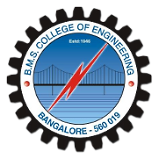
**B.M.S. COLLEGE OF ENGINEERING**

**(Autonomous College under VTU)**

**Bull Temple Road, Basavangudi, Bangalore - 560019**



**MACHINE LEARNING (23DS4PCMLG)**

**Alternative Assessment Report**

***Submitted by***

**Amrit Raj(1BM22AD005)**

**Sai Pramod (1BM22AD049)**

**Snehasish Kabi (1BM22AD057)**

***in partial fulfillment for the award of the degree of***

**BACHELOR OF ENGINEERING**

***in***

**ARTIFICIAL INTELLIGENCE AND DATA SCIENCE**

***Under the Guidance of***

Dr. LAKSHMI SHREE K

Assistant Professor

**Department of AI&DS, B.M.S. College of Engineering**

**2023-2024**

**DECLARATION**

We, hereby declare that the Alternative Assessment Report entitled IPL Score Prediction on Machine Learning (23DS4PCMLG) is a bonafide work and has been carried out by us under the guidance of Faculty, Dr. Lakshmi Shree K, Assistant Professor, Department of Artificial Intelligence and Data Science, B.M.S. College of Engineering, Bengaluru, in partial fulfillment of the requirements of the degree of Bachelor of Engineering in Artificial Intelligence and Data Science of Visvesvaraya Technological University, Belagavi.

I further declare that, to the best of my knowledge and belief, this report has not been submitted either in part or in full to any other university/college.

Candidate details:

|  |  |  |  |
| --- | --- | --- | --- |
| **SL. NO.** | **Student Name** | **USN** | **Student’s Signature** |
| 1 | Amrit Raj | 1BM22AD005 |  |
| 2 | Sai Pramod | 1BM22AD049 |  |
| 3 | Snehasish Kbi | 1BM22AD057 |  |

Place: Bengaluru

Date:

Certified that these candidates are students of Artificial Intelligence and Data Science Department of B.M.S. College of Engineering. They have carried out the **Alternative Assessment Report** entitled “IPL Score Prediction” on **Machine Learning (23DS4PCMLG).** The work is original and duly certify the same.

Faculty In-charge Signature

Date:

**INTRODUCTION**

In recent years, predicting the final score of a cricket match has gained significant interest among data scientists and sports analysts. Accurate predictions can be invaluable for various applications such as sports betting, game strategy development, and enhancing fan engagement. This project aims to predict the final score of a match based on the current score using advanced machine learning techniques.

For our analysis, we have collected data from various seasons of the Indian Premier League (IPL), one of the most popular T20 cricket leagues in the world. The dataset includes detailed information on matches, including current scores, overs, wickets, and other relevant features that can influence the final score. By leveraging this rich dataset, we aim to build a robust model that can provide accurate and reliable predictions of the final scores.

In this report, we will detail the entire process of our project, starting from data collection and preprocessing to model selection, training, and evaluation. We will also discuss the results and potential future improvements to enhance the accuracy of our predictions.

**AIM & OBJECTIVE**

**Aim**

The primary aim of this project is to develop a machine learning model that accurately predicts the final score of an IPL match based on the current score and other relevant match parameters.

**Objectives**

1. **Data Collection and Preprocessing**:
   * Gather historical match data from multiple IPL seasons.
   * Clean and preprocess the data to ensure it is suitable for analysis and modeling.
2. **Exploratory Data Analysis (EDA)**:
   * Perform a detailed analysis of the data to identify key patterns and correlations.
   * Visualize the data to gain insights that can inform the modeling process.
3. **Model Development**:
   * Explore various machine learning algorithms and select the most appropriate ones for the task.
   * Train and optimize the models using the collected data.
4. **Model Evaluation**:
   * Evaluate the performance of the models using relevant metrics such as Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE).
   * Compare the performance of different models to select the best-performing one.
5. **Result Interpretation and Visualization**:
   * Analyze the predictions made by the final model and compare them with actual match outcomes.
   * Create visualizations to effectively communicate the results of the model.
6. **Conclusion and Future Work**:
   * Summarize the findings and highlight the strengths and limitations of the current approach.
   * Suggest potential improvements and future directions for further research.

**PROBLEM STATEMENT**

In the realm of sports analytics, accurately predicting the final score of a cricket match is a challenging task due to the numerous variables involved. Traditional methods often fall short in capturing the complexity and dynamic nature of the game. With the advent of machine learning, there is an opportunity to leverage historical data to create models that can make more precise predictions.

The specific problem addressed in this project is to develop a machine learning model that can predict the final score of an IPL match based on the current score, overs completed, wickets fallen, and other relevant match parameters. This involves:

Collecting and preprocessing a comprehensive dataset from past IPL matches.

Analyzing the data to understand the key factors influencing the final score.

Building and evaluating various machine learning models to identify the best approach for accurate score prediction.

By solving this problem, we aim to provide a tool that can be used by analysts, coaches, and enthusiasts to gain deeper insights into match outcomes and enhance strategic decision-making during the game.

**PROJECT DESCRIPTION**

The project focuses on developing a machine learning model to predict the final score of an IPL match based on the current score and other relevant match parameters. This involves several key steps, each contributing to the overall goal of creating an accurate and reliable prediction model.

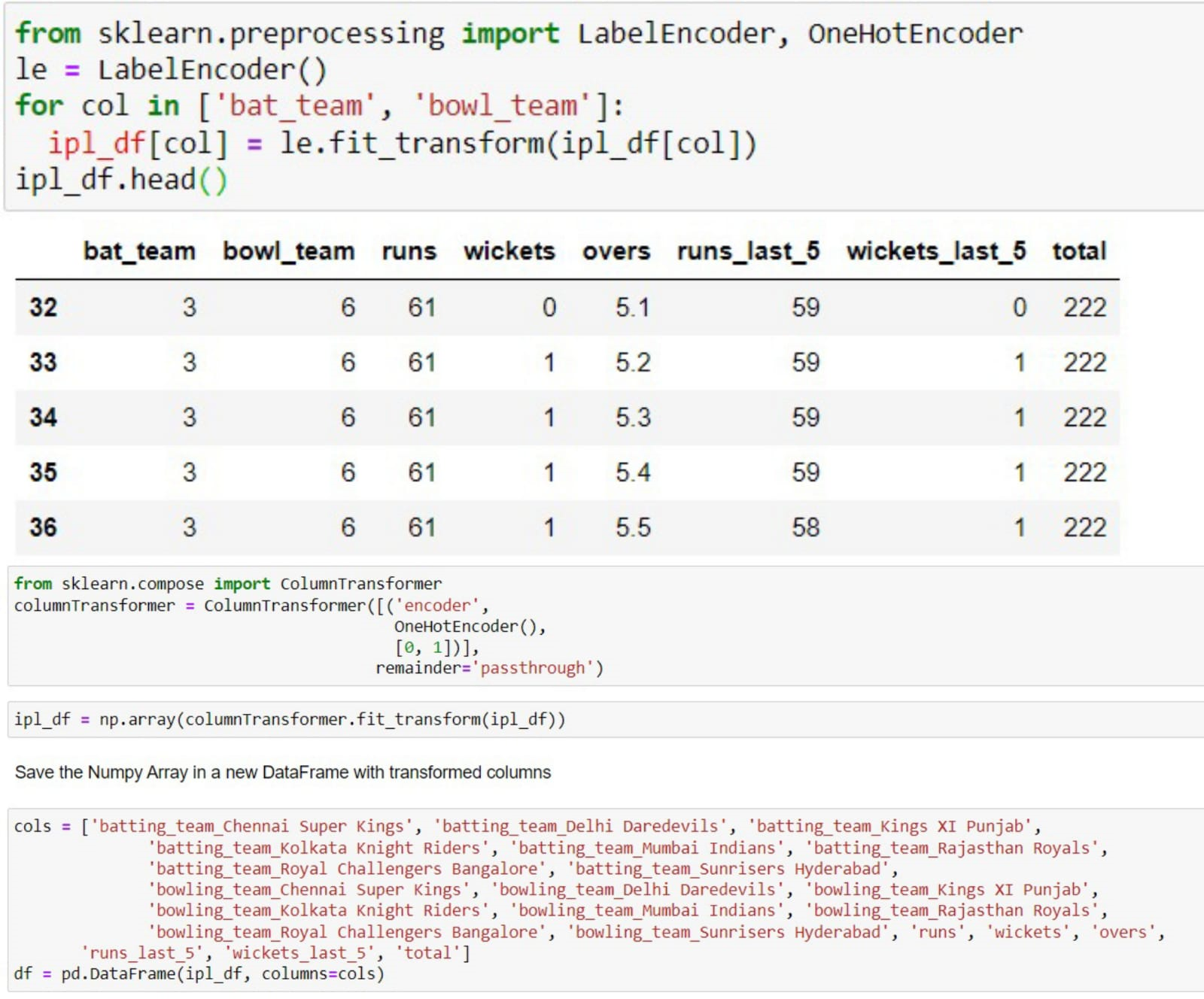
**Data Collection**

Historical match data from multiple IPL seasons is collected and stored in CSV format. The dataset includes various features such as the current score, overs completed, wickets fallen, player statistics, and other relevant match details.



**Data Preprocessing**

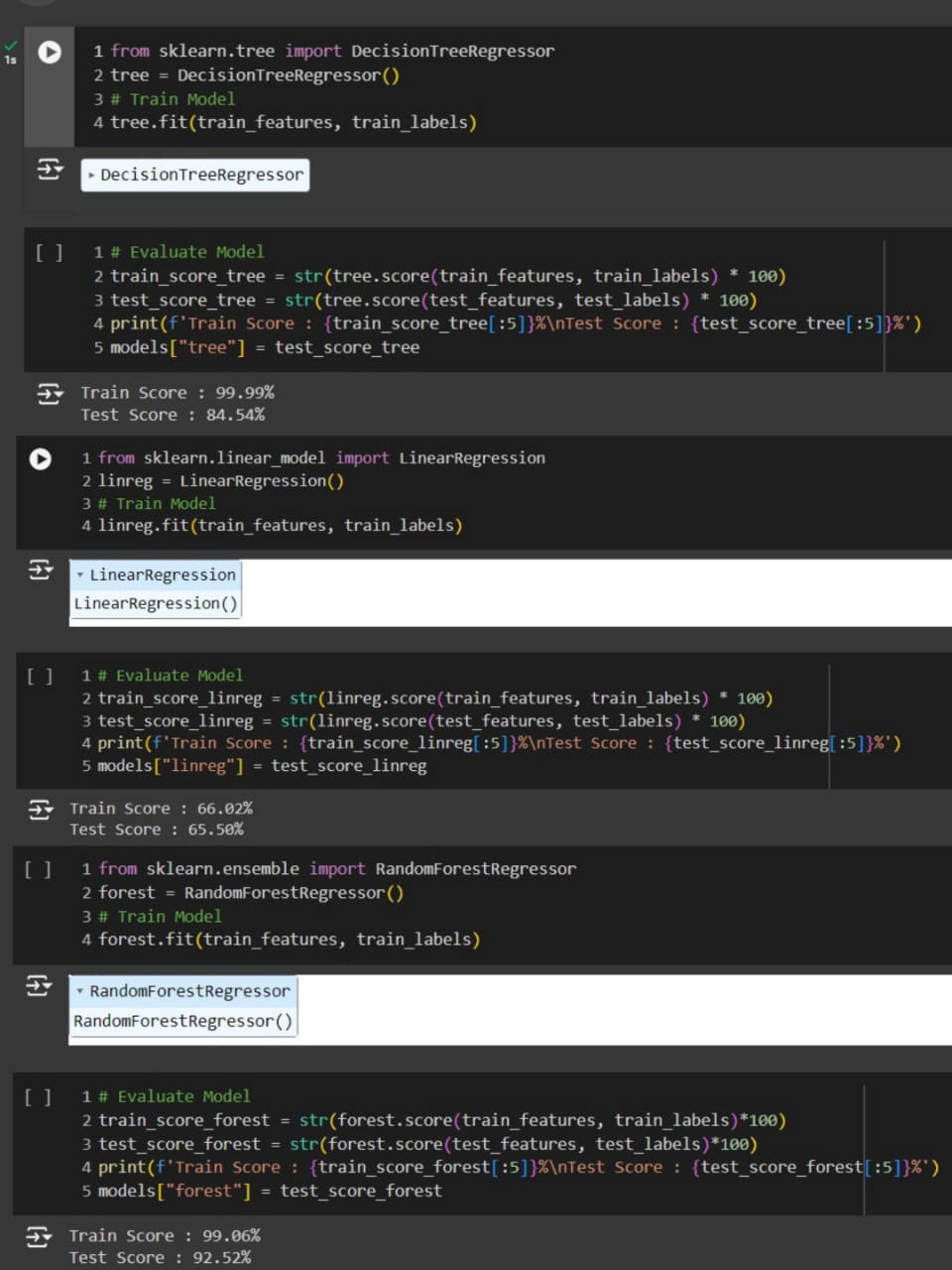
The collected data undergoes thorough preprocessing to ensure its quality and suitability for modeling. This includes handling missing values, encoding categorical variables, normalizing numerical features, and creating new features if necessary.

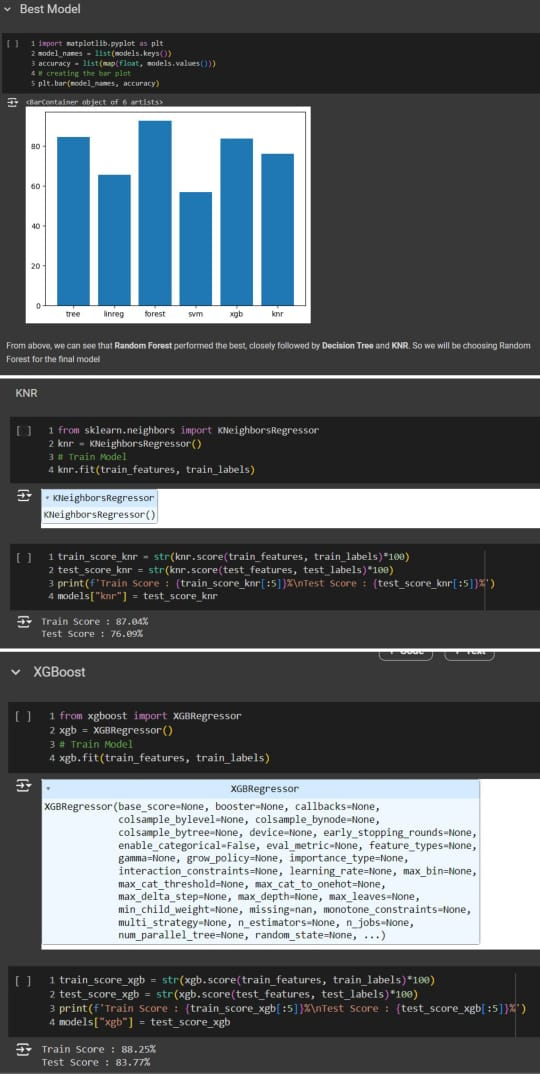
**Exploratory Data Analysis (EDA)**

A detailed analysis of the dataset is conducted to uncover important patterns, trends, and correlations among the features. Visualization techniques, such as scatter plots, histograms, and heatmaps, are used to gain insights into the data.

**Model Development**

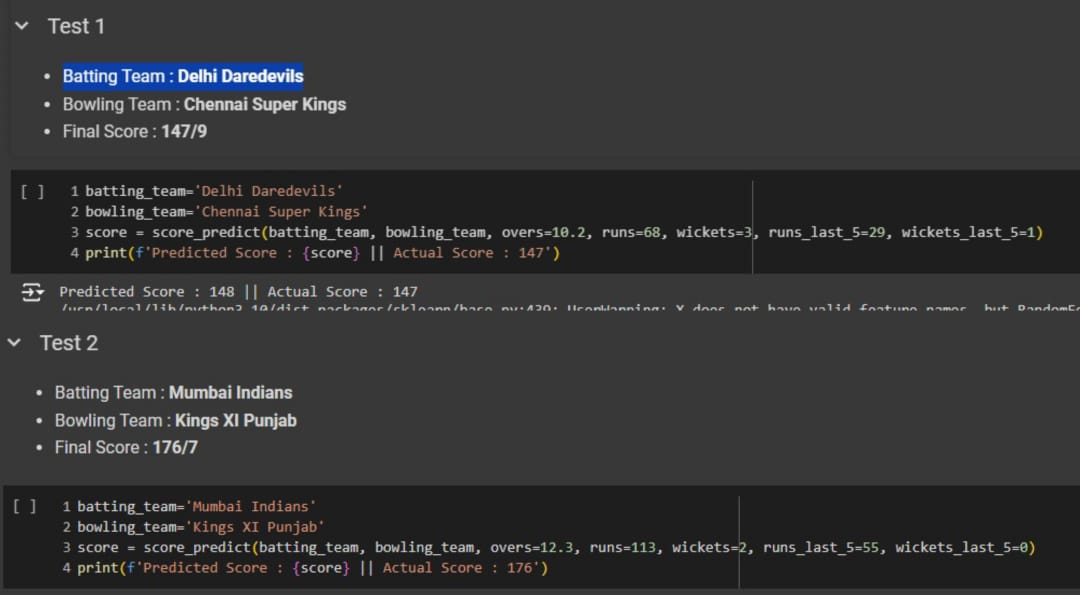
Initially, machine learning models like K-Nearest Neighbors (KNN) and Naive Bayes (NBT) are applied to the preprocessed data. To enhance the accuracy of the predictions, ensemble learning models such as Random Forest and AdaBoost are also employed. Each model is trained and fine-tuned to identify the most effective approach for predicting the final score.





**Model Evaluation**

The performance of each model is evaluated using metrics such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R-squared. Cross-validation techniques are employed to ensure the robustness and generalizability of the models.



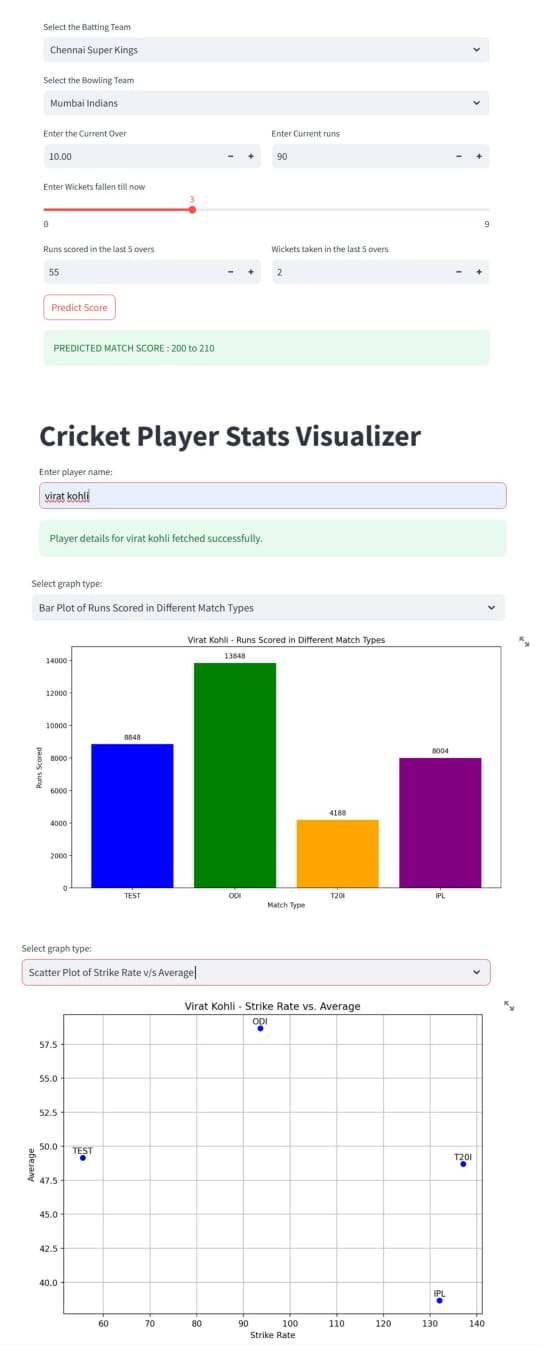
**Results and Visualization**

The predictions made by the final model are compared with the actual match outcomes to assess its accuracy. Visualizations, such as line graphs and scatter plots, are used to effectively communicate the model's performance and predictions.

**Conclusion and Future Work**

The findings of the project are summarized, highlighting the strengths and limitations of the developed model. Suggestions for potential improvements and future research directions are provided to enhance the accuracy and applicability of the prediction model.

**RESULTS**



The image displays a web interface for a cricket match score prediction and player statistics visualization tool. Here's a detailed explanation of the plotting:

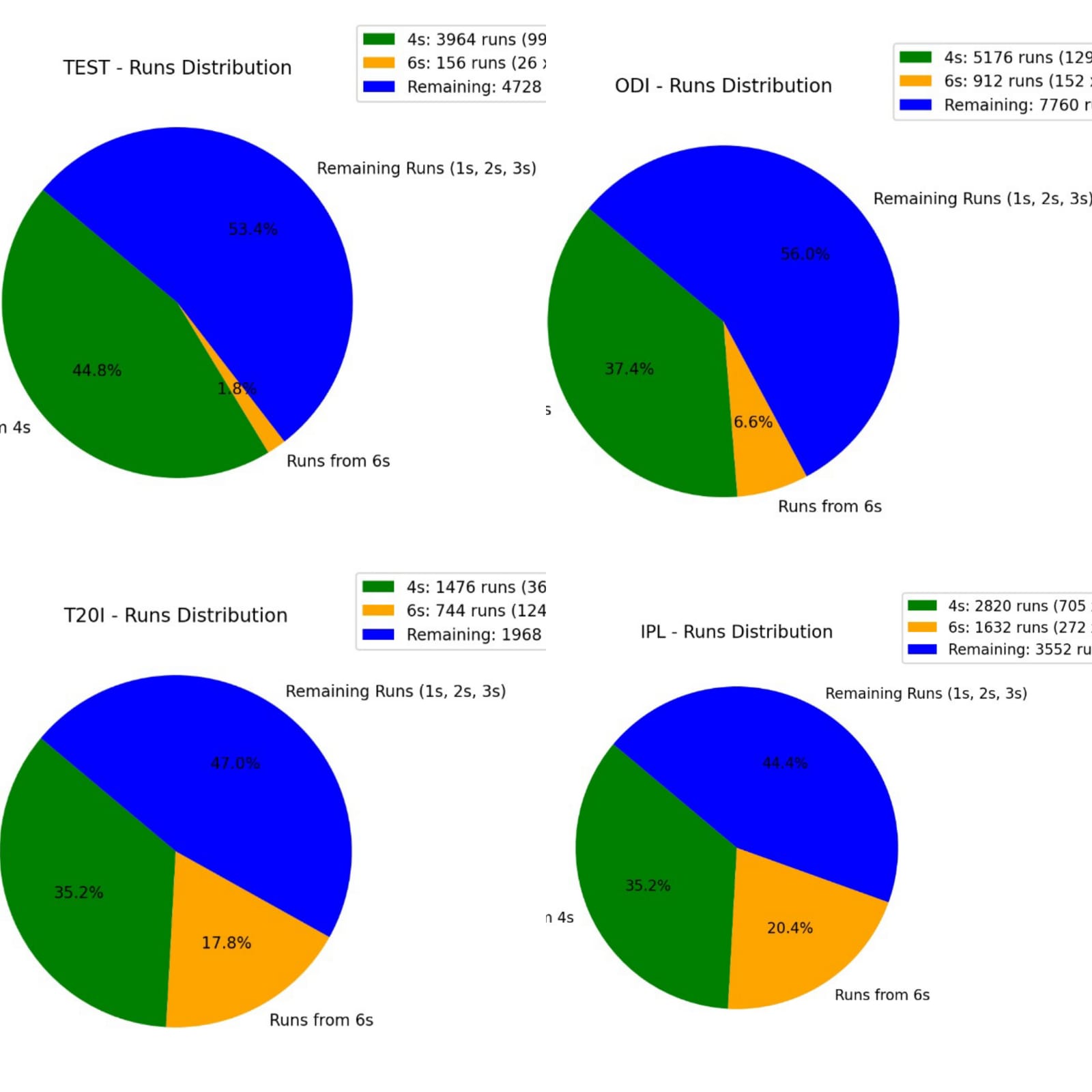
Score Prediction Section:

1. Input Fields:
   * Select the Batting Team: Dropdown to choose the batting team (e.g., Chennai Super Kings).
   * Select the Bowling Team: Dropdown to choose the bowling team (e.g., Mumbai Indians).
   * Enter the Current Over: Number input to specify the current over (e.g., 10).
   * Enter Current Runs: Number input to specify the current runs scored by the batting team (e.g., 90).
   * Enter Wickets Fallen Till Now: Slider to input the number of wickets fallen (e.g., 3).
   * Runs scored in the last 5 overs: Number input for runs scored in the last 5 overs (e.g., 55).
   * Wickets taken in the last 5 overs: Number input for wickets taken in the last 5 overs (e.g., 2).
2. Predict Score Button:
   * When clicked, it predicts the final match score based on the input data. The predicted score range is displayed (e.g., 200 to 210).

Player Statistics Visualizer:

1. Input Field:
   * Enter Player Name: Text input to specify the player’s name (e.g., Virat Kohli). The system confirms the successful fetching of player details.
2. Graph Type Selection:
   * Bar Plot of Runs Scored in Different Match Types: Dropdown to select the type of graph to be displayed.
3. Bar Plot:
   * The bar plot shows the runs scored by Virat Kohli in different match types:
     + TEST: 8548 runs
     + ODI: 13648 runs
     + T20I: 4188 runs
     + IPL: 8004 runs
   * Different colors are used for each match type to distinguish between them.
4. Graph Type Selection:
   * Scatter Plot of Strike Rate vs. Average: Dropdown to select another type of graph to be displayed.
5. Scatter Plot:
   * The scatter plot displays Virat Kohli's strike rate against his average for different match types:
     + Points are plotted for TEST, ODI, T20I, and IPL matches.
     + X-Axis (Strike Rate): Represents the strike rate.
     + Y-Axis (Average): Represents the batting average.
     + Each point represents the relationship between the strike rate and average for each match type.

This interface provides a comprehensive tool for cricket match predictions and visualizing player statistics using various graph types.



The image shows four pie charts representing the distribution of runs scored by Virat Kohli in different match types. Each pie chart breaks down the total runs into three categories: runs from 4s (boundaries), runs from 6s (sixes), and remaining runs (1s, 2s, and 3s). Here is the explanation of each pie chart:

**TEST - Runs Distribution:**

* **4s:** 3964 runs (99 boundaries) - 44.8%
* **6s:** 156 runs (26 sixes) - 1.8%
* **Remaining:** 4728 runs (1s, 2s, 3s) - 53.4%

**ODI - Runs Distribution:**

* **4s:** 5176 runs (129 boundaries) - 37.4%
* **6s:** 912 runs (152 sixes) - 6.6%
* **Remaining:** 7760 runs (1s, 2s, 3s) - 56.0%

**T20I - Runs Distribution:**

* **4s:** 1476 runs (36 boundaries) - 35.2%
* **6s:** 744 runs (124 sixes) - 17.8%
* **Remaining:** 1968 runs (1s, 2s, 3s) - 47.0%

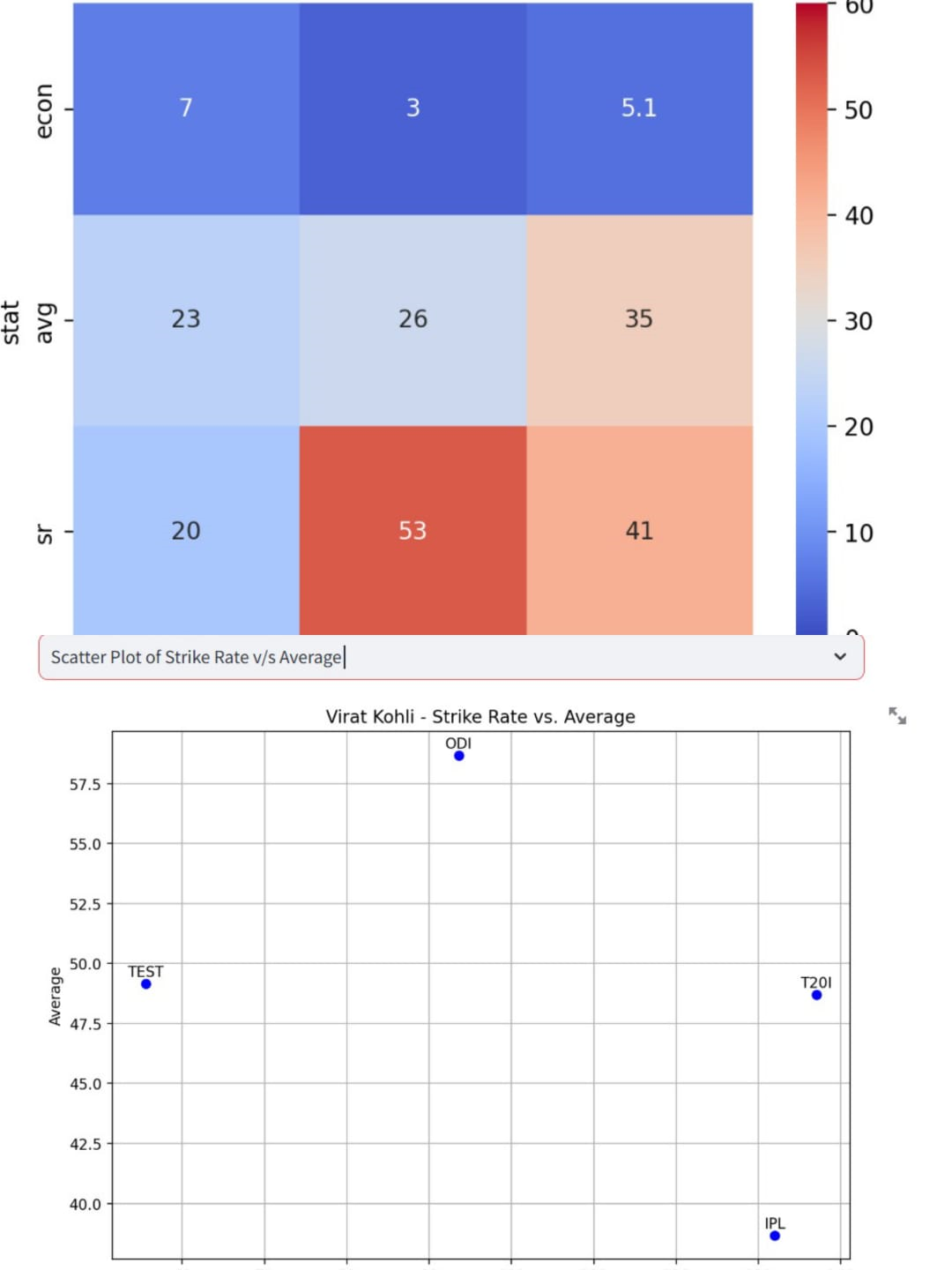
**IPL - Runs Distribution:**

* **4s:** 2820 runs (705 boundaries) - 35.2%
* **6s:** 1632 runs (272 sixes) - 20.4%
* **Remaining:** 3552 runs (1s, 2s, 3s) - 44.4%

**Observations:**

* In Test matches, the majority of runs come from singles, doubles, and triples, with a significant portion from boundaries.
* In ODI matches, over half of the runs come from singles, doubles, and triples, with a substantial contribution from boundaries and a smaller portion from sixes.
* In T20I matches, there is a notable increase in the percentage of runs from sixes, indicating a more aggressive batting approach.
* In IPL matches, there is a balanced distribution with a higher percentage of runs from boundaries and sixes compared to Test and ODI formats, reflecting the fast-paced nature of T20 cricket.

Each pie chart gives a visual representation of how Virat Kohli scores his runs in different formats of cricket, highlighting his adaptability and varied approach across formats.



**Heatmap:**

The heatmap represents a matrix of values with colors indicating the magnitude of those values. The axes and the matrix values are as follows:

* **Y-Axis (Rows):**
  + econ (Economy rate)
  + avg (Average)
  + sr (Strike rate)
* **X-Axis (Columns):** Categories that are not labeled in the provided image but could represent different scenarios or conditions.
* **Matrix Values:** The numerical values in each cell represent the data for the corresponding row and column. The color gradient from blue to red indicates the value's magnitude, with blue representing lower values and red representing higher values.
  + For example, in the econ row, the values are 7, 3, and 5.1, indicating the economy rates under different conditions.
  + In the avg row, the values are 23, 26, and 35, representing the averages.
  + In the sr row, the values are 20, 53, and 41, representing the strike rates.

**Scatter Plot:**

The scatter plot displays Virat Kohli's strike rate against his average for different match types:

* **X-Axis (Strike Rate):** Represents the strike rate.
* **Y-Axis (Average):** Represents the batting average.
* Each point on the scatter plot represents a different match type:
  + **TEST:** Lower strike rate (around 50) and average (around 50).
  + **ODI:** Higher strike rate (around 90) and average (around 57.5).
  + **T20I:** High strike rate (around 135) and slightly lower average (around 50).
  + **IPL:** High strike rate (around 130) and average (around 40).

**Observations:**

* The heatmap provides a visual comparison of economy rates, averages, and strike rates under different conditions, showing how these metrics vary.
* The scatter plot shows the relationship between strike rate and average for different match types, highlighting Virat Kohli's performance across different formats. Higher strike rates are associated with limited-overs formats (ODI, T20I, IPL), while lower strike rates are associated with the Test format, reflecting a more conservative approach.

These visualizations offer a comprehensive view of the player's performance metrics and how they vary across different match types and conditions.