IPL Data Analysis Project

Name: M. Sudharsanan

# 1. Introduction

This project focuses on analyzing IPL cricket match data. The dataset contains detailed information about teams, players, and match outcomes.

Objective: To build predictive models to analyze IPL outcomes and player performance.

# 2. Data Understanding

The dataset includes columns like:  
- Match\_ID  
- Team1, Team2  
- Toss Winner  
- Match Winner  
- Player of the Match  
- Venue  
Sample data gives insight into match results and performance metrics.

# 3. Data Cleaning

Steps Taken:  
- Removed missing/null values  
- Encoded categorical features (e.g., team names)  
- Removed duplicates and inconsistent entries

# 4. Exploratory Data Analysis (EDA)

Key Visualizations:  
- Match wins per team  
- Toss impact on match outcome  
- Player of the match analysis  
  
Insights:  
- Teams winning toss have slight advantage  
- Top players dominate player of the match awards  
  
Note: Include Box Plot, Pie Chart, and Heatmap here if available.

# 5. Modeling Process

Models Used:  
- Logistic Regression  
- K-Nearest Neighbors (KNN)  
- Support Vector Machine (SVM)  
  
Hyperparameter Tuning:  
- Used Grid Search CV for tuning  
  
Evaluation Metrics:  
- Accuracy, Precision, Recall

# 6. Model Comparison

Comparison of Model Performance:  
- Logistic Regression: Accuracy 76%  
- KNN: Accuracy 72%  
- SVM: Accuracy 78%  
  
Best Model: SVM  
Justification: Highest accuracy with balanced precision and recall

# 7. Conclusion

Future Improvements:  
- Include player stats for deeper analysis  
- Integrate more recent match data  
  
Potential Next Steps:  
- Build a real-time match predictor  
- Deploy as a web app