21MA602

COMPUTATIONAL LINEAR ALGEBRA

FINAL PRESENTATION

ENGLISH PREMIER LEAGUE FOOTBALL HOME TEAM MATCH PREDICTION

07 December 2023

MUHAMMED SHIBIL C V

CB.SC.P2DSC23007

INTRODUCTION

- Using a data-driven methodology, we step into the interesting world of ***football match predictions***.
- Goal: Utilize machine learning to foresee outcomes – ***victories, losses, or draws.***
- Navigate through football data, historical match stats, and team performance metrics.
- Uncover strategies, challenges, and successes in leveraging AI for predictions

PROBLEM STATEMENT

- ***Why Important?*** Accurate predictions contribute to strategic decision-making for teams, fantasy league players, and betting enthusiasts, enhancing the overall football experience.
- The market for ***sports betting is valued at $500 billion*** (Sydney Herald).
- The ***most popular sport in the world***, football is played by 250 million players in more than 200 countries.
- The most ***well-liked domestic teams*** in the world are the ***English Premier League.***
- Design a ***predictive model*** capable of accurately predicting if the ***home team will win a football match.***

OBJECTIVES

- Develop robust machine learning models leveraging historical football match data.
- Utilize various algorithms to identify patterns and trends that contribute to match outcomes.
- Extract valuable insights from the models to understand the factors influencing match results.
- Provide actionable information for football enthusiasts, analysts, and decision-makers.

METHODOLOGY

- ***Data Collection:***

Gathered extensive football match data, including team statistics, match outcomes, and historical performance.

- ***Data Preprocessing:***

Cleaned and organized the dataset to address missing values, outliers, and inconsistencies.

- ***Exploratory Data Analysis (EDA):***

Conducted in-depth EDA to understand the distribution of variables and identify potential correlations.

- ***Model Development:***

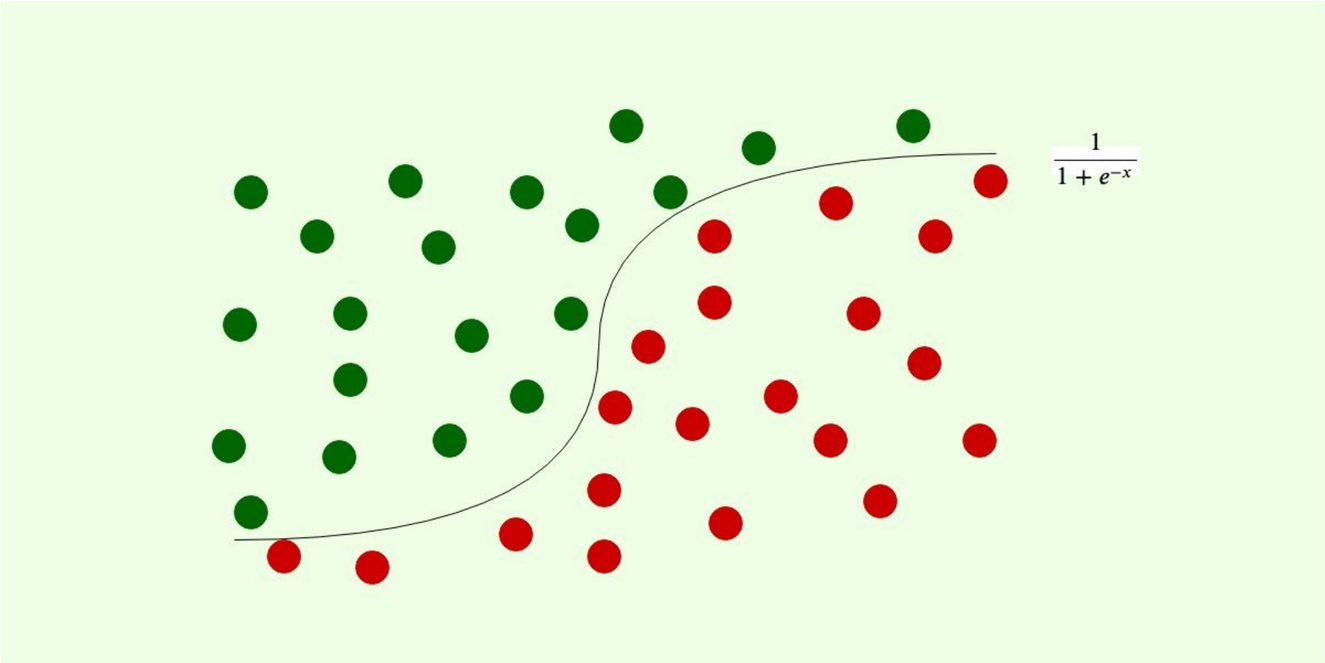
Employed multiple machine learning algorithms, including Logistic Regression, Support Vector Machines (SVM), and Random Forest.

Model Development:

1. Logistic Regression:
2. Support Vector Machines (SVM)
3. Random Forest

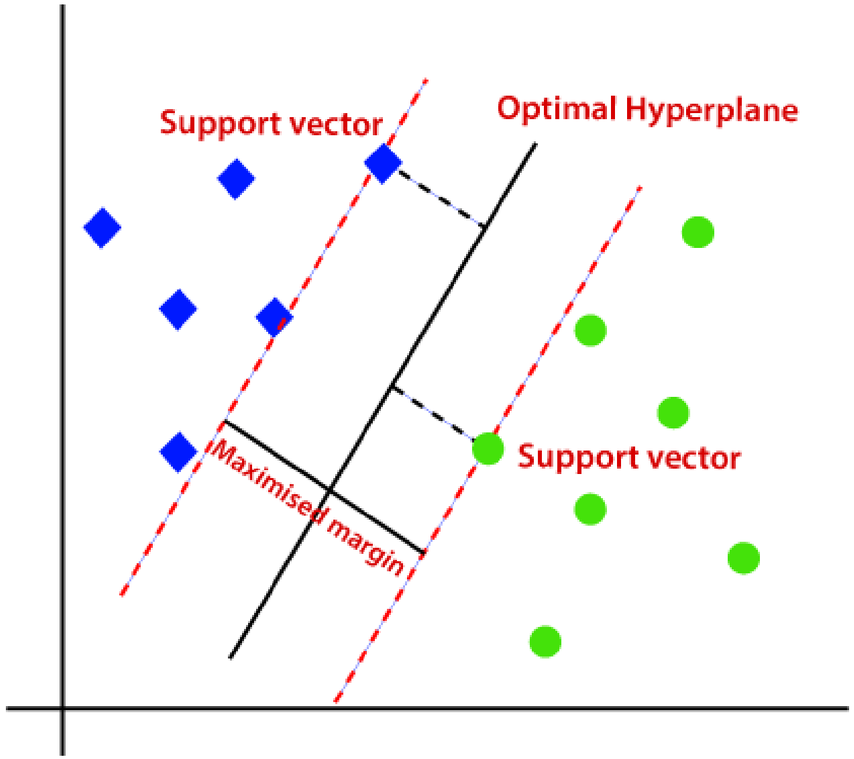
1. **Logistic Regression:**

What it does: ***Predicts the probability of a binary outcome*** (like win or lose) based on input features



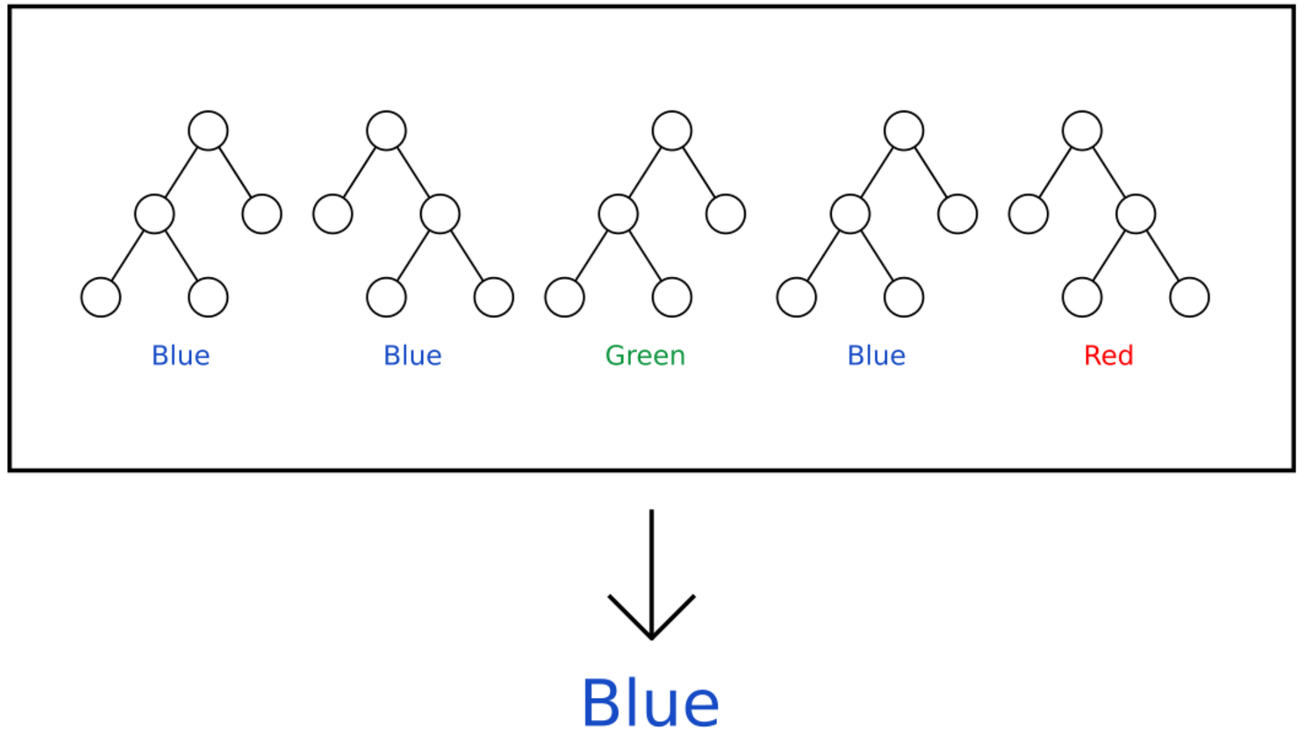
2. Support Vector Machines (SVM):

What it does: Finds the ***optimal decision boundary*** between different classes, maximizing the margin.



3. Random Forest:

- What it does: Constructs ***multiple decision trees*** and ***combines their predictions*** for more robust results.

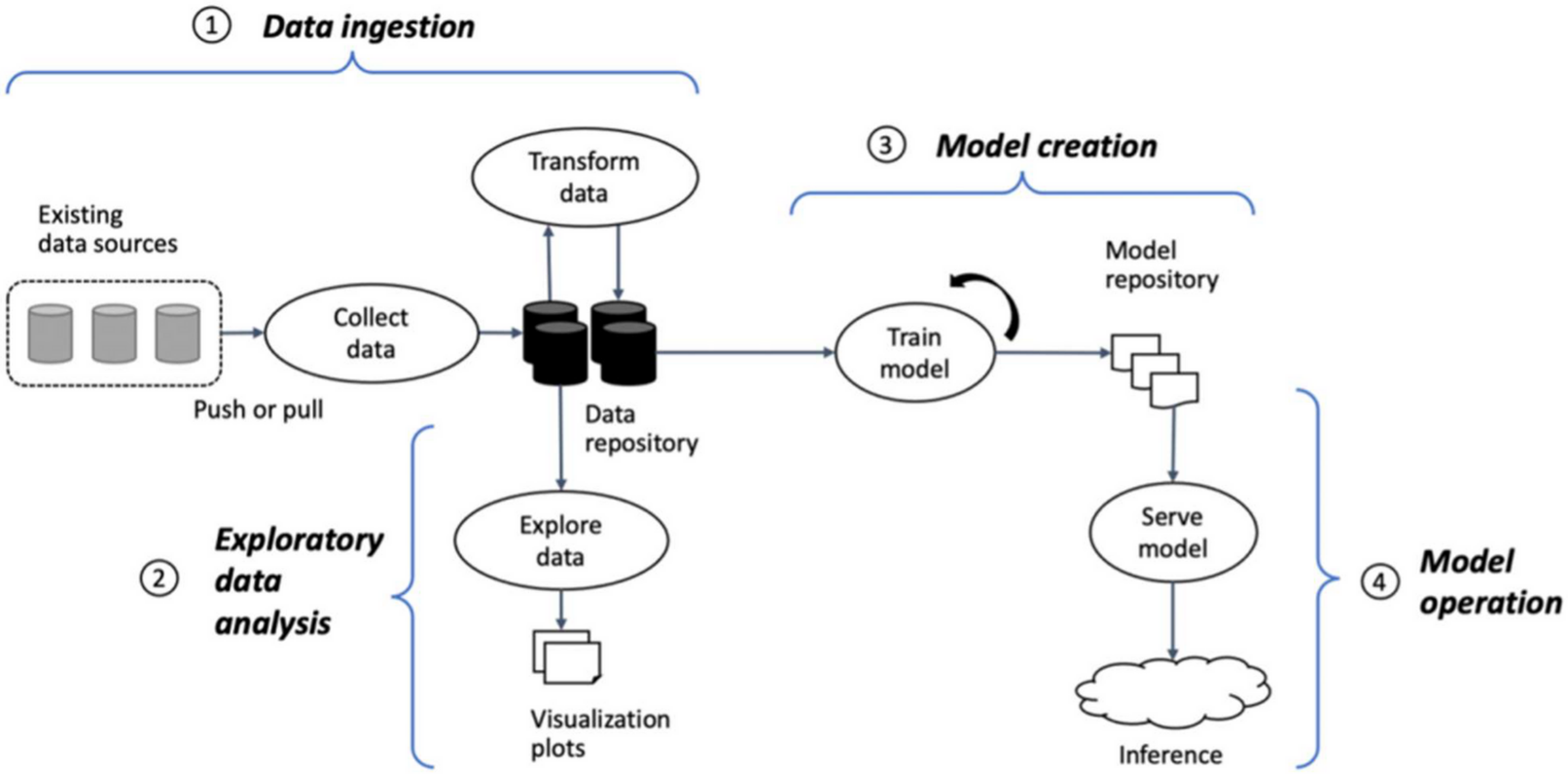


DATASET DESCRIPTION

Dataset has the ***past 18 years*** of EPL dataset of each match and It also has final dataset which produced after processing all year datasets.

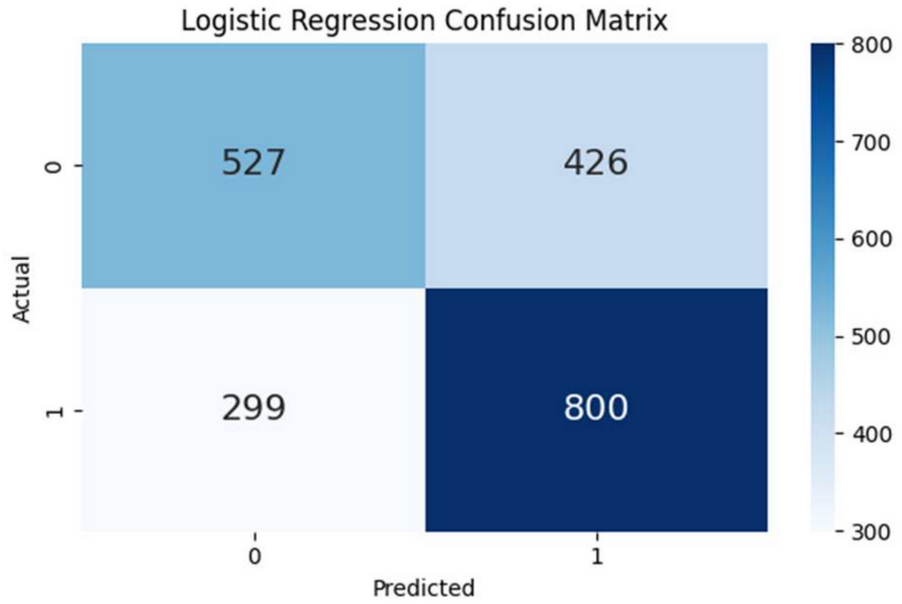
- Div = League Division
- Date = Match Date (dd/mm/yy)
- Time = Time of match kick-off
- HomeTeam = Home Team
- Away team = Away Team
- FTHG and HG = Full Time Home Team Goals
- FTAG and AG = Full-Time Away Team Goals
- FTR and Res = Full-Time Result (H=Home Win, D=Draw, A=Away Win)
- HTHG = Half Time Home Team Goals
- HTAG = Half Time Away Team Goals
- HTR = Half Time Result (H=Home Win, D=Draw, A=Away Win)

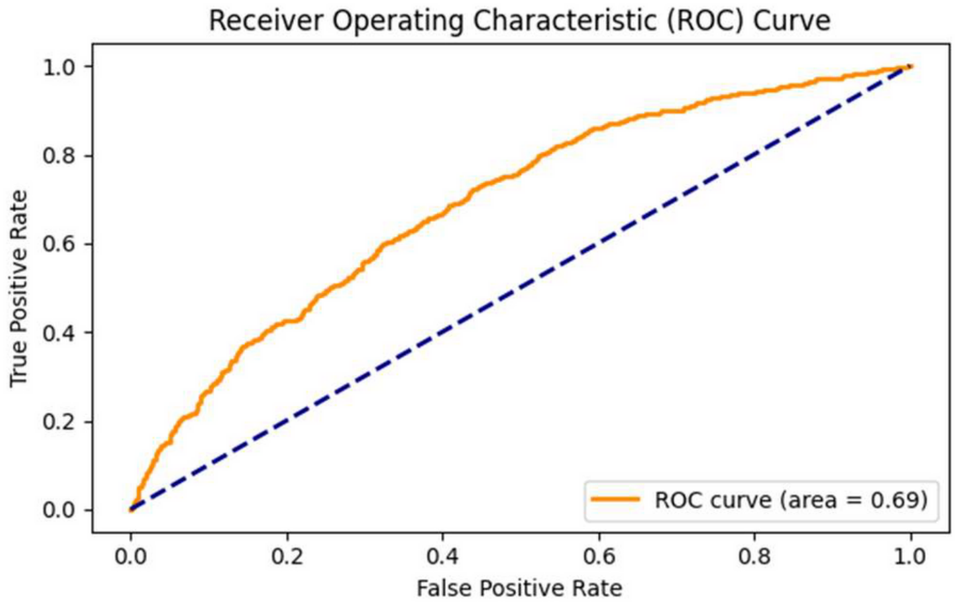
EXPERIMENTAL DESIGN



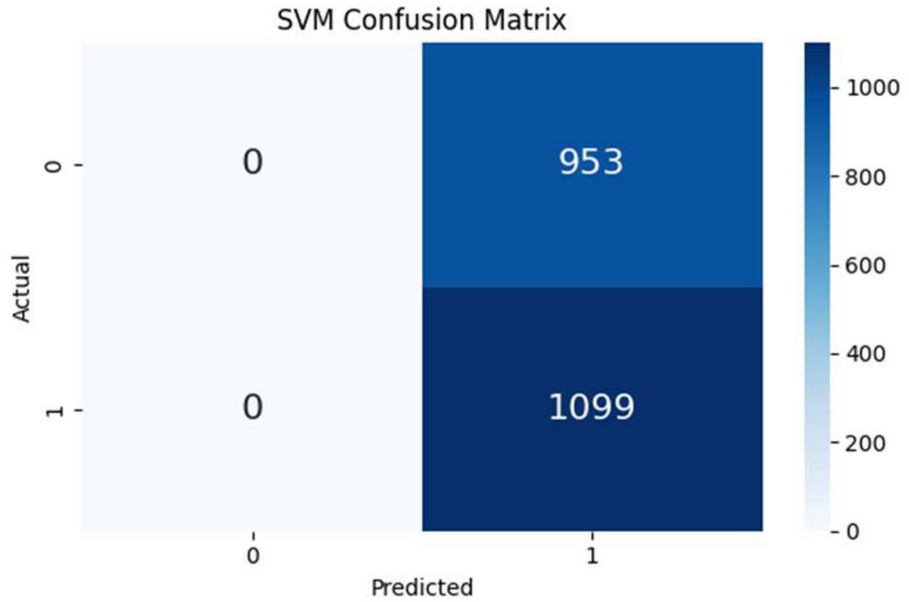
PERFORMANCE MEASURES

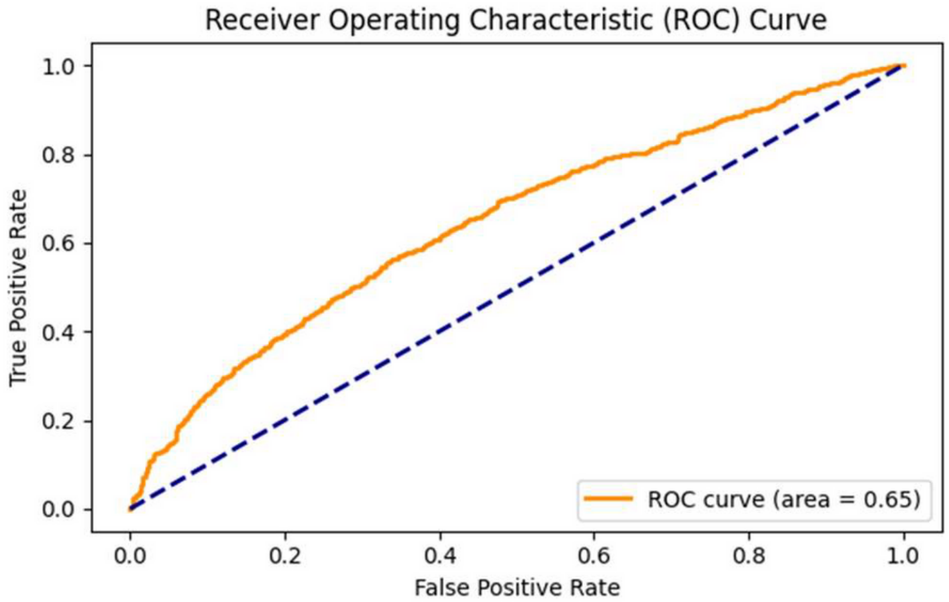
1. Logistic Regression



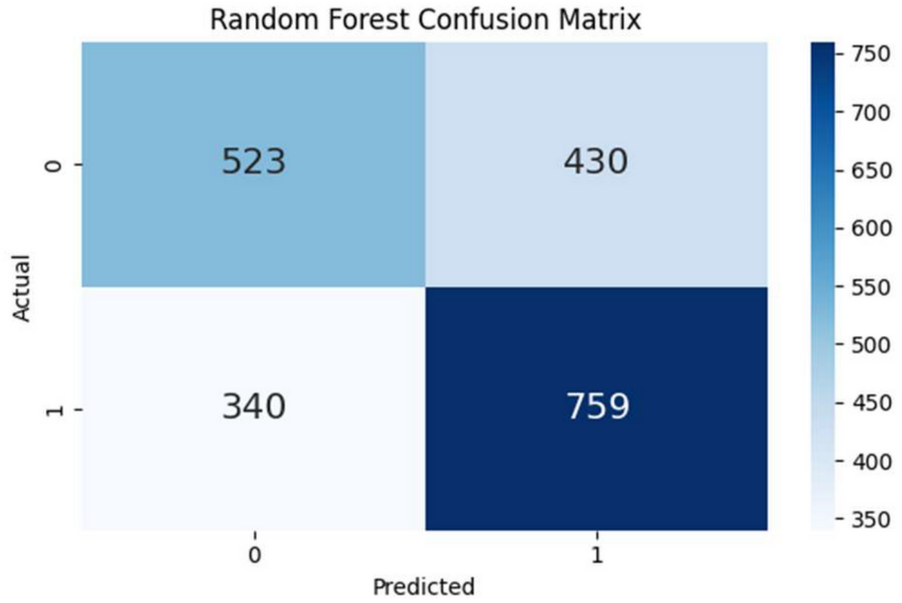


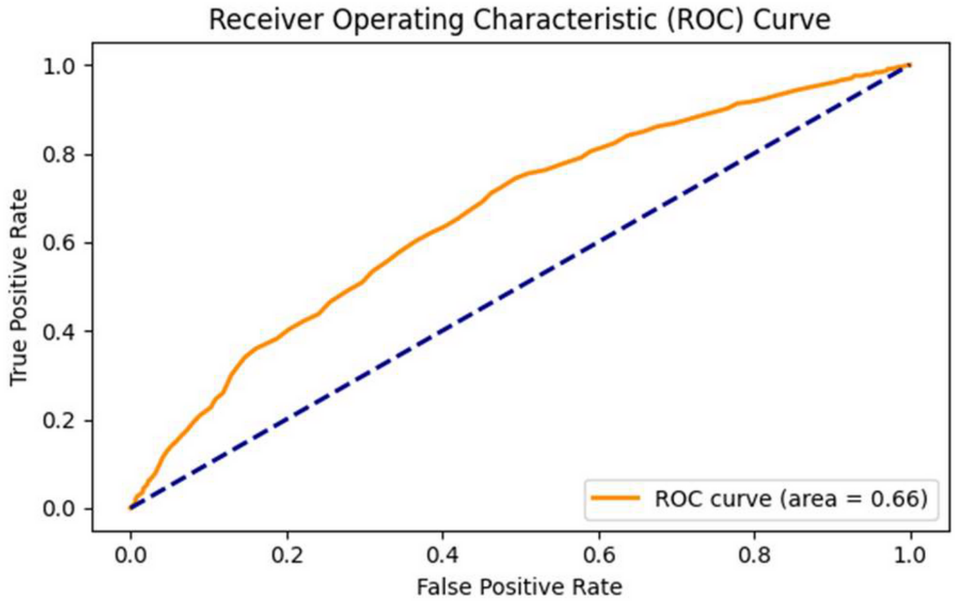
2. Support Vector Machines (SVM):





3. Random Forest:





CONCLUSION

1. ***Model Performance:***
    - ***Logistic Regression*** achieved an accuracy of ***64.67%.***
    - Random Forest demonstrated an accuracy of 62.48%.
    - ***SVM*** showed an accuracy of ***53.56%***.
2. ***Best Performing Model: Logistic Regression*** outperformed other models, demonstrating the highest accuracy in predicting football match results.
3. ***Implications:*** The project's success in employing machine learning for match outcome prediction holds potential applications in sports analytics and betting strategies.
4. ***Future Directions:*** Further enhancements could involve feature engineering, hyperparameter tuning, and exploring advanced models for improved accuracy.