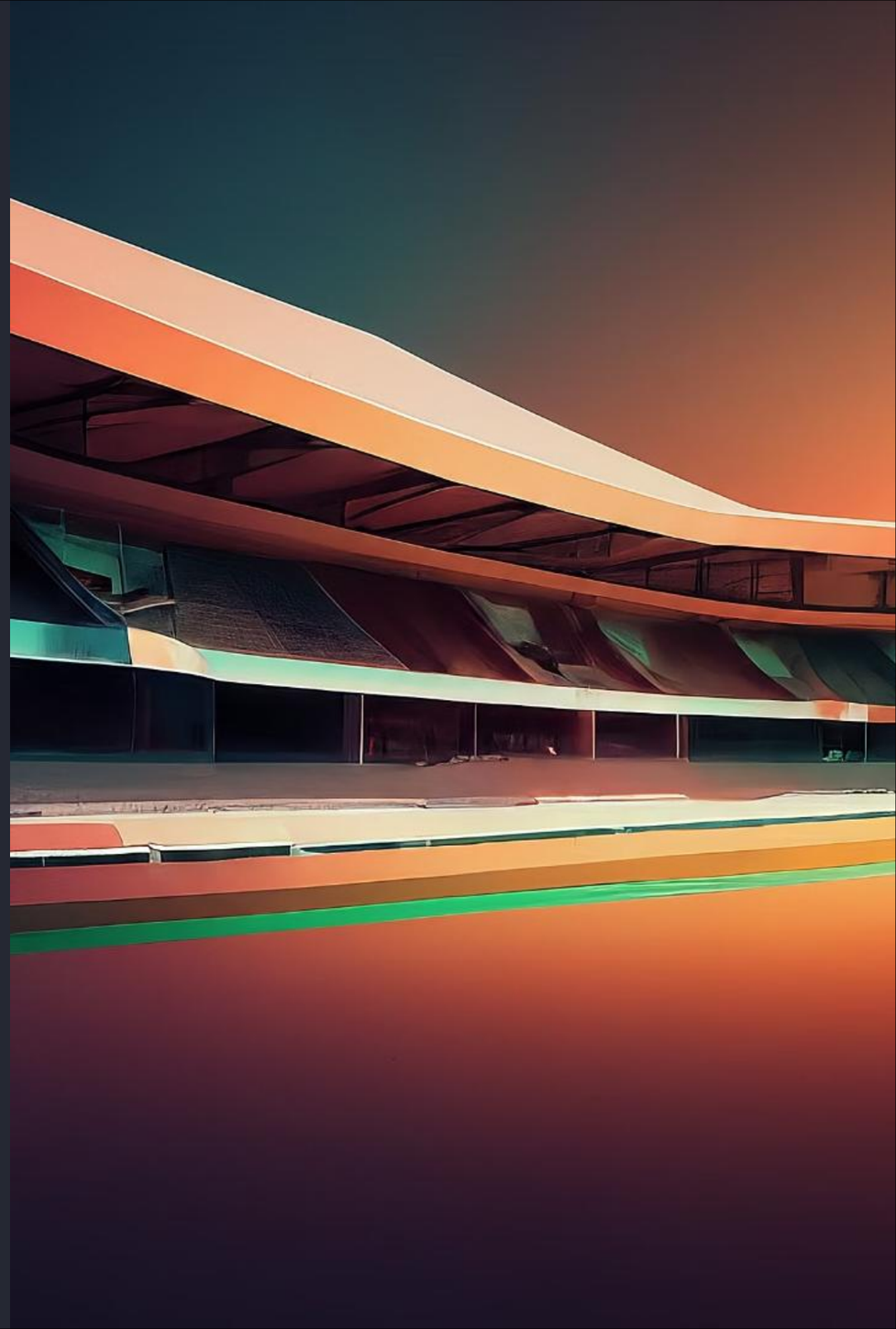
Deployment Next Steps:

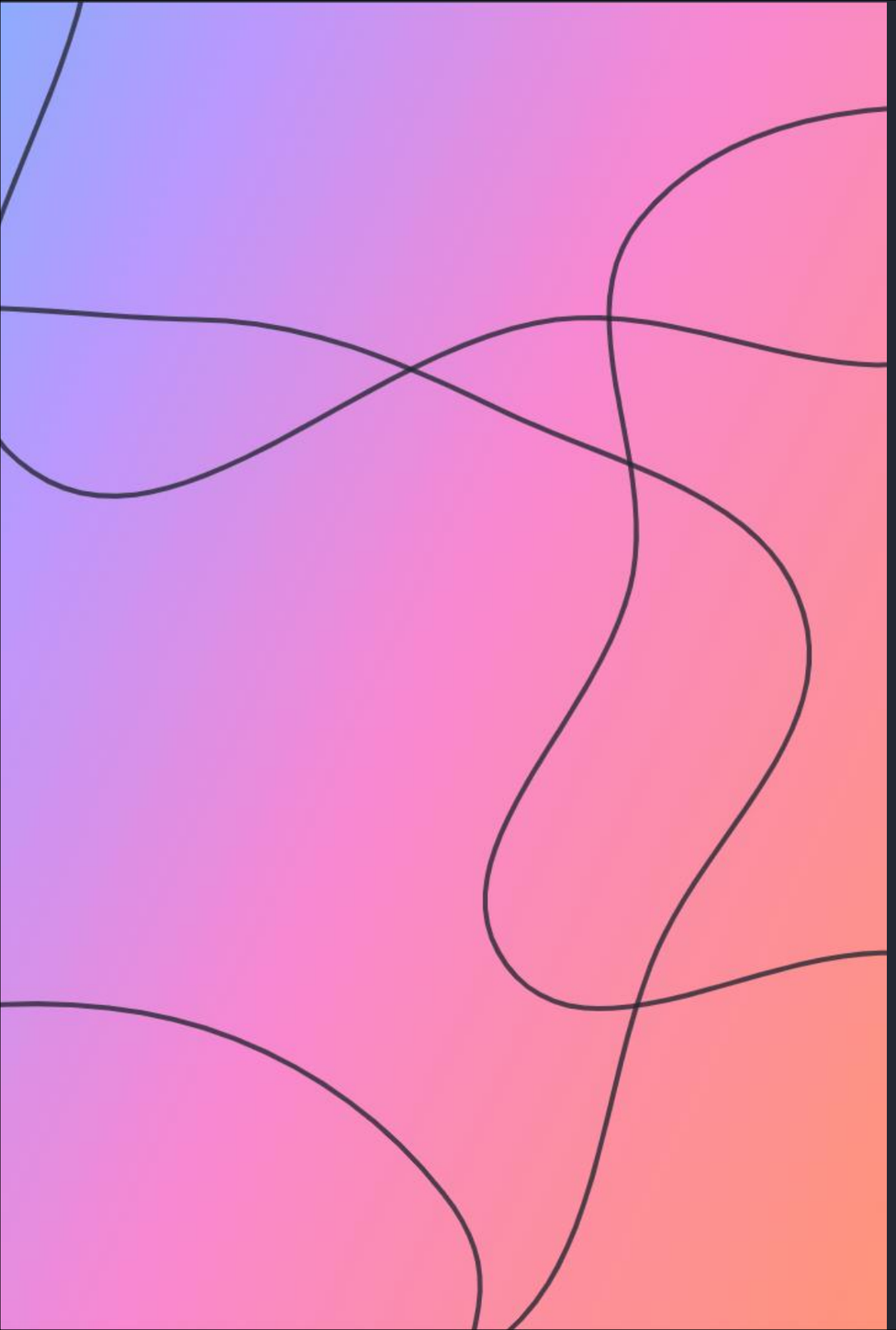
Integration: Embed the model into a live-match analytics platform.

Testing: Conduct real-time testing during matches to refine predictions.

Feedback Loop: Implement a system to learn from prediction errors and improve over time.

Outcome: A ready-to-deploy model offering valuable insights for teams and bettors, with a framework for continuous enhancement.





Thank You!

Thank you for joining us in this IPL Score Prediction journey. We hope you found the insights valuable and informative. If you have any further questions or would like to explore more, feel free to reach out to us.