Project Report: IPL Match Winner Prediction System

1. Project Title

IPL Match Winner Prediction System Using Machine Learning

2. Objective

The main objective of this project is to predict the winner of an IPL match using historical data and machine learning techniques. The system allows users to input match-related details (teams, toss decision, venue, etc.) and receive a predicted outcome based on trained models.

3. Introduction

The Indian Premier League (IPL) is one of the most-watched cricket leagues in the world. Given the data-rich environment, predicting match outcomes using machine learning provides a real-world application of data science. This project demonstrates how cricket match data can be analyzed and used to train predictive models for forecasting match results.

4. Problem Statement

The aim is to develop a web-based system that can:

- Analyze past IPL match data
- Train a model to learn patterns from this data
- Predict the winning team of a match based on user input
- Display predictions with a confidence score

5. Scope of the Project

- Use IPL historical data from 2008 onwards
- Train a classification model to predict match outcomes
- Create a simple web interface for interaction
- Offer visual insights into IPL match trends

6. Technology Stack

Component	Technology Used
Programming Language	Python
Data Handling	Pandas, NumPy
Visualization	Matplotlib, Seaborn
Machine Learning	Scikit-learn (Logistic Regression)
Frontend	Streamlit (Python-based Web Framework)
Deployment	Streamlit Cloud

7. Dataset Used

- **Source:** Kaggle - IPL Matches Dataset
- **Key Features:**
    - Team 1 & Team 2
    - Toss Winner
    - Toss Decision
    - Venue
    - Match Date
    - Winning Team (target label)

8. Methodology

□ Step 1: Data Collection

- Downloaded IPL match data CSV files from Kaggle.

□ Step 2: Data Preprocessing

- Handled null values and irrelevant columns.
- Encoded categorical variables (e.g., team names, venue).
- Selected relevant features for prediction.

□ Step 3: Model Selection

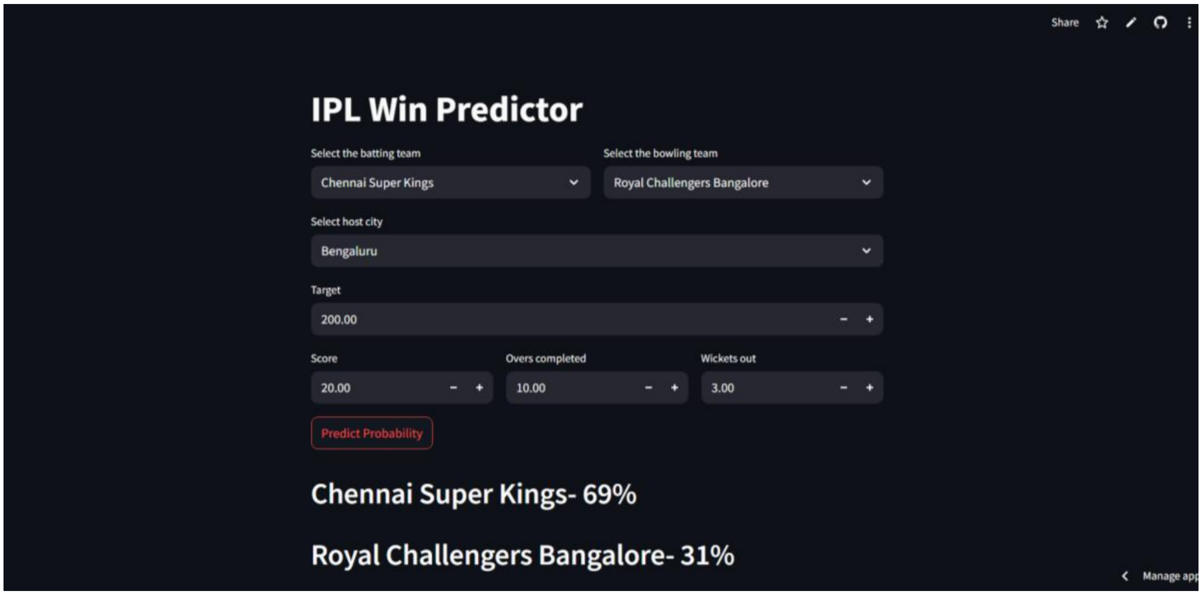
- Tried Logistic Regression.
- Used train-test split to evaluate accuracy.
- Final model chosen based on best performance

□ Step 4: Web Interface with Streamlit

- Created input widgets for selecting team names, venue, and toss.

- Displayed prediction results and confidence levels interactively.

9. Screenshot



IPL Win Predictor

Select the batting team: Chennai Super Kings

Select the bowling team: Royal Challengers Bangalore

Select host city: Bengaluru

Target: 200.00

Score: 20.00

Overs completed: 10.00

Wickets out: 3.00

Chennai Super Kings- 69%

Royal Challengers Bangalore- 31%

10. Evaluation Metrics

- **Accuracy**: 73.6% (based on model and training set)
- **Confusion Matrix**: Displayed to show true vs. predicted values
- **Cross-Validation**: Performed to ensure model stability

11. Results

The system successfully predicts match winners with high accuracy using historical match data. Users can simulate real match conditions through the interface and receive quick predictions along with confidence scores.

12. Challenges Faced

- Imbalanced dataset due to frequent wins by dominant teams (e.g., MI, CSK)
- Encoding of categorical data required manual mapping
- Determining features with actual predictive power (e.g., toss impact)
- Overfitting with certain models (Decision Trees)

13. Future Enhancements

- □ Include player-level stats (e.g., top batsmen, bowlers)
- □ Integrate live data feeds and weather info
- □ Use ensemble models like XGBoost or LightGBM
- □ Deploy as a mobile/web app using Flask or React + API
- □ Integrate with Cricbuzz or ESPN APIs for live predictions

14. Conclusion

This IPL Match Winner Prediction System successfully demonstrates how historical sports data can be leveraged using machine learning for outcome forecasting. The combination of accurate predictions, user interactivity, and scalability makes it a strong base for future sports analytics applications.

15. References

- Kaggle IPL Dataset: https://www.kaggle.com/datasets
- Scikit-learn Documentation: https://scikit-learn.org/
- Streamlit Documentation: https://docs.streamlit.io/
- Python Official Docs: https://docs.python.org/

Live Link:https://ipl-prediction-app-jnqcdr3th4ywbg7mnj6ums.streamlit.app/

Github Repository Link: rahul15-manch/ipl-prediction-app