

Predict the Overall Skill of Top 5% Football Player Using Football Video Game Data and Machine Learning Techniques

THIS RESEARCH USES DATA FROM FIFA 15 TO FIFA 23 AND MACHINE LEARNING TECHNIQUES TO PREDICT OVERALL FOOTBALL PLAYER SKILLS, FOCUSING ON IDENTIFYING THE BEST MODEL FOR PREDICTING THE TOP 5% OF PLAYERS. THE RESULTS SHOW THAT GRADIENT BOOSTING CONSISTENTLY OUTPERFORMS OTHER MODELS ACROSS ALL PLAYER POSITIONS, PROVIDING VALUABLE INSIGHTS INTO FOOTBALL PLAYER ASSESSMENT THROUGH DATA-DRIVEN APPROACHES.



INTRODUCTION & METHODOLOGY

INTRODUCTION

The FIFA video game series by EA Sports has established itself as a cornerstone for football enthusiasts, providing an immersive and realistic simulation of the sport. This study delves into player data spanning from FIFA 15 to FIFA 23, encompassing a broad range of over 100 attributes for each player. These attributes include personal information, such as age, nationality, and physical characteristics, as well as performance-related metrics, such as skill ratings, positional data, and statistical achievements.

OBJECTIVE

The goal is to predict the overall skill of football players using machine learning techniques

METHODOLOGY

- Data Extraction: FIFA 15-23 and FC24 dataset, >10,000 players, 110 attributes.
- Pre-processing: Handling null values, grouping data based on player positions.
- Data Mining: Analysis of player rating distributions.
- Modeling: Application of Random Forest, Gradient Boosting, and Linear Regression.
- Model Evaluation: Using RMSE, MAE, R², MAPE, and accuracy metrics.

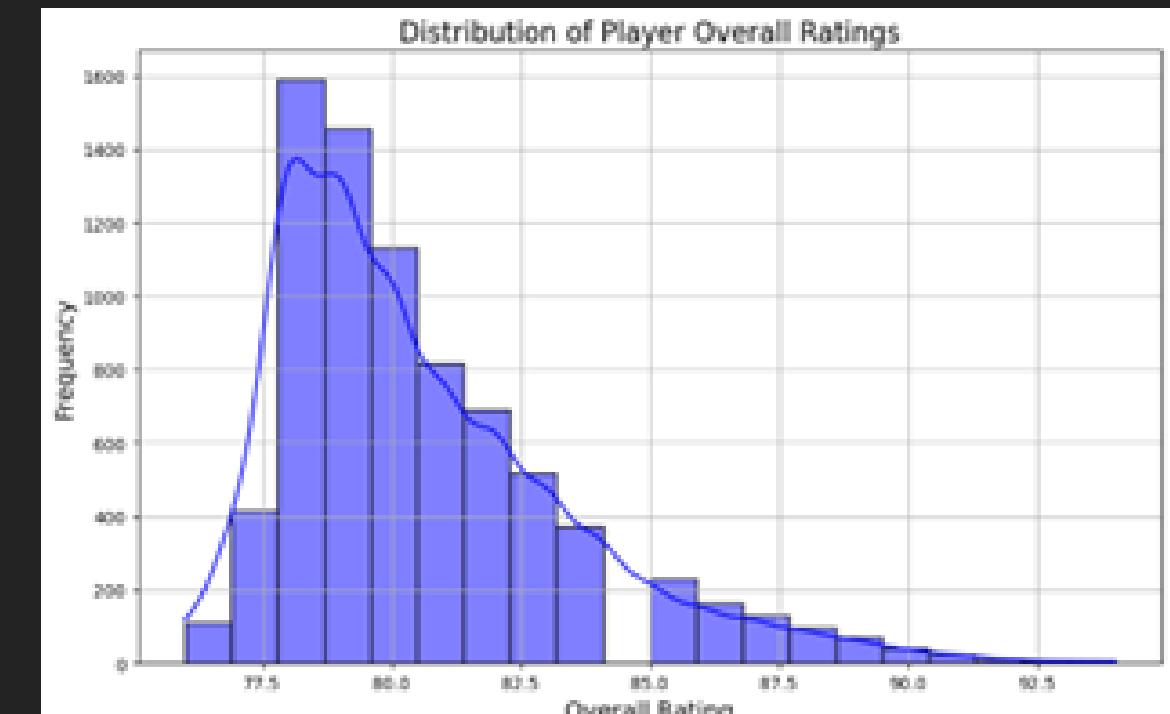
EDA

PREPROCESSING

RESULT & CONCLUSION

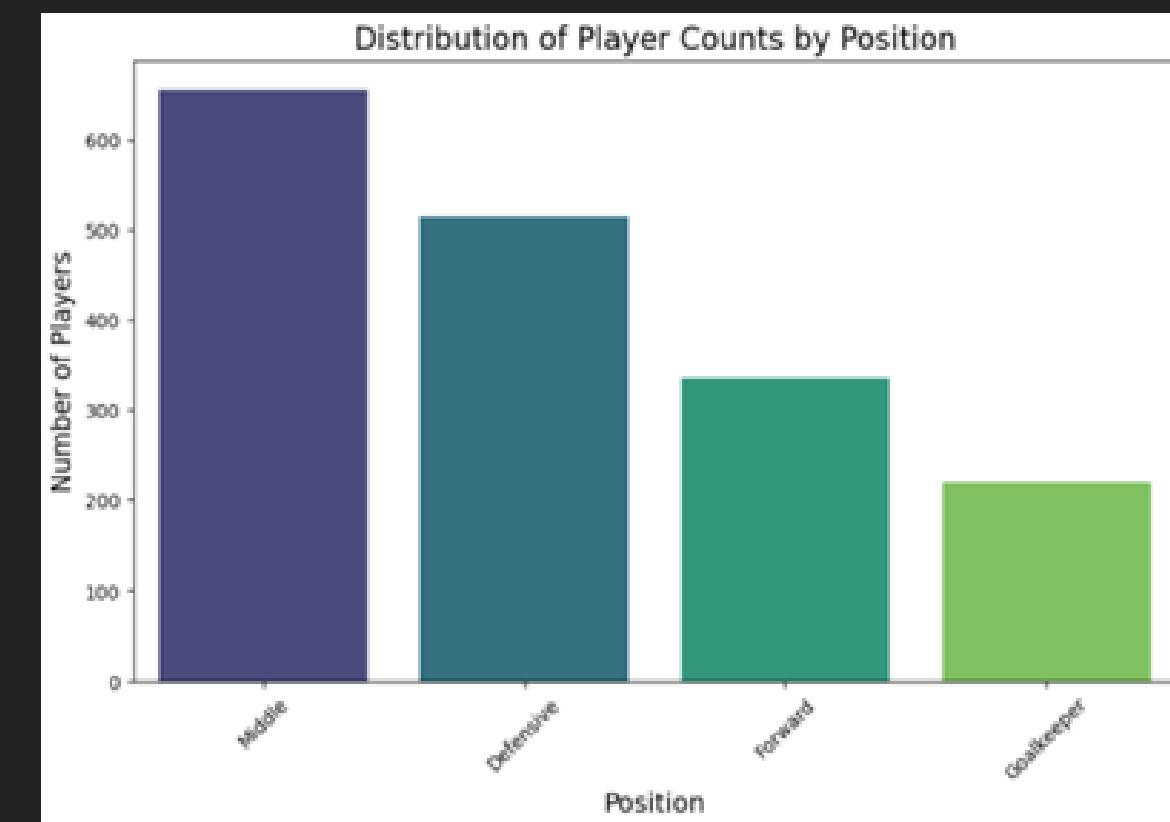
5% OVERALL DISTRIBUTION

The visualizations provide a clear overview of player "overall" ratings distribution, showing a right-skewed histogram with a peak between 77 and 80. Understanding these distributions helps in assessing model performance.



5% POSITION DISTRIBUTION

From a total of 1,726 players, we categorized them into 4 main position groups: Middle (655), Defensive (515), Forward (336), and Goalkeeper (220). This study focuses on the top 5% of players from each category, equivalent to about 86 players. This grouping simplifies various specific positions like ST, CB, CM, etc. into broader categories for more effective analysis.



ANALYSIS FROM EACH EVALUATION METRICS

- RMSE: Gradient Boosting (GB) consistently outperforms other models across all positions, e.g., GK (0.75), Defensive (0.87), Midfield (1.03), Forward (0.89).
- MAE: GB shows the lowest errors across positions, e.g., GK (0.58), Defensive (0.68), Midfield (0.78), Forward (0.63).
- R²: GB explains the most variance, e.g., GK (0.93), Defensive (0.80), Midfield (0.84), Forward (0.91).
- MAPE: GB achieves the lowest percentage errors, e.g., GK (0.73%), Defensive (0.86%), Midfield (0.96%), Forward (0.77%).
- Accuracy: GB shows the highest accuracy, e.g., GK (99.06%), Defensive (98.91%), Midfield (98.72%), Forward (98.89%).

CONCLUSION

The Gradient Boosting model is the most effective for predicting football player skills in FIFA games from 2015 to 2023, outperforming Random Forest and Linear Regression in all evaluation metrics and player positions.

