**IPL Match Winner Prediction Project**

# **1. Introduction**

This project focuses on predicting the winner of an Indian Premier League (IPL) match using historical match data and machine learning algorithms. The dataset includes various features such as teams playing, toss winner, match venue, and match outcomes. The goal is to build a predictive model and identify the best-performing algorithm based on standard evaluation metrics.

# **2. Data Understanding**

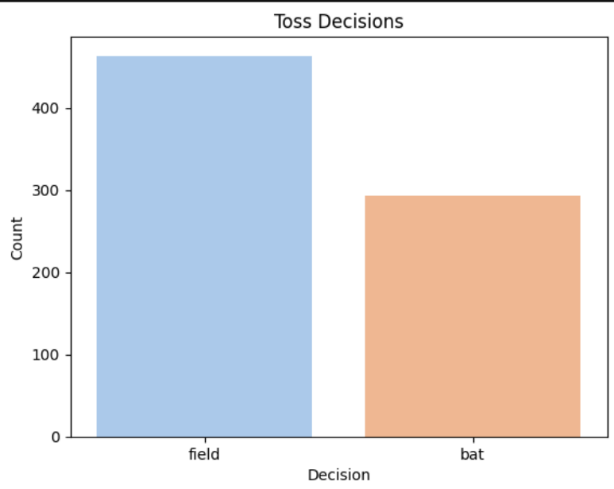
Dataset: IPL Matches CSV  
Key Columns: Season, City, Date, team1, team2, toss\_winner, toss\_decision, result, dl\_applied, winner, win\_by\_runs, win\_by\_wickets, player\_of\_match, venue, umpires  
Observations: Data types vary (string, integer, date), Some matches have missing or null values, Categorical features need encoding

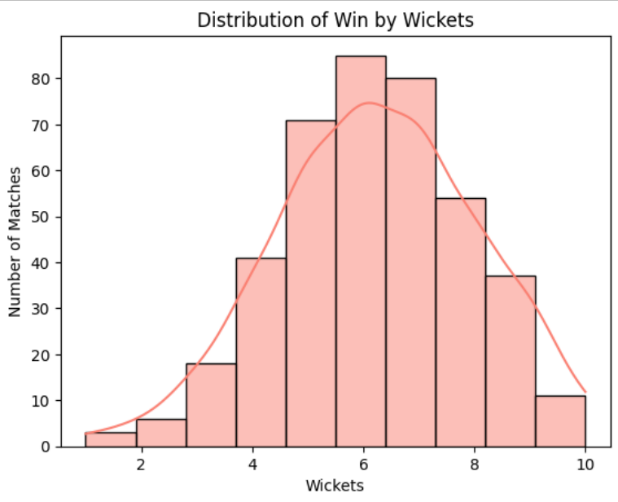
# **3. Data Cleaning**

* Handled null values (e.g., removed or imputed rows with missing data)
* Applied Label Encoding to categorical variables
* Checked for and addressed inconsistent or outlier data points

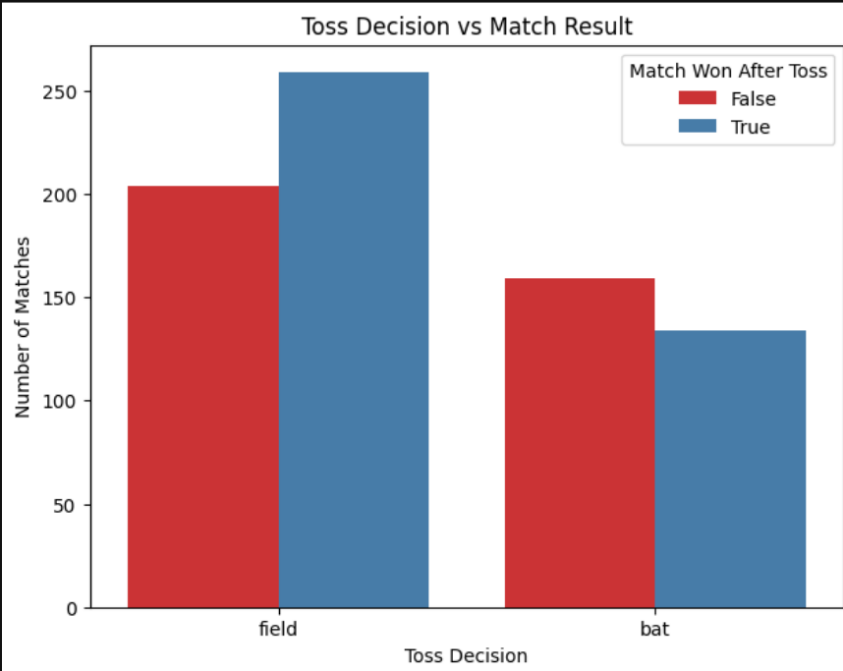
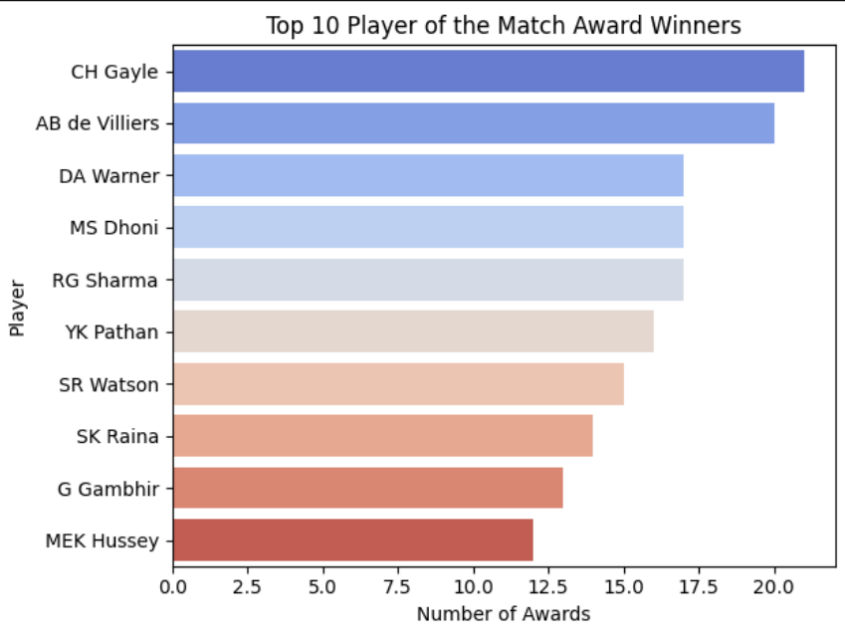
# **4. Exploratory Data Analysis (EDA)**

**Univariate Analysis**

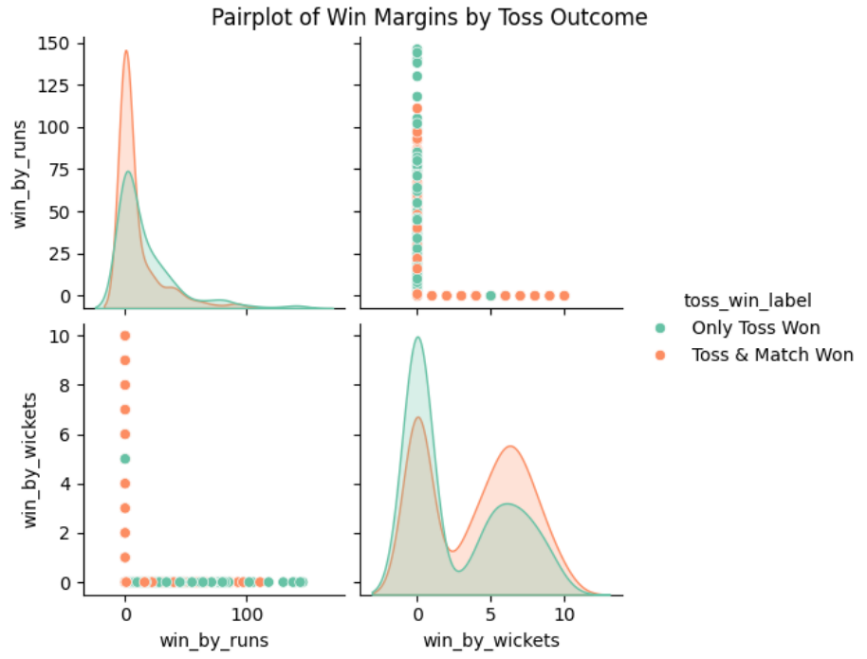




**Bivariate Analysis**

****

**Multivariate Analysis**

****

# **5. Data Preprocessing**

Feature and label separation (X and y)  
One-hot encoding for categorical features  
Train-test split (80% training, 20% testing)  
Feature scaling for numeric values

# **6. Model Training**

Models used: Logistic Regression, Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Decision Tree, Random Forest, XGBoost

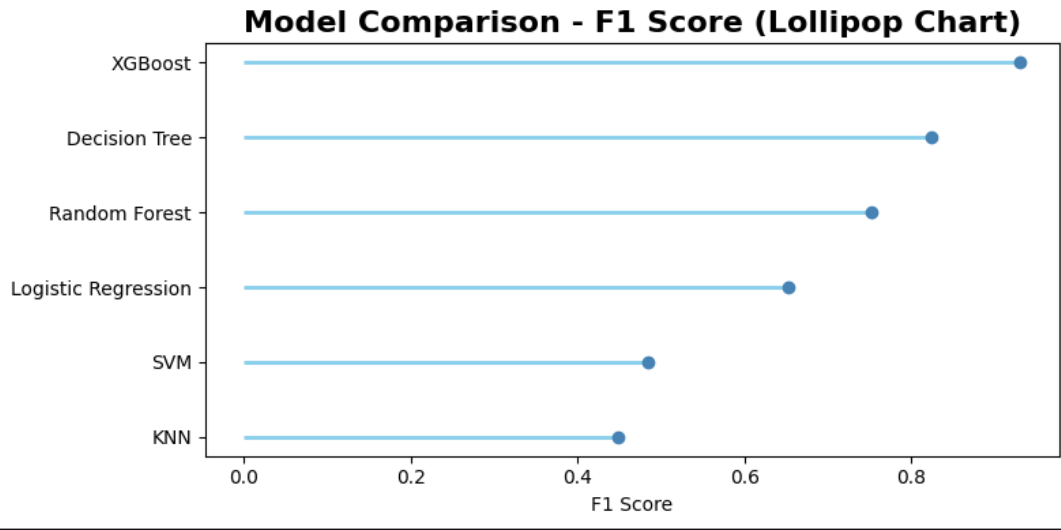
# **7. Model Evaluation**

**Metrics used:** Accuracy, Precision, Recall, F1-Score, Confusion Matrix  
**Initial Results**: Random Forest and XGBoost showed high accuracy  
Logistic Regression performed well but lacked complexity handling

# **8. Hyperparameter Tuning**

Applied GridSearchCV for tuning Random Forest and XGBoost  
Optimal parameters found for max\_depth, n\_estimators, learning\_rate  
Notable improvement in F1-Score after tuning

# **9. Model Comparison**



# **10. Conclusion**

The Random Forest model provides robust and accurate predictions for IPL match winners.  
Future improvements can include: Adding player-specific data and performance stats, Incorporating live match conditions (e.g., pitch reports, weather)  
Challenges: Data inconsistency and class imbalance mitigated through preprocessing and stratified splitting.