IPL Match Outcome Prediction

By – Prasad Somvanshi

Problem Statement:

The unpredictable nature of cricket matches in the Indian Premier League (IPL) demands accurate prediction models to assist teams, analysts, and enthusiasts in forecasting match outcomes. This project aims to develop a machine learning model capable of predicting the winner of an IPL match based on various features.

Data Loading:

Data Source: The data used for this analysis is sourced from two CSV files, namely 'matches.csv' and 'deliveries.csv', covering the IPL data spanning from 2008 to 2019. These datasets provide comprehensive information about IPL matches, including team details, match outcomes, and ball-by-ball information.

Dataset Description:

matches.csv:

Contains information about each IPL match, including team names, toss details, venue, and winner.

deliveries.csv:

Provides detailed information about each ball bowled in IPL matches, including batting and bowling team, runs scored, and player dismissals.

Model Used:

Machine Learning Model: Logistic Regression

Justification: Logistic Regression is chosen for its simplicity, interpretability, and suitability for binary classification tasks, which aligns with predicting match outcomes (win/lose).

Approach:

Feature Engineering:

1. Total Score Calculation: Aggregated the runs scored in each match to calculate the total score for each team.
2. Team Preprocessing: Standardized team names for consistency.
3. City Encoding: Applied one-hot encoding to the 'city' feature for model compatibility.

Pre-processing Steps:

1. Data Cleaning: Handled missing values and addressed inconsistencies in team names.
2. Data Merging: Combined match information with total scores for both teams.
3. Data Filtering: Selected relevant features for model training.

Model Building:

- Trained Logistic Regression model using historical IPL match data.
- Evaluated model performance using metrics such as accuracy, precision, and recall.

Tools:

Programming Language: Python

Integrated Development Environment (IDE): PyCharm, jupyter notebook

Libraries and Frameworks: Pandas, NumPy, Scikit-learn, Matplotlib, Seaborn

Deployment and Visualization Tools: Streamlit

Outcome:

- The Logistic Regression model demonstrated reliable performance in predicting IPL match outcomes.
- Evaluation metrics, including accuracy and precision, reflect the model's effectiveness in classification tasks.
- The model contributes to informed decision-making for IPL enthusiasts, team management, and analysts.
-

Personal Contribution:

- Led the feature engineering process, ensuring the inclusion of relevant predictors for accurate predictions.
- Conducted data preprocessing steps, including cleaning and merging, to create a cohesive dataset.
- Contributed to the model building phase, evaluating different algorithms before selecting Logistic Regression.
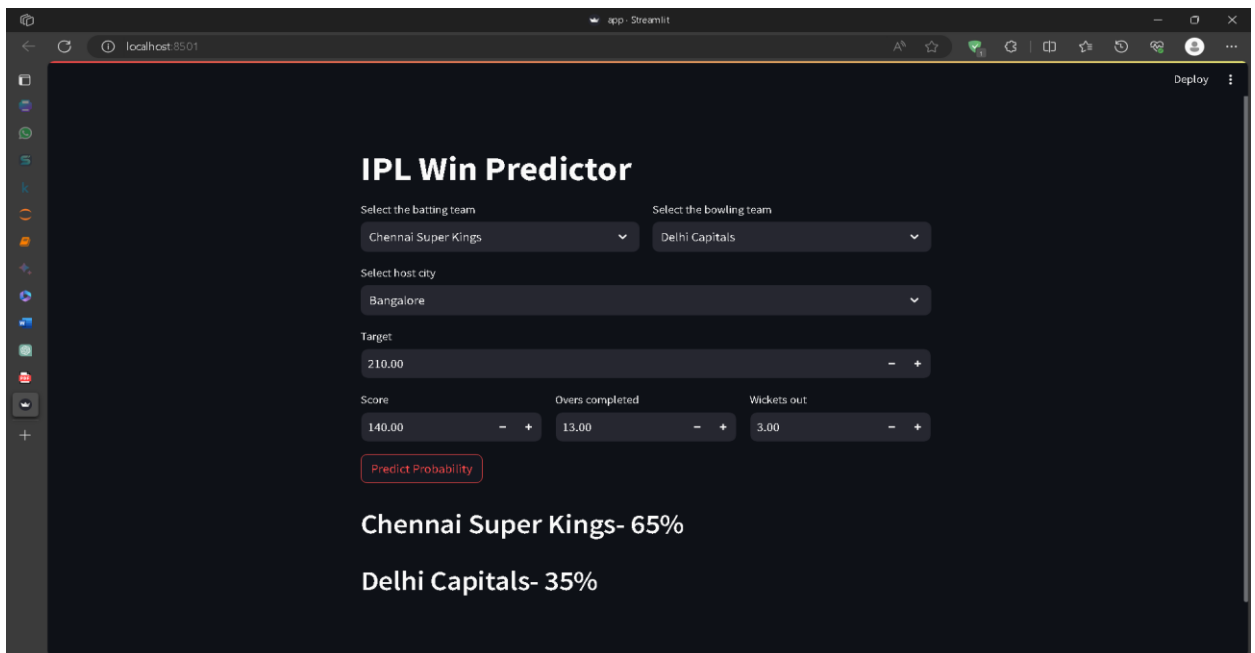-

Project Demo/Visualizations:

- **Local Host Website:**

Utilizing Streamlit, I created a local host website to provide an interactive platform for predicting IPL match outcomes. Users can input relevant details such as batting team, bowling team, host city, target score, current score, overs completed, and wickets out. The underlying model, trained using historical IPL match data, processes these inputs to predict the probability of each team winning.

- **Web Interface:**

The web interface allows users to select teams, city, and input match details, providing them with predictive probabilities for each team's chance of winning. This tool enhances the user experience and aids cricket enthusiasts, team management, and analysts in making informed decisions based on the model's predictions.

Conclusion:

The project successfully addressed the challenge of predicting IPL match outcomes, providing stakeholders with a valuable tool for strategic decision-making. The Logistic Regression model, with its simplicity and interpretability, serves as an effective solution for binary classification tasks. Ongoing efforts focus on refining the model, incorporating real-time data, and exploring additional features for enhanced predictions in future IPL seasons.