STARC '23 Abstract

The Ultimate IPL Score Predictor and Probabilistic Analysis Tool

Sport:	Cricket
Leader:	Sanath Kumar R [PES2UG21CS474]
Members:	1. Sufiyan Pasha [PES2UG21CS545] 2. Sujal S [PES2UG21CS548] 3. Sanath Kumar R [PES2UG21CS474]

Project Details

Background

Cricket fans worldwide eagerly follow the Indian Premier League (IPL), a prestigious Twenty20 cricket tournament held annually in India. With its high intensity and star-studded lineups, predicting the score of IPL matches becomes a captivating endeavor.

Accurate score predictions in IPL matches have immense value for teams, analysts, and fans. While traditional methods relied on subjective judgments, the advent of machine learning and data analysis has opened new avenues for developing data-driven models that can leverage historical IPL match data.

By incorporating factors such as the batting team, bowling team, current score, overs played, and other relevant variables, machine learning algorithms can discern patterns and relationships crucial for predicting the final score. This predictive capability can aid teams in strategic decision-making, enable analysts to assess player performance, and enhance the overall engagement and excitement of the IPL fans.

Goals, Objectives and Scope

1. Develop an accurate machine learning model that predicts the score of IPL cricket matches by considering factors such as batting team, bowling team, current score, and overs played.
2. Provide valuable insights and predictions to assist IPL teams in making informed decisions during matches, including setting target scores, adjusting strategies, and evaluating performance.
3. Enhance the overall engagement and enjoyment of IPL fans by offering real-time score predictions, analysis, and a deeper understanding of the game.
4. Collect and preprocess historical IPL match data, ensuring its quality and relevance to the prediction task.
5. Train and fine-tune machine learning algorithms using the prepared dataset to develop a predictive model that accurately forecasts the final score of IPL matches.
6. Create a user-friendly interface or API that allows users to input match details and receive real-time score predictions, facilitating easy access and usability.

Applications:

1. **Team Strategy and Decision-making:** IPL teams can use the score predictions to make informed decisions during matches, such as setting a target score, planning batting order, adjusting bowling strategies, and optimizing fielding positions. This can help teams maximize their chances of winning and improve overall performance.
2. **Player Performance Assessment:** Coaches, analysts, and team management can leverage score predictions to assess player performance in real-time. By comparing predicted

scores with actual outcomes, they can identify the impact of individual players on the team's overall performance and make data-driven decisions regarding player selection, training, and performance improvement.

3. **Fan Engagement and Fantasy Cricket:** Accurate score predictions can enhance the overall fan experience and engagement with the sport. Fans can use the predictions to participate in fantasy cricket leagues, make informed match predictions, and engage in discussions and debates about team strategies and performance.
4. **Broadcasting and Commentary:** Broadcasters and commentators can integrate score predictions into their coverage, providing valuable insights and analysis to viewers. Predicted scores can be used to create interactive graphics, statistical comparisons, and discussions around team strategies, player performances, and match dynamics.
5. **Sports Betting and Bookmaking:** Predicting cricket scores can have implications in the sports betting industry. Accurate score predictions can be utilized by bookmakers to set odds and help bettors make informed betting decisions. However, it's important to ensure responsible gambling practices and comply with relevant legal and ethical considerations.
6. **Cricket Analytics and Research:** The score prediction model can contribute to the field of cricket analytics and research. The predicted scores, combined with other match statistics, can be used to derive valuable insights, identify patterns, and analyze the impact of different factors on match outcomes. This can further fuel research in cricket strategy, player performance, and game dynamics.

Deliverables

Data Collection:

- Identify and select reliable sources for collecting IPL match data, such as official IPL websites, cricket statistics websites, or APIs.
- Develop a systematic approach to gather historical IPL match data, ensuring coverage of multiple seasons and teams.
- Implement data collection scripts or tools to automate the process and ensure regular updates of the dataset.

Data preprocessing:

- Clean the collected data by handling missing values, outliers, and inconsistencies using appropriate techniques.
- Perform exploratory data analysis (EDA) to gain insights into the dataset, identify patterns, and visualize distributions of variables.
- Convert categorical variables into numerical representations through techniques like one-hot encoding or label encoding.
- Split the dataset into training and testing sets, ensuring appropriate data partitioning to evaluate the model accurately.

Feature Engineering:

- Analyze the existing variables in the dataset and identify potential feature engineering opportunities.
- Create new features that capture the dynamics of an ongoing IPL match, such as run rate, required run rate, batting team's recent performance, bowling team's recent performance, and historical head-to-head records.
- Consider interactions or combinations of variables to derive new informative features.
- Evaluate the relevance and importance of the engineered features using techniques like feature importance, correlation analysis, or domain knowledge.

Model Selection and Training:

- Explore and compare different machine learning algorithms suitable for score prediction, such as regression models (linear regression, decision tree regression, random forest regression) or advanced techniques like gradient boosting or neural networks.
- Select the most appropriate model based on performance metrics, computational efficiency, and interpretability.
- Split the preprocessed dataset into training and validation sets for model training and validation.

Model Evaluation and Fine-tuning:

- Evaluate the trained model's performance using appropriate evaluation metrics such as mean squared error (MSE), root mean squared error (RMSE), and R-squared.
- Assess the model's ability to generalize by evaluating its performance on the validation set or through cross-validation techniques.
- Analyze the model's strengths, weaknesses, and areas for improvement based on the evaluation results.
- Fine-tune the model by adjusting hyperparameters, optimizing feature selection, or considering ensemble techniques to enhance its predictive capabilities.

User Interface Development:

- Design and develop a user-friendly interface or API that allows users to input match details easily, such as batting team, bowling team, current score, and overs played.
- Implement a responsive and intuitive interface with appropriate input validation and error-handling mechanisms.
- Integrate the developed score prediction model into the user interface to provide real-time predictions.
- Provide clear and concise instructions on how to use the interface and interpret the predicted scores.

Potential Obstacles

- Limited availability and quality of reliable IPL match data.
- Capturing and incorporating all relevant factors that influence cricket match outcomes.
- Achieving high accuracy in score predictions due to the uncertainty of cricket matches.

- Adapting the model to accommodate the dynamic nature of IPL, including team compositions, rule changes, and match dynamics.
- Ensuring transparency and interpretability of the model to gain user acceptance.
- Addressing legal and ethical considerations, particularly when using the model for sports betting or commercial purposes.

Project Approval

Suggestions

-

Approved by

POSITION

DD/MM/YYYY

Approved by:

POSITION

DD/MM/YYYY

Week - 1

Meetings:

Meeting 1 with Dr. Bharathi R on 5/06/23:

The project progress was shown and input was taken from ma'am for future implementation.

4 Meetings with Team members

Various problems were discussed and solved.

Progress Made:

- We successfully collected data on all IPL matches till 2021 more will be done in later weeks.
- A working Python code was written for our model.
- Many minor changes were added to the code.
- We referred to some research papers from Google scholars.
- We finished some courses on regression models.

Next Scheduled Faculty Review:

Monday 14/06/23 at 10:00 AM (online)

Work to do in the Upcoming Week:

- Adding data of matches played in years 2022 and 2023.
- Improving our ML model.
- Adding some new features to the web interface.
- Finish some other courses on clustering models.

Week - 2

Meetings:

Meeting with Sandesh Sir and Bharthi ma'am on 05/06/2023

Showed the progress so far and they suggested that we add graphs for predicted vs actual scores.

3 Meetings with Team members

Many features of the web interface were discussed and many tests were conducted on the model for checking its efficiency. We distributed tasks among ourselves.

Progress Made:

- Collected data for Gujrat Titans and Lucknow Super Giants teams and added them to the data set.
- Added many pages to the web interface such as About, Info etc
- Added Team logos, when the user selects a team the respective logo will be displayed on the screen.
- Graphs for the projected and actual score was added.
- The current run rate section was added.

Next Scheduled Faculty Review:

Monday 22/06/23 at 10:00 AM (online)

Work to do in the Upcoming Week:

- Second innings have to be added.
- Improving our ML model.
- Adding features to the web interface.
- Weather conditions should be taken into account for the model.