

Group Members:

Name: Vedant Bhawnani

SAP ID: 60018210069

Name: Shubh Harde

SAP ID: 60018220135

Football Player Replacement Finder

Background and Context

- **Player recruitment challenges:** Traditional methods are subjective, time-consuming, and rely heavily on manual scouting and video reviews.
- **Need for data-driven decisions:** Football clubs increasingly require automated, objective tools to enhance player evaluation and replacement processes.
- **Key factors in player valuation:**
 - **On-field performance metrics:** Goals, assists, passes, defensive actions.
 - **Off-field metrics:** Injury records, market trends, player popularity, and social media presence.
- **Modern football's data-driven shift:** Clubs are adopting advanced analytics and machine learning for recruitment, tactical decisions, and performance analysis (e.g., Expected Goals, passing networks).
- **Technological advancements:** Platforms like Opta, FBRef, and WyScout provide detailed real-time player data for in-depth analysis.

Examples of Advanced On Field Metrics

- Expected Goals (xG)
- Expected Assists (xA)
- Progressive Passes
- Key Passes
- Progressive Carries
- Expected Goals on Target (xGOT)
- Tackles + Interceptions
- Post-Shot Expected Goals (PSxG)
- PSxG - Goals Allowed

Introduction

- **Objective:**

Revolutionize football recruitment by automating player replacement decisions using **data analytics** and **machine learning**.

- **Key Features:**

Process large amounts of **player data** (on-field metrics, off-field factors like injuries and marketability).
Provide **precise, data-driven recommendations** for player replacements.

- **Technological Integration:**

Includes **predictive models**, **player similarity index**, and **real-time data extraction** from platforms like **FBRef**.

- **Impact:**

Helps clubs make **informed decisions** that align with both **tactical needs** and **financial constraints**.

Enhances **decision-making** in maintaining competitive performance and optimizing transfer spending.

Scope/Boundaries of the Project

- **Data Collection:**

Focus on publicly available data from sources like **FBRef** and **Transfermarkt**.
No integration of private club databases or manual data input.

- **Player Metrics:**

Evaluation of **on-field performance** (goals, assists, passing accuracy) and **off-field factors** (injury history, marketability).

- **Player Similarity Index:**

Measurement of how closely potential replacements match current players' attributes.
Focus on technical performance attributes only.

- **User Interface:**

Development of a user-friendly interface for analysts, scouts, and coaches.
No complex tactical planning tools or in-depth match analysis beyond replacement suggestions.

- **Financial Considerations:**

Integration of player price prediction models for estimating transfer costs.
No handling of direct financial transactions or integration with transfer systems.

- **Integration with Existing Club Systems:**

Standalone capabilities for player scouting and replacement.

Sr. no.	Paper	Key findings	Gap identification
1	Predict the Value of Football Players Using FIFA Video Game Data and Machine Learning Techniques	<p>Makes use of 4 models (Linear regression, multiple linear regression, regression trees and random forests)</p> <p>Random forests performs the best</p> <p>Dataset used: https://www.kaggle.com/datasets/stefanoleone992/fifa-20-complete-player-dataset </p>	<p>Uses data from FIFA game, which is subjective to fifa employs and not statistically backed</p> <p>Considers only 9 parameters.</p> <p>Future work stated is not useful to our project.</p>
2	Estimating transfer fees of professional footballers using advanced performance metrics and machine learning	<p>Uses data from sofifa.com and transfermarkt.com</p> <p>Makes use of linear regression and xgbTree</p> <p>xgbTree performs better than linear regression</p> <p>Work answers the question: "Given past performance what is the expected fee of the player", future work: "Given the performance of a player after a transfer, was the price paid reasonable?"</p> <p>Event data from: Instat and GIM performance ratings from: Tarak Kharat.</p>	<p>Uses only ML-based models and optimisations to predict values.</p>
3	Football Player Transfer Value Prediction Using Advanced Statistics and FIFA 22 Data	<p>Uses linear regression models, Gradient boosting and extreme gradient boosting work the best.</p> <p>Predicts prices by dividing the dataset into 4, goalkeepers, defenders, midfielders and forwards.</p>	<p>Uses data from FIFA 22 and Transfermarkt.</p> <p>Gaps: Only uses hard statistics to predict price, completely ignores stats other than contract length</p>

Sr. no	Paper	Reviews	Gap Identification
4	A Machine Learning Ensembling Approach to Predicting Transfer Values	Uses Athletic Performance Data, Financial Indicators, Player Popularity Metrics, Player demographics, Injury data and team and competition data Uses LightGBM with no transformation, root transformation and log transformation	High Value Transfers Lack of Buying Club Info Limited scope of injury and popularity metrics
5	Predicting the Football Players' Market Value Using Neural Network Model: A Data-Driven Approach	Uses Neural Networks and tests different neural networks by changing hyperparameters and activation functions,	Does not use in game statistics Uses data from Fifa 19 only.
6	Econometric Approach to Assessing the Transfer Fees and Values of Professional Football Players	Uses: Contractual Variables, Player Characteristics, Basic Performance Metrics, Club and Market Variables, National Team Participation and Contextual Factors, Uses Multiple Linear Regression	<ul style="list-style-type: none"> Does not handle soft factors such as player popularity Does not go into the depths of economic conditions of the buying and selling clubs Only uses data from top 5 European leagues, as a result, it is less generalised.

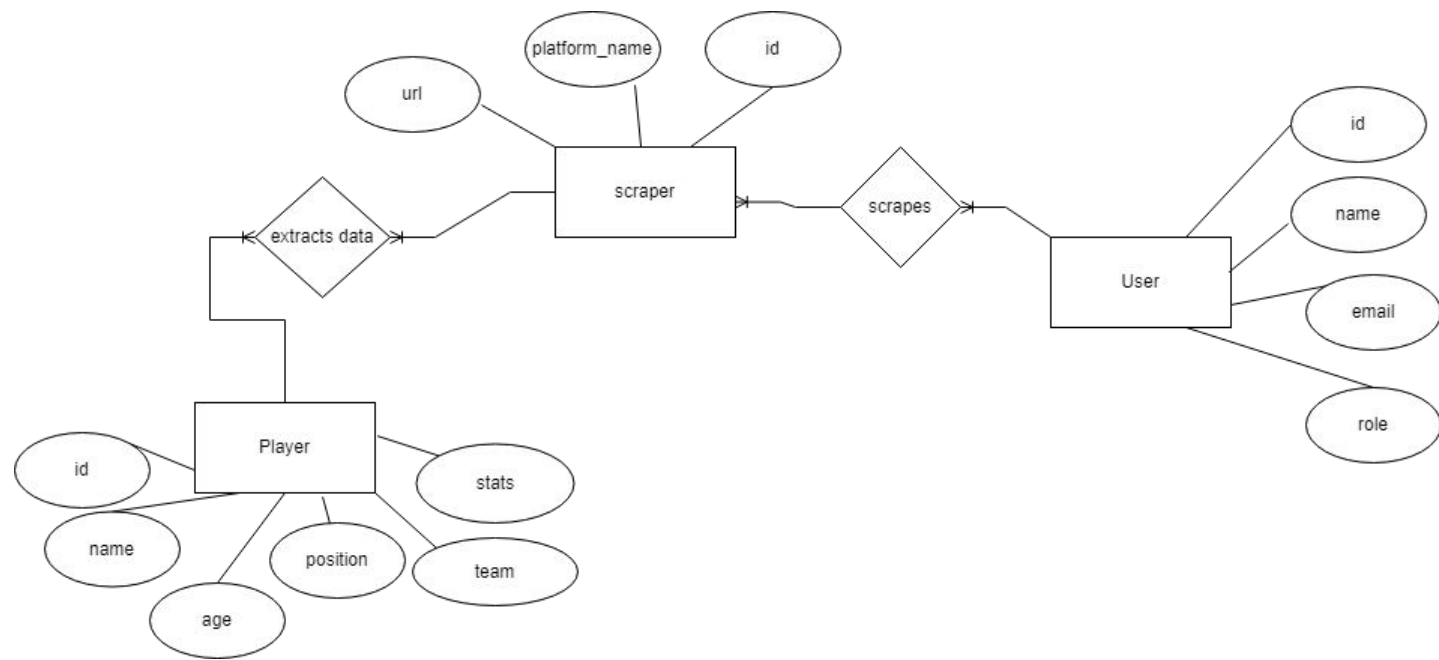
Existing System

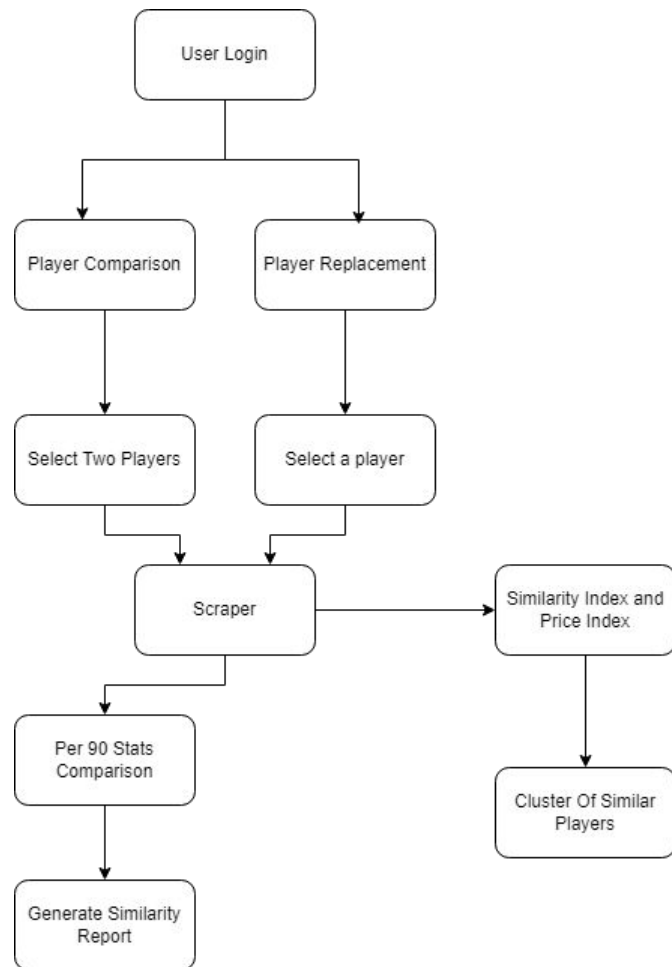
Flaws:

- Has only 2 seasons worth of data for each player (20-21, 21-22)
- Lets you select a team for replacement, which the player to be replaced does not even play for
- Lets you select position which the player to be replaced does not play in.
- Does not look at price as a parameter

Problem Definition

- In football, identifying suitable player replacements is a crucial yet complex task traditionally reliant on labor-intensive manual video reviews. This method is not only time-consuming but also subject to human biases and inconsistencies. Football clubs face significant challenges in optimizing player recruitment and tactical adjustments while adhering to financial constraints and performance goals.
- The need for a more efficient, objective, and strategic approach to player replacement is evident. Current practices lack the integration of advanced data analytics and automated systems that can enhance decision-making processes. There is a critical gap in leveraging comprehensive player statistics and machine learning to facilitate precise, data-driven player comparisons and replacements.





Results

Machine Learning Models

- Models used: Linear Regression, XGBoost, LightGBM, RandomForestRegressor etc.
- Best Model observed: XGBoost
- New method: Ensemble Learning with multiple models
- Improvement observed: 33%

Baseline model:

```
Mean Absolute Error: 6761496.415649414
R2 Score: 0.5802977711217809
Accuracy: 0.4889961242254811
Root Mean Squared Error: 14716142.610714935
```

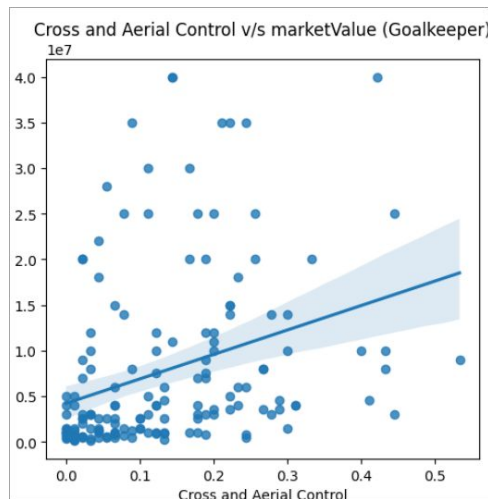
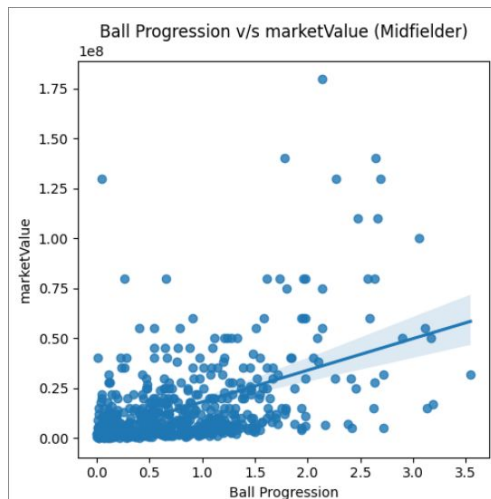
Ensemble learning:

```
Stacking Ensemble -> RMSE: 4595671.66, R2: 0.8816
Bagging Ensemble -> RMSE: 8693049.26, R2: 0.5764
Boosting Ensemble -> RMSE: 7150573.47, R2: 0.7134
```

Results

EDA and Key Insights

- Engineered new metrics for positions that had **stronger correlations** with market value compared to traditional stats.
 - Midfielders:** Metrics like **Ball Progression** and **Possession Retention** directly impacted their market value.
 - Forwards:** **Shooting Accuracy** stood out—teams want strikers who can consistently hit the target.
 - Goalkeepers:** The importance of the **Cross and Aerial Control** stat was evident, showing goalkeepers with better aerial control had higher market value.
- Increase of the R^2 score by **27.6%**



Results

Interactive Website

Key features of the website are:

- Multi-page and dynamic in nature
- Player Statistics can be explored with ease
- Graphical representation for comparison of players
- Leagues can be analyzed with interactive visualisations

