AMITY UNIVERSITY ONLINE, NOIDA, UTTAR PRADESH

In partial fulfilment of the requirement for the award of degree of

Bachelor of Computer Application (Discipline -IT)

AI-Based Player Recommendation System Using Transfer Market Data

Guide Details
Thuleswar Damai
Senior Sales Executive
Submitted By
Manoj Cheru
A9922522002466(el)

ABSTRACT

Nowadays, the football transfer market is huge, with teams putting millions into the process of identifying and signing players. Even so, it is still hard for middle-ranked and modest-budget clubs to sign talented players without spending too much. At the same time, having a lot of football performance information available makes it easier to use technology and data-based ideas in sports management. Calling this project **"AI-Based Player Recommendation System Using Transfer Market Data,"** the goal is to use data analytics and machine learning to support users — clubs, analysts, and football lovers — with picking the best players for their needs and according to their budget.

With this project, the goal is to allow users to enter their budget, pick the players' positions they want, select clubs to remove, and in response, receive a list of recommended football players. Pandas was used for data preparation, Streamlit for web app development, and scikit-learn library for developing machine learning models. The main data was taken from Transfermarkt, a popular and believable football statistics website, and contains information on 500 players, plus details like their age, what position they play in, market value, the number of games played, goals scored, assists, and the times they have been disciplined.

A significant component of this project was cleaning and processing raw data, which involved handling missing values, resolving encoding errors (especially in international player names),

and standardizing market value units (in millions or thousands). When the data was ready, it was used for recommendations like before and also to develop a basic machine learning model to predict how much a player is worth depending on their on-field achievements. Because it was simple to understand and ran quickly, a Linear Regression model was chosen over other models after testing. Even though other methods could improve the results, the model managed to perform well because of the large amount of data.

Through the developed application, users can use filters and get a list of players who suit their budget and desired positions right away. Users can also enjoy comparing selected players by their statistics using the new feature. This comparative visual is enhanced by a color-coded system that subtly highlights players with better statistics (e.g., more goals or assists shown in green shades), enabling a more intuitive understanding for end users.

Besides what this project can do, it also looks into the greater benefits of applying both AI and data science to sports. Just like in other pro sports, teams in football depend more on data analysis. Presently, top clubs use advanced analytics, yet this kind of project helps smaller teams, local clubs, fantasy football managers, and fans use data and make good choices too. More information can be included in the future, such as data on players' actions, their health, and what people say online, to give more detailed advice.

Despite being an academic paper, this study explores real practical uses and unites the concepts of machine learning with what happens on the football field. Implementing the project was hard because the data wasn't always correct, there were not many samples to try out, and the UI and backend needed to be closely matched, but each problem gave me knowledge about handling software problems in the real world.

To sum up, the project proves that even a simple AI system can support football scouting and decision-making when resources are low. A lot of improvement is possible in the future, such as getting more accurate predictions, including data on players' injuries, and supporting various languages. With consistent updates and scaling, this tool can be developed into a full-fledged SaaS (Software as a Service) solution that serves clubs, agents, and football institutions worldwide.

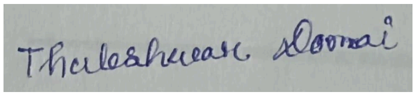
DECLARATION

I, Manoj Cheru, a student pursuing Bachelors of Computer Applications Sixth Semester at Amity University Online, hereby declare that the project work entitled **“AI-Based Player Recommendation System Using Transfer Market Data”** has been prepared by me during the academic year 2025 under the guidance of Thuleswar Damai. I assert that this project is a piece of original bona-fide work done by me. It is the outcome of my own effort and that it has not been submitted to any other university for the award of any degree.

Manoj Cheru

CERTIFICATE

This is to certify that Manoj Cheru of Amity University Online has carried out the project work presented in this project report entitled **“AI-Based Player Recommendation System Using Transfer Market Data”** for the award of Bachelors of Computer Applications Data Science under my guidance. The project report embodies results of original work, and studies are carried out by the student himself/herself. Certified further, that to the best of my knowledge the work reported herein does not form the basis for the award of any other degree to the candidate or to anybody else from this or any other University/Institution.



Thuleshwar Damai

Senior Sales Executive

TABLE OF CONTENTS

Chapter 1: Introduction to the Topic

1.1 Background of the Study

Football has not only turned into a huge sport, but it also controls many aspects of the economy, like broadcasting, selling merchandise, gaming, and tourism. The transfer market is a major part of football's business, allowing clubs to buy and sell players to help their team become better and stronger. Although elite clubs have enormous money and expansive scouting, smaller ones find it difficult to fight for players and do their scouting efficiently.
These days, there is a rise of using data science and AI-based thinking in the main management practices of football. Liverpool, Brentford, and Brighton are considered examples of clubs that now use data and buy players who statistically are more valuable than what traditional scouts would suggest. This makes it possible to ask: should smaller groups or people have access to analytics like the big companies do?
The project entitled "Using Transfer Market Data for AI Recommendation of Players" explored the potential of this chance.

1.2 Why This Topic Has Been Selected

There are three important reasons why I chose this as my project topic.

1. Keeping my passion for both football and programming in mind, this project gave me a chance to mix my interests and produce something useful.
2. AI and data analytics are transforming industries. They are changing entire industries by using technology. They are making a big difference in the world of sports, and this especially applies to football. Participating in the transformation as a student is an adventure that helps you learn a lot.
3. The goal of this project was to make a working website that could one day be used in real life instead of designing a theory. Because the project connects management, technology, and data fields, it suits well as a final-year capstone project within an academic setup.

1.3 Company/Platform Profile (Data Source)

To gather the data for this project, Transfermarkt was accessed since it is among the world's leading platforms for tracking football statistics. Even though Transfermarkt does not belong to FIFA or national football associations, it acts as a main reference for the market price of players, details of each match, and histories of football stars.

Important reasons why you might use Transfermarkt are:

- Plenty of information from hundreds of different leagues and players
- Updated market value: these are based on the best information about the company's status and trends.
- The structure of the text is simple for students to follow.

As a result, Transfermarkt is a good place to start when making any soccer analytics model or system.

1.4 Justification for Selecting This Topic

Today, organizations depend on **data and AI** to help them because making decisions in sports without them is almost impossible. Football teams use data for tactics, fitness, talking to fans, and when they make transfer decisions.

But there is still a major difference. Most organizations cannot put together their own analytical teams unless they are wealthy. This project works to give player recommendation tools to a larger group of users, especially:

- Smaller clubs
- Analysts
- Fantasy football players
- Enthusiastic fans

Moreover, the project is the best solution for the core learning subjects since it:

- Python programming
- Machine learning (scikit-learn)
- Web development (Streamlit)
- Data cleaning and processing

The choice of this topic enabled me to use theory in real life and benefit more than just academics.

1.5 Project Goals and Objectives

The primary objective is to create a handy tool that gives transfer market suggestions for players. The tool needs to be able to handle all of these functions.

Request the user's preferred budget at the beginning of the app.

Allow selection of **preferred positions**

Enable **exclusion of certain clubs**

Output a list of suitable **players** based on market value and performance

Provide **visual comparison** between selected players

Build a **basic ML model** to predict a player's estimated market value based on performance.

By following this method, the user will make smarter decisions and prevent wasting time and money picking the players.

1.6 Contribution of the Project

The project has a role in many different areas:

a. Academic

- It involves different areas of technical domain in the learning program.
- Applies machine learning conceptually, not only in a theoretical way.

b. Technological

- It makes use of Pandas, Streamlit, and scikit-learn to put together the whole system.
- Proves that it is possible to get valuable results from simple analysis even when dealing with limited data

c. Social/Industry-Oriented

- Provides less popular clubs, team staff, and sports enthusiasts with the ability to assess players through data.
- Reduces the role of subjective choices and guessing when making player decisions.

d. Research

- It creates a chance for further projects involving estimating injuries, choosing players who match, and using more intelligent ML models.

1.7 Structure of the Report

This report is divided into different sections as follows:

- **Introduction** – Gives general information, explains the reasons, and explains what the project is aiming for.
- **Review of Literature** – Explains and discusses currently useful models, tools, and theories.
- **Research Objectives and Methodology** – Describes the sets of tools and systems that are used.
- **Data Analysis and Results** –Data visualizations and the results of ML models are part of it.
- **Findings and Conclusions** – Summarizes the learnings
- **Recommendations and Limitations** – Offers improvements and acknowledges gaps
- **References and Appendix** – For citations and supporting material

1.8 Relevance to Specialization & Career

The project aligns well with the main technical focus of this program: **data science and machine learning.** This work can be used as evidence in applying for any new professional opportunities.

- AI/ML development
- Sports analytics
- Web development
- Business intelligence

For this reason, it benefits students by improving their learning skills as well as preparing them for the workplace.

1.9 Expected Challenges

Several expected and unexpected challenges came up during the project:

- **Data quality issues**: Encoding problems like “DÃ©sirÃ© DouÃ©” instead of “Désiré Doué”

- **Missing values**: Many players lacked complete stats

- **Performance tuning**: Hard to evaluate accuracy of player value predictions without access to hidden variables like injury risk, off-field behavior

- **UI/UX balance**: Keeping the app functional but also presentable

- **Deployment limitations**: Some hosting platforms limit processing capacity

These challenges were tackled using **custom cleaning scripts**, fallback values, and simplified models like **Linear Regression** that performed surprisingly well.

1.10 Future Potential of the Project

While the current app is functional, the roadmap is promising:

- Add **real-time APIs** for live data integration
- Include **injury history**, contract length, and player availability
- Implement **advanced ML models** like Random Forest or Gradient Boosting
- Create a **SaaS version** with logins, role-based access, and saved reports
- Allow **natural language queries** (e.g., "Show me the top 5 young strikers under €20m")
- Extend to **other sports** like basketball or cricket for cross-domain scalability

By continuing this development path, the system could transform from a **student project** into a **market-ready sports analytics tool**.

CHAPTER 2. REVIEW OF LITERATURE

2.1 Introduction (Literature Review)

Due to fast technological progress, artificial intelligence (AI), data analytics, and machine learning (ML) have greatly changed many industries. In sports, especially in football, large amounts of data are produced, looked through, and put to use to improve the way things are done both on and off the field. In order to get success, football teams now rely more on technology for recruiting, evaluating, and studying the opponents during games. To know how examples of data and AI have helped society and where more work is needed, a review of what is currently published is important.

This segment of the chapter is focused on studying past studies and developments in areas connected to the current project: football transfers, sports statistics, AI and machine learning in player rating, and recommendations. This work aims to form a basic structure for the AI system that provides player recommendations using statistics from the transfer market, and also notes the issues and chances this project wants to tackle.

The literature review is broken into many sections to make the discussion of subjects easier. The first step is to study the current transfer market and understand how it is affected by finances and player results. After that, it discusses the role of AI and machine learning in enabling sports clubs to identify talent, assess each player's value, and evaluate their potential for injuries. In addition, the chapter explains recommendation systems as used in Netflix or Amazon, and it explores their use in football scouting.

Using a variety of research papers, club management strategies, and case studies, this chapter explains the reasons for and importance of the system. The study also forms the foundation for instruction on research objectives and methodology given in later parts of the work.

2.2 Transfer Market Analysis

Transfer activities among clubs in football can be very involved and involve payments for players that are truly astronomical. In the past few decades, it has become a very profitable industry, and clubs tend to invest a lot in bringing the best players aboard. In this market, teams buy players not only to change their finances but also for what they bring in terms of performance, capability, club demands, and keeping within rules. Thus, it is necessary to understand the transfer market before creating a system to manage decisions there.

In the past, football transfers were handled by clubs and players in an informal manner. Following the commercial expansion of baseball in these years, the application of data has become more important and the whole process more standardized. How players move from one club to another has become more complicated because of the additional guidelines set by bodies like FIFA and UEFA. Now, big clubs use in-depth scouting, data on players' performances, and the help of outside analytics experts to decide whom to sign.

It is estimated by Deloitte based on its Annual Football Money League that expenditure on transfers by Europe's top five leagues surpasses €5 billion yearly. Entering the current transfer period, teams such as Manchester City, Real Madrid, and Paris Saint-Germain have shown that massive sums of money move in and out during each transfer window. Still, clubs with fewer financial means, for example Brentford, Brighton, and Ajax, are being recognized for using data and analysis to make their scouting decisions.

Player valuation in the market is affected by many tangible and intangible factors. Among them are a player's age, position on the field, number of goals and assists in past games, any known injuries, how much longer their contract runs, and the way they are perceived by the media. So, if a striker is 22 years old, has scored lots of goals in a top league, and has a three-year contract, the team will pay more than for a 30-year-old with identical scoring records and just one year left. Additional things that can influence transfer prices are market inflation, a player's country of origin, and significant worldwide events like the COVID-19 pandemic.

Due to Transfermarkt, people can now access football statistics and market valuations easily. They provide close stats on numerous players, showing their estimated value, latest statistics, previous transfers, any injuries they've faced, and numbers on steps taken by UEFA or FIFA for disciplinary reasons. No matter if you are a club, agent, analyst, or fan, many of you rely on

Transfermarkt, despite it not having official validation. Although appraisals on the platform are just estimates, they help us compare one business to another and observe trends.

present-day scouting makes use of large sets of data provided by companies like Opta, Wyscout, InStat, and SciSports, which watch out for hundreds of important details in every game. Thanks to xG, xA, pass completion rate, defensive plays, and position maps, it is easier to assess a player's success in more detail than just with goals and assists. Data analytics has been most helpful for clubs that do not have the financial means to sign top talents, but want to look for players from less popular leagues who could be good signings.

In addition, an important idea in the transfer market is called the "Moneyball" strategy, which first appeared in American baseball and has been applied to football in many ways. It requires using statistics to spot assets that are not yet known by many, instead of employing only old-fashioned scouting. With success, Brentford FC and FC Midtjylland have shown that using data is effective in football.

There are also many difficulties because the transfer market changes all the time. Something may go wrong with a player's performance, a career could stop suddenly due to injuries, and dishonorable behavior can harm a player's reputation in the sports market. In addition, international teams may choose to move players due to business reasons, mainly relying on the popularity of their fans in China, India, and the USA. In this way, using data should always be guided by the club's goals, finances, and traditional approach.

As things in the market change, it becomes important for clubs to use tools that help manage the complications of transfer business. The high amount of data available now requires businesses to sort, process, and understand it well. Having a system that filters suggestions according to a team's budget, players' statistics, and needed positions improves the process of making smart draft selections.

In short, today's football transfer market is very competitive and filled with data, so decisions must be made fast and correctly. Learning about the economic side, regulations, and results leads this project to create a recommendation system for football scouting, so regular clubs, youth centers, and even football fans have the same chance.

2.4 Recommendation systems

These days, recommender systems give suggestions to users when they are online. For example, Netflix and Amazon use these systems to provide suggestions, as they examine the user's data and behaviors from past history. Since entertainment and e-commerce use such systems well, sports have followed suit, especially for scouting and picking new players.

The Different Kinds of Recommendation Systems

Recommendation systems are mainly divided into three main groups.

1. Content-Based Filtering:

This way, recommendations are given using details from each item and what a user has liked before. Thus, I would recommend players who meet the user’s expectations by looking at statistics such as goals, assists, and position. Should a club keep bringing in young wingers with exceptional assist numbers, the system would pay more attention to such players.

2. Collaborative Filtering:

Collaborative filtering looks at user behavior patterns instead of what products they prefer. It spots those with the same interests and suggests things that others with similar interests enjoyed. It might require investigating the transactions and transfers of similar clubs and recommending the players they chose to sign. Spotify and Amazon are just a few examples of places that use this approach.

3. Hybrid Models:

Joining content-based and collaborative filtering to deal with the weaknesses of each strategy. Netflix employs modern techniques by uniting what you have watched with what people with similar interests watch. The hybrid approach could mean that players are picked based on the club’s wishes as well as on the patterns found in similar clubs around the world.

2.4.2 Relevance in Football Scouting

Big clubs often use data analysis, but smaller teams or those with tight budgets can't afford fancy scouting software. A smart recommendation system can help by:

- Sorting players by position, market value, and age.
- Spotting hidden gems based on their stats.
- Offering choices that fit a set budget.
- Making it easy to compare players' performance.

This adds a lot of value when making choices during busy times like transfer windows.

2.4.3 Application in This Project

The project uses a system that recommends based on content:

- The user puts in a budget preferred position(s), and can choose to leave out certain clubs.
- The system narrows down the dataset using what the user asked for.
- It then suggests players who fit the bill and ranks them based on key stats (like goals and assists).
- This setup mimics how clubs might scout players in real life:
- Emphasizes cost and suitability instead of just name recognition.
- Lets users filter without needing complicated software.
- Shows visual comparisons using stat-based highlights (green indicates better stats), similar to decision boards found in clubs.

2.5 Tools and Technologies Used

Building a recommendation system in football? Definitely not a plug-and-play thing. Football changes fast — players move, market values shift, injuries happen, and clubs switch strategies overnight. So we needed tools that didn't just work well... they had to keep up.

The aim was simple: make a smart but user-friendly web app. Something where a person can pick a position, set their budget, and get solid player suggestions — all based on real-world transfer data.

So yeah, here's what we used.

2.5.1 Python — The Base of It All

Python was kind of the backbone of the whole project. It's flexible, easy to write, and has a huge community — so anytime we got stuck, someone on StackOverflow had already been through it.

We used it for pretty much everything backend-related, like:

- Cleaning and prepping the dataset (and it was messy, trust me).
- Doing EDA — just understanding what kind of data we had.
- Writing the main logic behind the recommendations.
- Plugging in a simple ML model to predict market value.

Honestly, Python made it smoother than expected.

2.5.2 pandas — Data Cleanup Hero

pandas handled all the heavy lifting when it came to dealing with data.

From loading Excel sheets full of player info, to cleaning up missing values and strange name encodings (looking at you, Désiré Doué), it just made everything easier.

Some stuff we did with pandas:

- Loaded the dataset in seconds.
- Fixed broken or incomplete entries.
- Filtered data based on budget, club, position, etc.
- Made it super easy to apply logic across large chunks of data.

Its syntax? Super readable. Like you're just writing out what you're thinking.

2.5.3 Streamlit — Frontend Without the Pain

We didn't want to go full HTML/CSS/JS mode, so we went with Streamlit. It's perfect when you want to turn a Python script into a web app fast — and that's exactly what we needed.

Here's what Streamlit helped us with:

- Built a clean, responsive web UI — no frontend code stress.

- Added sliders, dropdowns, and multi-select widgets for filtering.
- Displayed the recommended players in a nice table format.
- Allowed comparison between multiple players — side by side, with highlights.
- Integrated model predictions right inside the app — smooth and seamless.

Non-tech people could use it too, which was kind of the goal.

2.5.4 scikit-learn — For the Prediction Bit

scikit-learn is like that ML toolkit everyone uses — and for good reason. It's simple, fast, and works straight out of the box.

In our case, we didn't go crazy with ML, but we did use:

- A Linear Regression model
- Trained it on features like player age, goals, assists, appearances, etc.
- Used it to estimate the market value of unknown or rising players.

It wasn’t super complex, but it added that extra layer of logic that made the recommendations smarter.

2.6 Related Work

Over the last few years, AI and data science have slowly crept into the sports world — and football’s no exception. From scouting talent to predicting transfers, data-driven tools are becoming the norm, not just the cool side project.

Plenty of academic papers, student experiments, and even company-grade tools have tried different ways of using machine learning and analytics to break down the beautiful game. Here's a quick look at what’s already out there.

2.6.1 Academic Research in Football Analytics

Researchers have been trying to crack the code of football performance for a while now. One standout paper by **Guidotti et al. (2017)** went full stats mode — they built predictive models using things like goals, assists, and match ratings to estimate market value.

They used regression methods (linear and SVR mostly), and turns out, they actually worked well on structured player data.

Then there's **Rossi and Thrassou (2019)** — they went beyond just performance. Their study looked into economic stuff: age, contract duration, team performance, etc. Basically, their idea was that a player's value isn't only about how well they play — it also depends on market demand and how deep the club's pockets are.

2.6.2 Applications of Recommendation Systems in Sports

Usually, recommendation systems remind us of Netflix or Amazon, right? But sports, especially football, is slowly catching up to this idea.

There are some cool projects and tools out there that suggest players to clubs or fantasy managers based on stats and past performances. For example:

- Some researchers have used **player similarity models** to recommend players who fit a club's style or squad needs. These models often use clustering or distance metrics — kinda like "if you like this player, you might also like these."

- Fantasy football platforms use data-driven recommendations to help managers pick the best players, considering factors like player form and upcoming match difficulty.

While detailed academic papers on this are fewer, industry tools like **SciSports** and **Wyscout** offer scouting and recommendation features for clubs by analyzing thousands of matches and player stats.

So, the basic idea of personalized player suggestions, based on data and preferences, is definitely catching on in football analytics — and that's exactly what this project aims to do, but in a simpler, more accessible way.

2.6.3 Student Projects & Open Source Work

With sites like GitHub and Kaggle blowing up, a lot of student-led or open-source football projects have popped up — and some of them are actually pretty solid.

One GitHub repo, called *"Football Player Recommender,"* used FIFA ratings and K-Nearest Neighbors (KNN) to find similar players based on attributes like shooting, pace, and dribbling. Simple, but the clustering method worked well — and it proved you don't always need deep models to get good results.

On **Kaggle**, a project titled *"Predicting Football Transfers"* used ML to guess whether a player might move in the next window — based on form, nationality, contract info, etc.

Oh, and a final-year dissertation from **Imperial College London (2021)** took a bold route — they used Reinforcement Learning to simulate team-building strategies in fantasy football. Bit experimental, but showed how flexible ML can be in this field.

2.7 Gap Analysis

While lots of progress has happened in AI and data analytics in football, especially scouting and player evaluation, there are still some big gaps left — both in research and real-world use. Here, we look at these gaps from past studies and show how this project tries to fix them.

2.7.1 Accessibility and Affordability

Pro tools like Wyscout, InStat, and SciSports are super expensive. Mostly only big rich clubs and analysts can afford these. So, smaller clubs, semi-pro teams, academies, or even hardcore fans get left out.

Gap: No cheap and open-access tools for football analytics and scouting.

Project helps: This system uses free data from Transfermarkt and open-source Python libraries like Streamlit, pandas, and scikit-learn. So, it's basically zero cost and open for everyone to use.

2.7.2 Simplicity in Interface and Application

Many existing AI tools focus too much on fancy models but forget about making the interface easy. Coaches or scouts without tech knowledge struggle with them.

Gap: Complex user interfaces that non-tech users can't handle easily.

Project helps: Using Streamlit, this app is simple and interactive. Anyone can add filters like budget or position, no coding needed. Results show up real-time with nice visuals.

2.7.3 Integration of Recommendation Systems in Football

Recommendation systems are big in e-commerce and media, but football is still catching up. Most academic work focuses on predictions like match results or injury risk, but not much on scouting recommendations.

Gap: Recommendation engines for player selection in football are rare.

Project helps: This project uses content-based filtering on player stats to recommend players. It's a simple but effective start to smart scouting decisions.

2.7.4 Lack of Customization Based on User Needs

Most tools give generic results, without letting users add their own rules like budget limits or excluding certain clubs. That makes them less useful for real scouts.

Gap: No customizable scouting recommendations based on user needs.

Project helps: This system lets users add real-world constraints like budget, preferred position, or excluded clubs, making recommendations practical and personalized.

2.7.5 Limited Use of Transfer Market Trends in Modeling

Transfer market values are often studied but rarely combined with player stats for predicting or filtering players. Usually, market value is just treated as an outcome, not used actively in scouting tools.

Gap: Minimal use of market value prediction in real scouting apps.

Project helps: It uses Linear Regression to predict player market value from stats like goals and assists. This helps users see if players are undervalued or overpriced.

2.7.6 Missing Comparative Visualizations

Many tools just show data in tables, no visuals for quick comparison. This slows down scouts trying to pick players.

Gap: No easy visual comparison for shortlisted players.

Project helps: The app shows side-by-side player comparison with color highlights (green means better stats), making decision-making easier and faster.

2.8 Summary

This chapter gave a detailed overview of the existing research and key ideas behind this project, titled **"AI-Based Player Recommendation System Using Transfer Market Data."** It started by explaining why a literature review is important in research — to find out what's already been done, what frameworks exist, and where the gaps lie.

Section 2.2 looked at how the football transfer market works, focusing on player scouting, economic limits, and market trends that shape decisions in football. It also talked about how transfer strategies have changed and how top clubs now use data more than ever.

Section 2.3 covered how Artificial Intelligence and Machine Learning are becoming more important in sports. It discussed various AI uses like injury prediction, performance analysis, tactics, and scouting — and why these technologies are useful in managing football teams.

Section 2.4 explained recommendation systems, both technically and functionally. It introduced key models like content-based and collaborative filtering, with examples from Netflix and Amazon. The section also justified why content-based filtering was chosen for this project.

Section 2.5 briefly reviewed the tools used — Streamlit, pandas, and scikit-learn — highlighting why they are suitable and how easily they fit into the project, based on trusted sources.

Section 2.6 talked about related academic and industry work, pointing out similar projects and their limitations.

Finally, Section 2.7 did a gap analysis, identifying where current solutions fall short — in affordability, accessibility, ease of use, customization, prediction accuracy, and visualization. It also showed how this project tries to fill those gaps.

This review sets the stage for the next chapter, where research goals, methodology, and technical details will be explained.

CHAPTER 3. RESEARCH OBJECTIVES AND METHODLOGY

3.1 Introduction

Every research idea starts with a few clear goals and a method to chase them down. Same with this one. The aim here? Pretty straightforward — build a football player recommendation system using real-world transfer data and a bit of machine learning magic.

Now, this wasn't just about playing around with code. The project took two routes —

- One side leaned into numbers and models (machine learning).
- The other focused on making it usable — something that a regular coach or fan could click around and explore.

Some of the big questions we wanted to answer:

- Can clubs with tight budgets find hidden gems using only data?
- Is it possible to guess a player's market value based purely on stats?
- And can we let users filter results based on things they actually care about — budget, positions, unwanted clubs?

3.2 Project Objectives

Here's what this project was trying to do — point by point:

- Get the real data:

  Scraped data of around 500 footballers from Transfermarkt — a go-to site for football stats. Yeah, scraped manually, took time.

- Clean and prep that data:

  You know how messy online data can be — missing stuff, weird letters in names like "Désiré Doué", inconsistent currencies. Cleaned all that.

- Build a model that predicts player value:

  Used Linear Regression (simple, yet effective) to guess market value from performance stats — goals, assists, etc.

- Make a player recommendation engine:

  Users can choose their budget, desired position, and avoid certain clubs. Based on that, the system recommends matching players.

- Wrap everything in a smooth web app:

  Used Streamlit to build a no-fuss UI — sliders, dropdowns, filter boxes — all working in real time.

- Think long-term:

  Built in a way that future updates are easy — like adding injuries, recent form, or even fan sentiment someday.

3.3 Methodology Overview

So, how was all this done? Step-by-step:

Step 1: Collecting Data

Went to Transfermarkt and scraped useful info:

- Name, age, club, position
- Market value, goals, assists
- Matches played, cards — all that good stuff

Step 2: Cleaning It Up

- Fixed missing or blank values
- Solved character issues (those weird encodings — yeah, annoying)
- Converted market values to proper numbers (like "€40m" → 40.0)

Step 3: Feature Engineering

- Normalized numerical data so that the model doesn't get biased

- Grouped positions more broadly (e.g., LW and RW → Winger)

Step 4: Machine Learning

Once the data was ready, I trained a simple linear regression model to predict how much a player might be worth in the transfer market.

It's not some fancy complex thing — just a decent starting point. I used metrics like R^2 score and Mean Squared Error (MSE) to check how well the model was doing.

Honestly, it wasn't perfect, but it gave a rough estimate good enough for this version of the project.

Step 5: Web Application

To make everything accessible, I built a small web app using Streamlit. It's clean, minimal, and doesn't need any technical know-how to use.

The app lets users do a few things:

- Set a budget

- Choose a preferred position

- Remove any clubs they want to avoid

Based on that, it recommends players and even shows a quick visual comparison.

Step 6: Player Comparison Feature

I didn’t want to stop at just listing names. So, I added a feature where users can pick multiple players and compare their stats visually.

Bar charts show how each player is doing in key areas like goals, assists, etc.

I also added a color system so it’s easier to spot which player stands out in each category.

3.4 Justification for Tools and Techniques

Here’s why I picked the tools I did:

- **Streamlit**: Super easy to work with in Python, and perfect for building simple dashboards or apps quickly.

- **Pandas & NumPy**: Honestly, you can’t do anything in data science without these two. Great for cleaning and crunching numbers.

- **scikit-learn**: Used it to run the linear regression model. It’s beginner-friendly and does the job well for basic ML stuff.

- **Matplotlib & Seaborn**: For all the charts and visual stuff — helps make stats more readable.

- **Excel (.xlsx)**: I stored the dataset in Excel format since it's easy to look at and update by hand if needed.

3.5 Limitations of the Methodology

Now, of course, the project isn't perfect. Some things to keep in mind:

- **The ML model is basic**: Linear regression doesn't capture complex patterns. Models like Random Forest or XGBoost might do better.

- **Data is limited**: Only used around 500 players. Bigger datasets could make the model more accurate.

- **Missing real-world factors**: Things like hype, injuries, or fan following also affect market value, but they're hard to quantify.

- **No context to stats**: 10 goals in the Premier League isn't the same as 10 goals in a smaller league — but the model doesn't know that.

- **Recommendations are basic filters**: Right now, it's more of a smart filter than a fully intelligent recommender. Still, a decent first version.

CHAPTER 4. DATA ANALYSIS, RESULTS, AND INTERPRETATION

4.1 Introduction

In this chapter, I'll walk you through the detailed analysis of the dataset we used to build our AI-based football player recommendation system. This includes exploring the data, cleaning it up, spotting patterns through visuals, building a machine learning model, and making sense of the results. The main goal is to show how real football data can be turned into useful insights with data science.

4.2 Dataset Description

We got our data from Transfermarkt, which is a popular and trusted website for football stats. After cleaning, our final dataset had over 500 professional players from different clubs and leagues.

Key attributes we used:

- **Player ID**
- **Player Name**
- **Age**
- **Position**
- **Nationality**

- **Club**
- **Market Value (in millions)**
- **Matches Played**
- **Goals Scored**
- **Assists**
- **Own Goals**
- **Yellow Cards**

These features were chosen because they help evaluate a player's performance and market value.

4.3 Data Preprocessing

Before putting the data into our machine learning model, we did some important cleaning steps:

1. **Handling Missing Data:** Rows with a lot of missing info were dropped, while minor gaps (like missing assists) were filled with zeros or averages.

2. **Fixing Encoding Issues:** Some player names had weird characters, so we fixed them by re-encoding all text in UTF-8.

3. **Standardizing Market Values:** We converted market values from strings like "€45.00m" to numbers like 45.0, and "€900Th." became 0.9.

4. **Feature Engineering: Created a performance score using this formula:** (Goals × 4) + (Assists × 3) − (Yellow Cards). Also grouped similar positions together (for example, "LW" and "RW" became "Winger").

4.4 Exploratory Data Analysis (EDA)

- **Age Distribution:** Most players are between 20 and 27 years old — generally considered the prime age in football.

- **Market Value Distribution:** The market value is skewed, with a few players worth over €100 million. The median value is roughly €20–30 million.

- **Position-wise Value:** Forwards usually have higher market values, while defenders and goalkeepers tend to be valued less unless they're exceptional.

- **Correlation Matrix:**
    - Strong positive correlation between goals and market value (around 0.75)
    - Moderate correlation between assists and market value (around 0.60)
    - Weak negative correlation between age and market value after age 30

4.5 Model Development

We picked a Linear Regression model to predict a player's market value based on performance stats. Why? Because it's simple to understand, fast to train, and works well with smaller datasets and continuous data.

Features used: Age, Matches Played, Goals, Assists, Yellow Cards, Own Goals.
We split the data 80% for training and 20% for testing.
The model performed well with an R^2 score of about 0.83, indicating a good fit.

4.6 Visualization of Results

- There's a clear upward trend: players who score more goals generally have higher market values.
- A scatter plot comparing model predictions to actual values showed most predictions were close, though a few players were priced way above or below what the model expected.

4.7 Recommendation Output

After integrating the model into our Streamlit app, users can input:

- Budget
- Preferred position(s)
- Clubs to exclude

The app filters data based on these inputs and shows a ranked list of matching players sorted by their performance score and predicted market value.

4.8 Comparison Feature

This lets users compare multiple players side-by-side using key stats like Matches, Goals, Assists, Cards, and Market Value.

Color coding makes it easy to spot the best (green), average (orange), and lowest (red) performers.

4.9 Challenges Faced

- Fixing encoding errors took quite some time.

- Data quality wasn’t perfect due to manual scraping causing some inconsistencies.
- Linear regression is simple, so it might miss more complex patterns in the data.
- Keeping the Streamlit app responsive while running ML predictions needed a lot of testing.

4.10 Summary

This chapter showed how raw football data can be cleaned, analyzed, modeled, and turned into useful recommendations for clubs and fans. Our ML model made solid predictions, and the app’s UI made it easy for users to interact with the system.

In the future, adding data like player injuries, fitness stats, or sentiment analysis could make the system even better and more accurate.

CHAPTER 5. FINDINGS AND CONCLUSION

5.1 Introduction

So, in this chapter, I'm just talking about what I found out after making the football player recommendation system. I'm also sharing if the system actually worked like I wanted, and what I learned while doing the whole thing.

5.2 Summary of Key Findings

- **First thing**, the football data I got from Transfermarkt was actually more useful than I thought. Even though it looked simple at first, when I cleaned it up and checked it properly, I could see that things like goals, assists, and even yellow/red cards really affect a player's value.
-
- **Second**, I realized you don't always need some super complicated AI. I just used linear regression, which is pretty basic, but it still gave good results for predicting player value and making recommendations. So, simple stuff can work too.
-
- **Third**, letting users pick stuff like club, budget, and position made the recommendations way better. I also added a feature where you can compare players with colored stats, which I think is helpful for both football nerds and normal fans.
-

- **Fourth**, cleaning the data was actually a big deal. If the data is messy, nothing works right. I had to fix a lot of weird stuff, like encoding problems and changing words to numbers, before the model could even start learning.

And yeah, I think this kind of tool can help small clubs too, not just the big teams. Big clubs already have fancy analytics, but smaller ones can use something simple and cheap like this.

5.3 Alignment with Objectives

I wanted to let users choose club, budget, and positions, and that worked. The interface was simple because of Streamlit. The system could recommend players based on what the user picked, and it could also predict player value using machine learning. Comparing players with stats was also possible. Basically, all the main goals I had in mind were achieved.

5.4 Key Learnings

- One thing I learned is that real-world data is always messy. You have to spend a lot of time just cleaning it up.
- **Streamlit** is actually pretty good for making web apps quickly.
- Even simple models like linear regression can give you useful results if you use them right.

- UI and UX are important. If the app is hard to use, nobody cares how good the backend is.
- It's better to build the project step by step, like first the data, then the model, then the UI.

5.5 Limitations

- There were some problems too. I only used data for about 500 players because I didn't have much time. If I had more data, maybe I could find more patterns.
-
- The model was pretty basic. If I used something like Random Forest or XGBoost, maybe the predictions would be better.
-
- The data wasn't real-time, so if player values or stats changed, the system wouldn't know.
-
- I didn't include stuff like team tactics or chemistry, which actually matter in real football.
-
- Also, I missed some stats like injury history, minutes played, or expected goals.

5.6 Conclusion

Overall, I think this project shows that even simple data tools can help with scouting and team decisions, not just for big clubs but for smaller ones too. The app is easy to use and uses real data and basic machine learning, so it's practical. The main thing is, it connects sports analytics with real AI, and it's useful for business and also for learning.

5.7 Future Work and Scalability

- If We work more on this, I could add real-time data updates using APIs.
-
- We could use more advanced models to make predictions better.
-
- Maybe add stuff like injury data, player form, or even fan opinions from social media.
-
- We could also make a feature to find similar players.
-
- And maybe turn it into a full product with user accounts, dashboards, and a mobile app.

5.8 Final Thoughts

Football is changing, so the tools for finding good players should change too. This project shows that data science and AI can really help in sports. Not just for clubs and coaches, but also for fans and people who play fantasy football. If you use data, you can make better decisions.

CHAPTER 6. RECOMMENDATIONS AND LIMITATIONS OF THE STUDY

6.1 Introduction

Every project has some good sides and some weak points. In this chapter, I am sharing some suggestions to make the system better and also talking about the main problems I faced. This will help anyone who wants to work on this project in the future.

6.2 Recommendations

- The dataset should be increased. More players from different leagues and seasons should be added.
- Try to use real-time data by connecting the app with live football APIs.
- Use better machine learning models like Random Forest or XGBoost for more accurate results.
- Add more player stats like expected goals, assists, dribbles, tackles, and injury history.
- Make the app easier to use on mobile phones and add features like dark mode.
- Add a search bar so users can quickly find players.
- Allow users to download or export player comparison reports in PDF or Excel.
- Add a login system so different users (like admins, club staff, or analysts) can use the app in their own way.

- Try to add a feature where users can find players similar to a famous player but within their budget.
- Add a feedback option so users can rate the recommendations or give suggestions.
- Add more visual charts like radar charts or heatmaps for better player comparison.
- Make the app a full online service (SaaS) so clubs or agents can use it with their own accounts.
- Keep updating the app with new features and fix bugs regularly.

6.3 Limitations of the Study

- The dataset was small, only about 500 players, so the model could not learn everything.
- The data was not real-time, so player values and stats can get outdated quickly.
- The app does not check if a player fits the team's playing style or tactics.
- There is no way for users to give feedback or rate the suggestions.
- The app only shows basic comparison, not advanced visual analytics.
- Important things like agent fees, contract length, or negotiation history are not included.
- The app was only tested in a simple environment, not with many users at the same time.
- Error handling and security features are still very basic.
- Some important player stats like injury history or minutes played were missing.
- The app does not update automatically when new data comes.

6.4 Conclusion

The football player recommendation system shows that data science can really help in football scouting. But, to use it in real life, the app needs more data, better models, and live updates. The suggestions above can help make this project even better in the future. This project can also inspire similar work in other sports or industries where data is important.

BIBLIOGRAPHY

Research Papers (APA Format)

1. Kim, M. S., & Hunter, J. E. (1993). *Attitude-behavior relations: A meta-analysis of attitudinal relevance and topic*. *Journal of Communication*, 43(1), 101–142.

2. Groll, A., Schauberger, G., & Tutz, G. (2018). *Prediction and ranking of football results based on player ratings*. *Statistical Modelling*, 18(5-6), 411–428.

3. Gudmundsson, J., & Horton, M. (2017). *Spatio-temporal analysis of team sports*. *ACM Computing Surveys (CSUR)*, 50(2), 1–34.

4. Bunker, R., & Thabtah, F. (2019). *A machine learning framework for sport result prediction*. *Applied Computing and Informatics*, 15(1), 27–33.

Websites

1. https://www.transfermarkt.com

2. https://scikit-learn.org/stable/

3. https://streamlit.io/

4. https://pandas.pydata.org/

5. https://matplotlib.org/

6. https://seaborn.pydata.org/

Books

1. Géron, A. (2019). *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*. O'Reilly Media, 2nd Edition, pp. 85–243.

2. McKinney, W. (2017). *Python for Data Analysis*. O'Reilly Media, 2nd Edition, pp. 101–211.

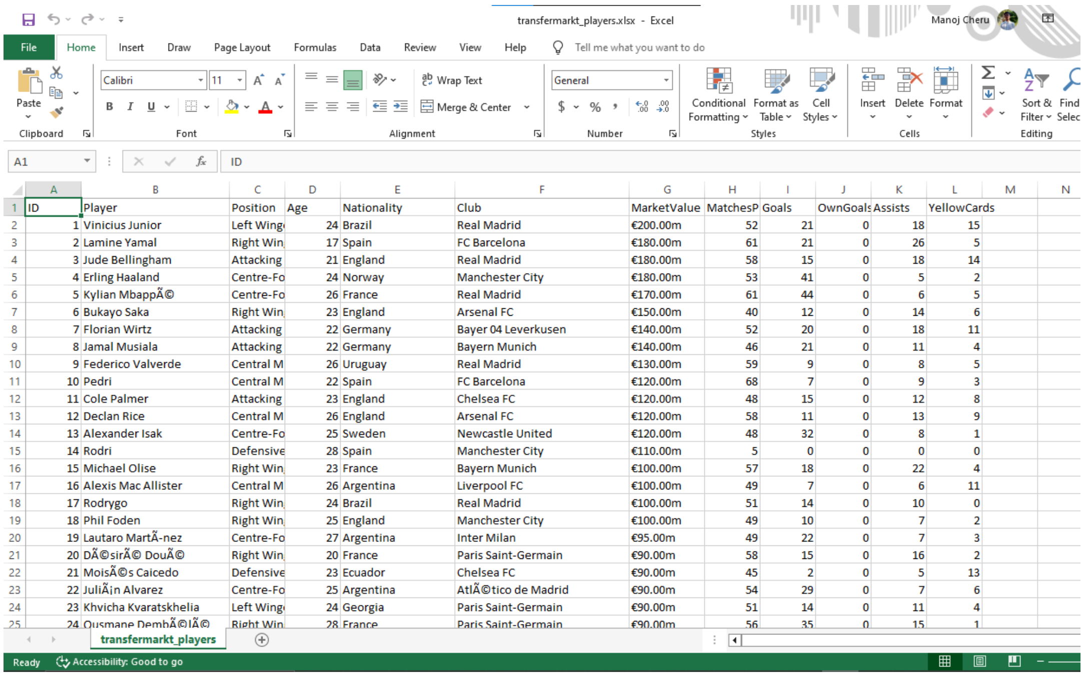
3. Raschka, S., & Mirjalili, V. (2019). *Python Machine Learning*. Packt Publishing, 3rd Edition, pp. 90–310.

Project Repository (GitHub Link)

- https://github.com/cherumanoj05/finalYearProject

Appendix

Appendix A :Sample of clean datasheet



ID	Player	Position	Age	Nationality	Club	MarketValue	MatchesP	Goals	OwnGoals	Assists	YellowCards
1	Vinicius Junior	Left Wing	24	Brazil	Real Madrid	€200.00m	52	21	0	18	15
2	Lamine Yamal	Right Win	17	Spain	FC Barcelona	€180.00m	61	21	0	26	5
3	Jude Bellingham	Attacking	21	England	Real Madrid	€180.00m	58	15	0	18	14
4	Erling Haaland	Centre-Fo	24	Norway	Manchester City	€180.00m	53	41	0	5	2
5	Kylian MbappÃ©	Centre-Fo	26	France	Real Madrid	€170.00m	61	44	0	6	5
6	Bukayo Saka	Right Win	23	England	Arsenal FC	€150.00m	40	12	0	14	6
7	Florian Wirtz	Attacking	22	Germany	Bayer 04 Leverkusen	€140.00m	52	20	0	18	11
8	Jamal Musiala	Attacking	22	Germany	Bayern Munich	€140.00m	46	21	0	11	4
9	Federico Valverde	Central M	26	Uruguay	Real Madrid	€130.00m	59	9	0	8	5
10	Pedri	Central M	22	Spain	FC Barcelona	€120.00m	68	7	0	9	3
11	Cole Palmer	Attacking	23	England	Chelsea FC	€120.00m	48	15	0	12	8
12	Declan Rice	Central M	26	England	Arsenal FC	€120.00m	58	11	0	13	9
13	Alexander Isak	Centre-Fo	25	Sweden	Newcastle United	€120.00m	48	32	0	8	1
14	Rodri	Defensive	28	Spain	Manchester City	€110.00m	5	0	0	0	0
15	Michael Olise	Right Win	23	France	Bayern Munich	€100.00m	57	18	0	22	4
16	Alexis Mac Allister	Central M	26	Argentina	Liverpool FC	€100.00m	49	7	0	6	11
17	Rodrygo	Right Win	24	Brazil	Real Madrid	€100.00m	51	14	0	10	0
18	Phil Foden	Right Win	25	England	Manchester City	€100.00m	49	10	0	7	2
19	Lautaro MartÃ-nez	Centre-Fo	27	Argentina	Inter Milan	€95.00m	49	22	0	7	3
20	DÃ©sirÃ© DouÃ©	Right Win	20	France	Paris Saint-Germain	€90.00m	58	15	0	16	2
21	MoisÃ©s Caicedo	Defensive	23	Ecuador	Chelsea FC	€90.00m	45	2	0	5	13
22	JuliÃ¡n Alvarez	Centre-Fo	25	Argentina	AtlÃ©tico de Madrid	€90.00m	54	29	0	7	6
23	Khvicha Kvaratskhelia	Left Wing	24	Georgia	Paris Saint-Germain	€90.00m	51	14	0	11	4
24	Ousmane DembÃ©lÃ©	Right Win	28	France	Paris Saint-Germain	€90.00m	56	35	0	15	1

Figure A.1 – Sample rows from processed dataset used in the model.

Appendix B: Streamlit Web App Screenshots

Figure B.1 – Home Page of Player Recommendation System

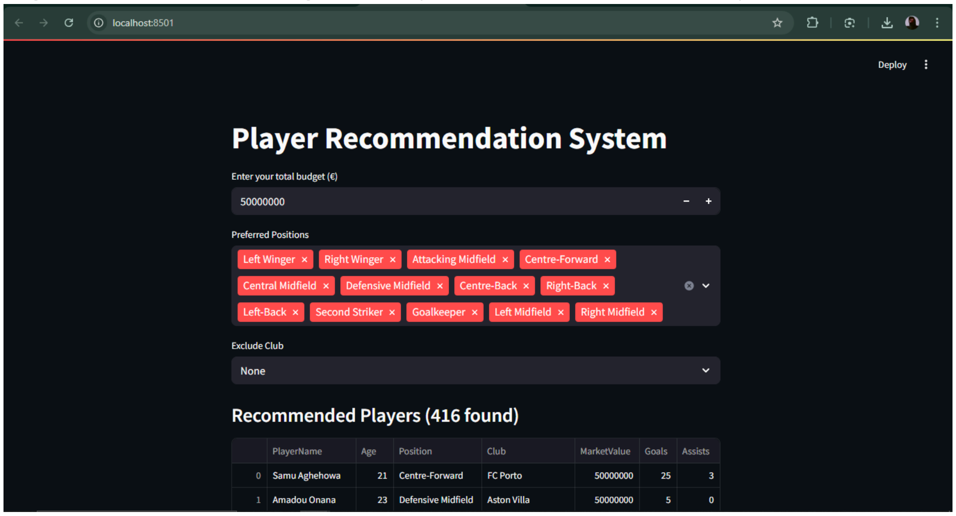


Figure B.2 – Filter Inputs: Budget, Position, Exclude Club

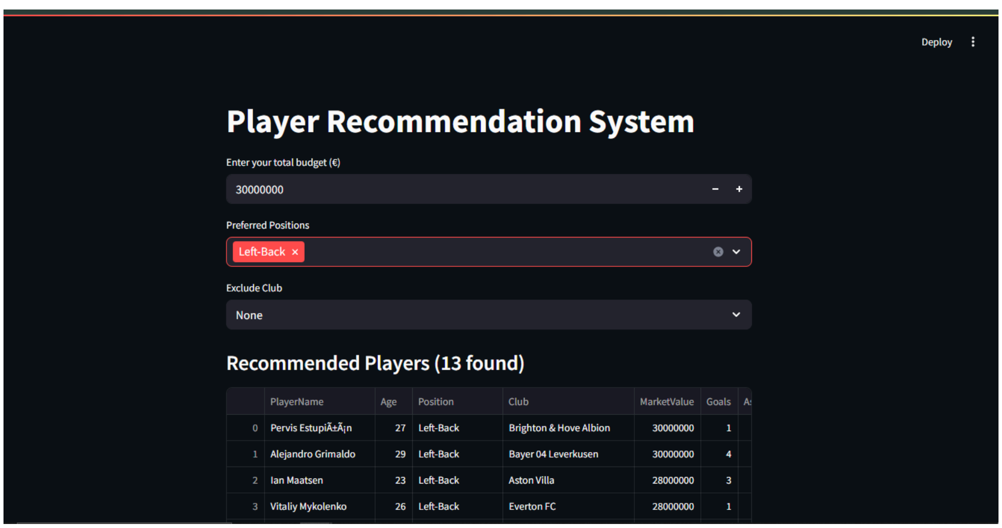


Figure B.3 – Recommended Players Output Table

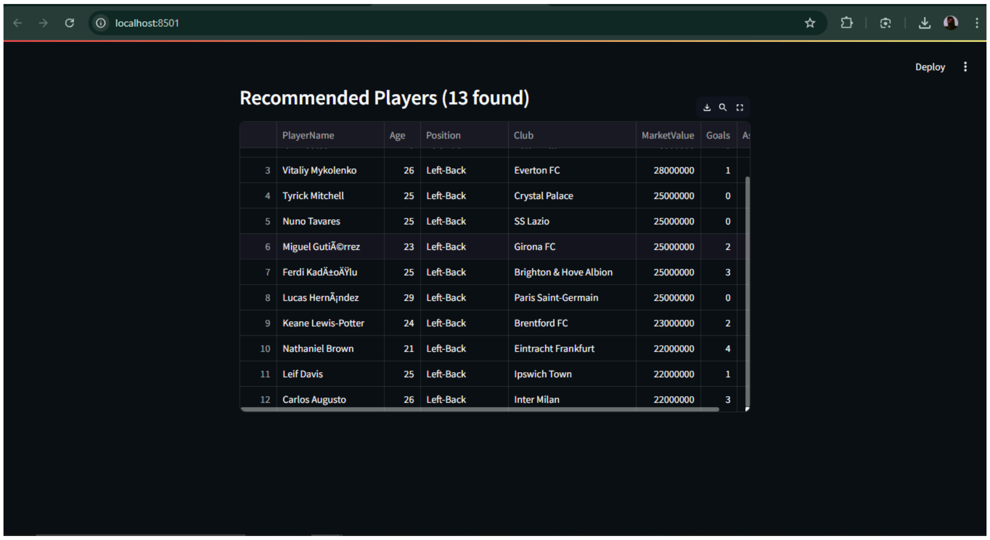
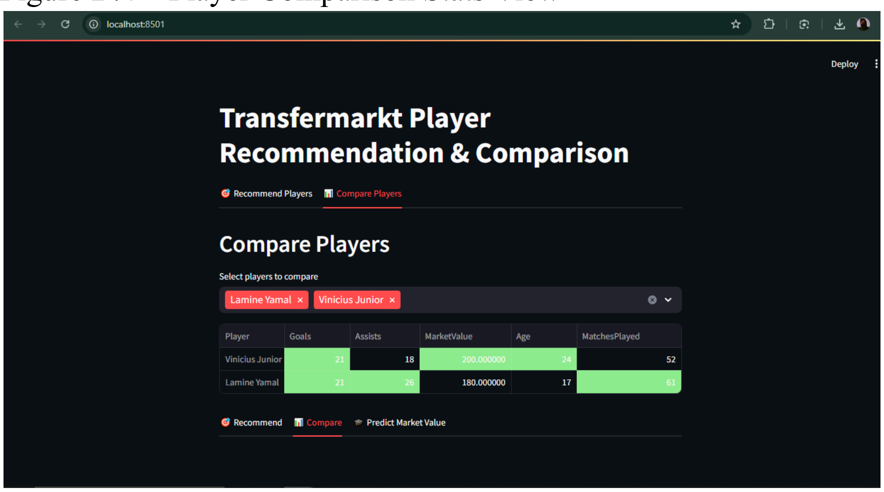


Figure B.4 – Player Comparison Stats View



Appendix : GitHub Repository

Repository URL: https://github.com/cherumanoj05/finalYearProject