



Corso di laurea in Data Science and Management

Cattedra di Data Visualization

# Analysis, Simulation, and Visualization of Football Player Transfers: A Case Study on Juventus FC

Prof. Blerina Sinimeri

---

RELATORE

Prof. Alessio Martino

---

CORRELATORE

Matr. Antonio Marco Vita

---

CANDIDATO

Anno Accademico 2022/2023

# Index

1. INTRODUCTION .....	2
1.1 CONTEXTUALISATION OF THE IMPORTANCE OF FOOTBALL PLAYER TRANSACTIONS .....	2
1.2 THE ACCUSATION OF THE CAPITAL GAINS CASE AGAINST JUVENTUS FC .....	3
1.3 PLAYERS TRANSFERS UNDER SCRUTINY .....	4
1.4 THESIS OBJECTIVES .....	5
2. ECONOMIC IMPORTANCE AND IMPACT OF THE FOOTBALL PLAYER TRANSACTION MARKET .....	5
2.1 OVERVIEW OF MACHINE LEARNING TECHNIQUES.....	10
2.2 MACHINE LEARNING IN THE CONTEXT OF THE FOOTBALL PLAYER MARKET .....	13
3. METHODOLOGY .....	17
3.1 OVERVIEW OF DATA USED IN THE RESEARCH: SOURCES, CHARACTERISTICS.....	17
3.1.1. UEFA TEAM RANKINGS (UEFARANKTEAM3.CSV): .....	17
3.1.2. SPI GLOBAL RANKINGS (SPI_GLOBAL_RANKINGS.CSV):.....	18
3.1.3. FOOTBALL DATA FROM TRANSFERMARKT (GAMES.CSV): .....	19
3.1.4 TRANSFERS.CSV .....	20
3.2 DATA CLEANING AND PRE-PROCESSING .....	21
3.3 EXCLUDED VARIABLES .....	27
3.4 FINAL DATASET EXPLANATION.....	28
3.5 SELECTION OF THE MACHINE LEARNING MODEL AND PERFORMANCE EVALUATION TECHNIQUES.....	30
4. ANALYSIS OF JUVENTUS TRANSFERS FEES CASE .....	33
4.1 APPLICATION OF THE MODEL TO ESTIMATE TRANSACTION PRICES .....	33
4.2 COMPARING PREDICTED PRICES WITH ACTUAL TRANSACTION PRICES.....	35
5. SUMMARY OF THE FINDINGS .....	39
5.1 APPLICATION OF THE TOOL FOR EVALUATION OF PLAYERS AND TEAMS .....	40
5.2 SUGGESTIONS FOR FUTURE DEVELOPMENTS AND POSSIBLE IMPROVEMENTS OF THE ALGORITHM .....	41
6. BIBLIOGRAPHY .....	<b>ERROR! BOOKMARK NOT DEFINED.</b>
7. SUMMARY .....	49

# 1. Introduction

## 1.1 Contextualisation of the Importance of Football Player Transactions

In the context of modern football, player transactions represent a key aspect for clubs and sports organisations. The acquisition of new talent is considered a strategic investment to improve the performance of teams and achieve sporting and financial goals. These transactions involve a complex network of players, including clubs, agents, intermediaries and sports organisations, creating a dynamic and competitive market.

Football clubs devote significant financial resources to secure the services of the most promising or already established players in the football landscape. However, the transaction pricing process is influenced by multiple factors, including the individual performance of the players, demand from purchasing clubs, age, experience, role on the pitch, popularity and growth potential of the player, as well as market and competitive environment considerations<sup>1</sup>.

An accurate assessment of a player's value is crucial for teams that is planning to buy or sell a player, as it can influence market strategy, budget and financial expectations. Moreover, agents and brokers could also benefit from a machine learning algorithm that can provide reliable estimates of transaction prices, enabling them to make more informed decisions and optimise their operations.

As a result, the ability to predict players' transaction prices has become a crucial element for clubs and insiders in the transfer decision-making process. The ability to make accurate predictions can provide a competitive advantage in the transfer market, enabling clubs to optimise financial resources and build competitive teams.

This Master's thesis in Data Science and Management aims to address this challenge by developing a machine learning algorithm for the prediction of player transaction prices<sup>2</sup>

---

<sup>1</sup> Transfer (association\_football) - Wikipedia

<sup>2</sup> Repository with full code: <https://github.com/marcovita33/FBFeesCalculator>

by using data from different sources. Several variables will be considered, including player characteristics (such as age and position), team performances (such as games won or rounds of competition reached) and other variables relevant to assessing the value of a player on the market.

## **1.2 The accusation of the capital gains case against Juventus FC**

The official announcement of the capital gains investigation by the competent authorities was a crucial moment in the case involving Juventus FC. The event attracted the attention of the media and public opinion, generating a wide discussion on the club's financial practices.

The start of the investigation took place on 26 November 2021, when the competent authorities made public their intention to closely examine Juventus FC's financial operations, particularly those related to capital gains. The specific allegations made against the club concern the conscious and repeated use of capital gains to manipulate financial statements without involving real money flows, thus constituting 'sporting unfairness'<sup>3</sup>.

In the case of the charges brought against Juventus in the so-called 'Capital Gains Case', the Turin Public Prosecutor's Office has contested multiple fraudulent conduct concerning the preparation of the financial statements for the years 2019-2020-2021. The allegation is that 'fictitious capital gains' corresponding to EUR 155 million were included in the financial statements. These capital gains were allegedly generated through fictitious valuations of owned players sold in exchange transactions with other clubs. It is alleged that the values attributed to the sold players were inflated in order to register an unjustified capital gain. This practice would have allowed Juventus to hide the huge losses that had characterised the previous corporate management, further aggravated by the economic effects of the pandemic.

This fraudulent conduct alleged by the public prosecutor's office is part of a broader context, in which several football clubs in recent years have sold players for exorbitant

---

<sup>3</sup> Plusvalenze fittizie e manipolazione del mercato: il caso Juventus F.C.- F. Maraschin

amounts of money in order to record capital gains in their balance sheets. This practice, known as 'administrative doping', has attracted the attention of Italian prosecutors. It is a phenomenon that is carried out mainly through the sale of young talents from the clubs' nurseries, exploiting their reduced value in the balance sheet in order to obtain a higher capital gain if they are sold.

The accusations have aroused great media interest and raised questions about compliance with financial rules in football. Juventus FC, being one of the most important and influential clubs in football, attracted particular attention.

### **1.3 Players transfers under scrutiny**

In the context of the Juventus FC capital gains case, several player exchanges have been under close scrutiny by the competent authorities. The thesis will focus on two specific exchanges that have been under scrutiny<sup>4</sup>.

One of the main exchanges considered was that between two midfielders Arthur and Pjanic<sup>5</sup>, which took place in 2020. This transaction raised questions about the valuation of the players involved and the actual monetary value of the transaction that seemed too high for those players. The authorities sought to determine whether the valuation of the players was accurate and whether the transaction was conducted in a fair and transparent manner.

Another trade that attracted the attention of the authorities was that between Canelo and Danilo<sup>6</sup>, which took place the previous year. In this case, too, the valuations and financial dynamics of the exchange were closely scrutinised. The objective was to verify whether the players' valuations were appropriate and whether the transaction was conducted in a transparent manner, in compliance with financial and regulatory rules.

The third was the "maxi-trade" with Genoa, in which Juventus acquired teenage midfielder Niccolò Rovella in return for Manolo Portanova and Elia Petrelli, two of their

---

<sup>4</sup> Plusvalenze Juventus: quali sono quelle sospette. I giocatori (e le squadre) coinvolti – Il Giorno

<sup>5</sup> Explaining Why Juventus Are So Willing To Swap Miralem Pjanic For Barcelona's Arthur Melo - Forbes

<sup>6</sup> Il senso dello scambio Danilo-Canelo - Ultimouomo.com

under-23 players. This trade has been referred as "scandalous"<sup>7</sup> because none of the players involved ever played for the team that acquired them, and the only one who did, Manolo Portanova, was given a six-year prison sentence for a major violation including non-consensual conduct after serving just one season.

## **1.4 Thesis objectives**

The objective of this project is to develop a Python tool that can accurately predict the transaction fees involved in football transfers. With the help of this tool, the capital gains associated with transfers in Juventus FC will be evaluated, as briefly introduced in Section 1.3. The thesis is structured as follows: Chapter 2 will provide the reader with theoretical background information on the football player transaction markets. It will specify the key actors and forces involved when teams acquire players, give a brief introduction to machine learning techniques, and provide examples of how machine learning has already been applied in similar cases. Chapter 3 will cover all the steps leading up to the development of the tool. This includes data gathering, cleaning, and preprocessing, as well as explaining the data choices. It will present the final dataset, provide data visualizations, and give an overview of the performance achieved using various methods. Chapter 4 will focus on applying the previous findings to the Juventus FC capital gain case and provide an interpretation of the results. Chapter 5 will suggest how the tool can be utilized in other situations and offer suggestions for future developments. Finally, Chapter 6, concludes this work, making an overview of the thesis' principal conclusions.

## **2. Economic importance and impact of the football player transaction market**

---

<sup>7</sup> Juve, Rovella è la madre di tutte le plusvalenze: le cifre e i valori 'scandalosi' di Portanova e Petrelli – calciomercato.com

The football player transaction market plays a role of significant economic importance in the context of the football industry. The sums involved in player transactions can reach astronomical values, with clubs investing large sums of money to acquire prominent talent. These transactions have a direct impact on the financial balance sheet of the teams involved, which face acquisition costs, high salaries and other financial burdens associated with players.

In addition, football player transactions affect the competitiveness of the teams themselves. A strategic acquisition can improve a team's overall performance, enabling it to compete at the highest level and achieve sporting success. On the other hand, a major divestment can lead to a reduction in a team's competitive strength, affecting its positioning in national and international competitions<sup>8</sup>.

The economic impact of football player transactions also extends beyond the playing field. A successful transaction can lead to increased revenue for the club involved. For example, the acquisition of an internationally renowned football player can attract new sponsors and investors, increase the sale of official club jerseys, generate more income from match ticket sales, and increase the team's media visibility. These factors contribute to a positive value cycle for the club, which can generate additional investment opportunities and financial growth.

Transaction prices of football players can be influenced those factors that reflect the complexity of the market. Understanding and evaluating the factors is essential to accurately predicting transaction prices and correctly assessing a player's value in the market. These factors interact with each other and the final price of a transaction is determined by the negotiation between the parties involved. Below are some of the main determinants that influence transaction prices in the football player market<sup>910</sup>:

1. Age and experience: A footballer's age and professional experience are key factors in determining his selling price. Young football players with high growth potential tend to

---

<sup>8</sup> Bhatia, Adil. Do Football Transfers Actually Improve Performance?. Bachelor's thesis, Harvard College. (2020)

<sup>9</sup> Determinants of football players' valuation: A systematic review  
- Wiley Online Library (2023)

<sup>10</sup> How do you value a player - Forbes

have higher prices because they are considered long-term investments. On the other hand, older footballers, even if they have extensive experience, may be transacted at lower prices because of the decline in their physical and technical abilities.

2. Current performance and skills: A footballer's current performance, such as goals scored, assists provided, defensive statistics, and overall performance on the field, have a significant impact on his selling price. Players who stand out for their technical, tactical and physical skills and who demonstrate consistency in performance usually command higher prices.

3. Popularity and marketability: A footballer's popularity, his ability to attract fans, sponsors, and media, and his marketability, which includes factors such as the number of jerseys sold, tickets sold, and followers on social media, directly influence his market value. Players with a strong media presence and a high degree of commercial appeal tend to have higher prices.

4. Role and way of playing: The role a footballer plays on the field and the way he plays influence his market value. Players who occupy strategic positions, such as strikers or creative midfielders, and who are distinguished by an attractive and disruptive style of play are often valued at higher prices than other roles.

5. Owning team and league: The home team and the league in which a player plays can influence his selling price. Teams at a high level or participating in international competitions tend to have players valued at higher prices. At the same time, leagues that are considered more competitive and prestigious, such as the Premier League or Serie A, can positively influence a player's value compared to lesser-known leagues.

6. Demanding team and league: The team interested in acquiring a player and the league in which it operates can influence the transaction price. Teams and leagues with higher financial resources and with the goal of reinforcing their team may be willing to invest larger sums in acquiring a talented footballer. In addition, the level of competitiveness of the prospecting league and its international visibility may affect the selling price. For example, a team from a top league might be willing to pay a higher price for a player from another, less competitive league in order to improve its sporting performance and global visibility.



7.Length of current contract: The remaining length of the player's contract with his current team can influence the transaction price. A longer contract could mean a higher sale price, as the team that owns the rights to the player has the opportunity to benefit from the player's performance for a longer period. Conversely, an expiring contract might push the transaction price down, as the current team might prefer to sell him rather than risk losing him for zero parameters.

8.Market demand: The overall demand for players at a given time and position affects transaction prices. If there are many teams interested in a specific footballer, competition among them may increase the selling price. Conversely, limited demand may lead to a reduction in price.

In the football player transaction market, prices are determined through complex mechanisms of supply, demand and bargaining. These mechanisms are at the heart of the process of determining a player's value and play a key role in negotiations between clubs and players<sup>11</sup>.

The determination of a player's price is influenced by the dynamics between supply and demand in the market. Supply represents the availability of a player in the market, that is, the willingness of the home club to dispose of the player. Demand, on the other hand, represents the buying clubs' interest in acquiring the player. The interaction between supply and demand creates a balance that determines the selling price.

Negotiations and bargaining play a crucial role in the pricing process. During negotiations, the parties involved, namely the clubs and the players' agents, try to reach an agreement on the selling price.

The concepts of "transfers" and "negotiations" represent two key aspects that deserve in-depth analysis. While they are terms often used interchangeably, it is important to understand their differences and the implications they have in the context of player trades.

A transfer occurs when a player moves from one team to another<sup>12</sup>. It can involve a variety of situations, such as moving from one club team to another, transferring from one league to another, or changing teams within the same league. During a transfer, an agreement

---

<sup>11</sup> The valuation of human capital in the football player transfer market – van den Berg, Erasmus University (2011)

<sup>12</sup> Transfer (association\_football) - Wikipedia

must be reached between the club that holds the rights to the player (source club) and the club that wishes to purchase the player (destination club). This agreement may cover the player's selling price, contract terms, termination clauses, brokerage fees, and other aspects related to the transaction. In some cases, transfers may also involve a player who is free of contractual obligations, allowing for greater flexibility in negotiations.

On the other hand, negotiations represent the process of discussing and agreeing on the details of a transfer<sup>13</sup>. During negotiations, the parties involved, namely the home and destination clubs, the players themselves, and any agents or intermediaries, discuss the financial, contractual, and logistical aspects of the transaction. This process can be complex and require a series of meetings, discussions and exchanges of proposals to reach an agreement satisfactory to all parties involved. Negotiations may involve aspects such as the player's sale price, contract length, bonuses, termination clauses, and other conditions that may affect the overall value of the transaction.

While transfers represent the physical movement of a player from one team to another, negotiations represent the process of negotiating and agreeing on the details of that transfer. It is important to understand these distinctions because the success of a transfer depends largely on the ability of the parties involved to conduct effective negotiations and reach an agreement that meets their interests and needs.

As sports evolved new figures came out, agents, also known as players' agents or procurers, act as representatives of players, offering support and advice during the transfer process. Agents and intermediaries play a crucial role in facilitating negotiations and transactions between clubs and players<sup>14</sup>.

The role of agents in football is very extensive and covers a range of activities, including researching transfer opportunities for players, negotiating contracts with clubs, managing the financial aspects of transactions, and mediating between the parties involved. Agents work closely with players to understand their preferences, career goals, and financial expectations, seeking to best meet their interests.

Intermediaries, on the other hand, play a similar role to agents but in a more limited way. Agents may be involved in negotiations, offering advisory services and support in

---

<sup>13</sup> The art of negotiation – Sky Sports

<sup>14</sup> Role of agents and intermediaries in sports - iPleaders

managing negotiations. However, unlike agents, intermediaries do not exclusively represent players and may be involved in multiple transactions and with different parties.

Agents and intermediaries have an extensive network of contacts in the football world, including clubs, managers, coaches, and other agents. This network gives them access to insider information about the transfer market, potential career opportunities for players, and the needs of clubs. By leveraging this network, agents and brokers can identify the best transfer opportunities for players and try to maximize their market value.

However, the role of agents and intermediaries in football has been the subject of debate and criticism. It's debated that their influence in the transfer process could lead to market distortions, influencing transaction prices and creating conflict of interest situations. Therefore, it is important that agents and intermediaries act transparently, ethically and in compliance with the rules and regulations set by football organizations.

## **2.1 Overview of machine learning techniques**

Over the years, the application of machine learning in every domain of human activity has gained increasing interest and relevance.

Machine learning represents a computational approach that allows systems to automatically learn from past data and adapt to new situations, it's wide range of application make possible to analyse complex data, identify patterns or making predictions or decisions without being explicitly programmed.

In machine learning there are two main approaches, Supervised and Unsupervised learning<sup>15</sup>.

Supervised approaches are based on the use of labeled training data, used to train predictive models that can estimate values or classify data. During the training phase, the model tries to learn the relationship between the variables. Among the common algorithms used for transaction price prediction, can be mentioned<sup>16</sup>:

---

<sup>15</sup> Introduction to Machine Learning, fourth edition. Adaptive Computation and Machine Learning series - Ethem Alpaydin - MIT Press (2020)

<sup>16</sup> Shalev-Shwartz, S., & Ben-David, S. (2014). Understanding Machine Learning: From Theory to Algorithms. Cambridge: Cambridge University Press

Linear regression: The linear regression algorithm is one of the simplest and most popular in the field of price prediction. It is based on the assumption that there is a linear relationship between the independent variables  $x$  and dependent variable  $y$ , which is always a quantitative value. The linear regression model estimates regression coefficients representing the effect of each independent variable on dependent one. Linear regression can be used for demand forecasting, sales prediction, and resource allocation. It's used in field like social sciences, economics and finance.

Logistic Regression: Logistic regression is used for binary classification problems where the mission is to predict one of two classes. It models the relationship between the input features and the probability of belonging to a particular class using a logistic function, in fact here  $y$  variable is categorical or qualitative. Apart from binary classification tasks, logistic regression can be used for multi-class classification problems, the model is widely used in fields such as healthcare, marketing, and fraud detection.

Support Vector Machines (SVM): SVMs are classification and regression algorithms that try to find an optimal hyperplane that separates the data into different classes or estimates the relationship between the independent variables and the dependent one. The goal is to find the maximum margin that separates the data and, consequently, estimate the dependant variable. SVMs can be used for anomaly detection, fraud detection, and image classification.

Decision trees: Decision trees are machine learning algorithms that use a tree structure to make decisions. They recursively partition the data based on feature values to make predictions. Decision trees are interpretable and can handle both numerical and categorical data. The model can be utilized for recommendation systems, customer profiling, and predictive maintenance.

Random Forest: The random forest is a machine learning algorithm that combines the output of several decision trees to obtain a more accurate final prediction. Each tree is trained on a different subset of the data and features, and the final prediction is made by aggregating the predictions of individual trees. It is widely used in areas such as healthcare, finance, and customer churn prediction.

Gradient Boosting: Gradient boosting algorithms are powerful aggregate methods that combine multiple predictive models. These algorithms iteratively build a series of models, each correcting the mistakes of the previous model. Gradient boosting algorithms have

achieved good performance in various domains, including web search ranking, recommendation systems, and anomaly detection.

**Neural networks:** Neural networks are computational models inspired by the workings of the human brain. They consist of interconnected layers of artificial neurons, enabling them to learn complex patterns and representations. Neural networks can be used for various tasks, including image recognition, natural language processing, time series analysis and autonomous vehicles.

On the other hand, unsupervised approaches rely on the use of unlabeled. These approaches try to identify patterns, relationships or groupings within the data without external supervision. Unsupervised algorithms explore intrinsic features of the data to identify hidden or similar structures among data. Unsupervised learning algorithms aim to cluster similar data points together, identify anomalies or outliers, reduce the dimensionality of the data, or generate new samples that resemble the original data. It's used for exploratory analysis, data mining, and finding hidden structures in the absence of labeled data. Some examples of unsupervised algorithms used include.

**K-means Clustering:** K-means clustering is a method that groups data points into clusters based on their similarities. It works by defining a predetermined number of clusters, represented by 'K'. The algorithm assigns each data point to the cluster with the nearest mean (centroid) and updates the centroids iteratively. The process continues until the assignments no longer change significantly. K-means clustering helps in finding groups or patterns in data by minimizing the distance between data points within each cluster and can be used for customer segmentation, market segmentation, and document clustering.

**Hierarchical Clustering:** Hierarchical clustering is a method that creates a hierarchy of clusters by merging or splitting them based on their similarities. It starts by considering each data point as a separate cluster and then successively merges or divides clusters based on their similarities. This process forms a tree-like structure called a dendrogram, which allows us to visualize the relationships between clusters at different levels of granularity. Hierarchical clustering helps in identifying clusters and subclusters in a dataset and it's mostly used for taxonomic classification and market segmentation.

**Principal Component Analysis (PCA):** PCA is a technique used for dimensionality reduction. It aims to transform a high-dimensional dataset into a lower-dimensional space while preserving the most important information. PCA identifies the directions (principal

components) along which the data varies the most. These principal components are orthogonal to each other, and each component captures a certain amount of variance in the data. By projecting the data onto a subset of these principal components, PCA helps in reducing the complexity of the data while retaining the essential patterns and relationships. It's used in image compression and feature extraction.

## **2.2 Machine Learning in the context of the football player market**

Machine learning has opened up new perspectives in the context of the football player market, enabling clubs and analysts to leverage advanced algorithms to predict transaction prices with greater accuracy. This computational approach has proven to be particularly effective in analysing vast amounts of heterogeneous data and discovering hidden patterns and relationships.

Literature review plays a key role in the context of football player price prediction. Exploring previous research and the methodologies used allows us to gain an in-depth overview of the trends, approaches, and challenges present in this area.

In recent years, numerous studies and research have been conducted to understand and predict transaction prices in the football player market. Efforts by the scientific community and industry experts have helped to provide an in-depth overview of the factors influencing prices and have led to the development of increasingly sophisticated predictive models.

One of the main areas of research concerns the identification of key factors influencing football player prices. Studies have investigated a wide range of variables, including a player's age, past performance, popularity, role and position on the field, the league in which the player plays, and the team from which the player comes. The goal has been to identify the most significant variables and develop models that take these factors into account to accurately predict transaction prices.

Some research has focused on the use of econometric models for player valuation<sup>17</sup>. These models rely on economic and statistical theories to estimate transaction prices. Variables

---

<sup>17</sup> Poli R, Besson R, Ravenel L. Econometric Approach to Assessing the Transfer Fees and Values of Professional Football Players. *Economies*. (2022)

considered include teams' financial data, such as budgets and revenues, as well as the performance of the players themselves. However, these traditional approaches may be limited in capturing the complexity of the football player market, which is also influenced by emotional and market factors.

Another line of research has focused on using machine learning and artificial intelligence techniques for football player price prediction. These approaches are based on the analysis of large amounts of data, including statistical data on players, market data, performance information, and data from social media<sup>18</sup>. Machine learning algorithms are trained to identify hidden patterns in the data and develop accurate predictive models.

A real-world example of the application of machine learning in the football player market is the use of predictive models based on regression algorithms to estimate the value of a football player. These models take into consideration a number of variables, such as the player's individual performance, match statistics, demographic characteristics, scouts' evaluations, and more. Through the analysis of these factors, the models can provide an estimate of a player's selling price, thus supporting clubs' buying or selling decisions<sup>19,20</sup>.

Another concrete example is the use of unsupervised clustering algorithms to identify groupings of players with similar characteristics. These algorithms analyse available data, such as past performance, technical and tactical characteristics, age, and other attributes, in order to identify homogeneous groups of players. This analysis can help clubs identify desired player profiles or identify emerging talent to monitor<sup>21</sup>.

Another example involves the use of neural networks to predict transaction prices. Neural networks are machine learning algorithms that emulate how the human brain works and are able to learn from complex data. These networks can analyse historical data on players' transactions, considering variables such as age, performance, playing position, experience and other characteristics. By training the network on a large data set,

---

<sup>18</sup> Užík M, Warias R, Glova J. Management of Transfer Prices in Professional Football as a Function of Fan Numbers. *Mathematics*. (2022)

<sup>19</sup> <https://www.footballtransfers.com/en/values/players>

<sup>20</sup> [https://www.footballbenchmark.com/methodology/player\\_valuation](https://www.footballbenchmark.com/methodology/player_valuation)

<sup>21</sup> A robust method for clustering football players with mixed attributes - D'Urso, De Giovanni, Vitale – Springer (2022)

predictive models can be obtained that can estimate future transaction prices with good accuracy<sup>22</sup>.

To predict transaction prices in the football player market, both traditional and innovative approaches have been used.

Traditional approaches in modelling transaction prices often rely on econometric models, such as linear regression and time series models. These models seek to identify causal relationships between transaction prices and a number of explanatory variables, such as a player's age, past performance, position on the field, league affiliation, and other relevant factors. Such models can provide a good understanding of the factors that influence transaction prices, but they may not be able to capture the complexity of the data and nonlinear patterns.

On the other hand, innovative approaches in transaction price modelling use machine learning and artificial intelligence techniques to analyse large amounts of data and recognize complex patterns<sup>23</sup>. Some of the commonly employed algorithms include artificial neural networks, decision trees, support vector machines, and deep learning models such as convolutional and recurrent neural networks. These algorithms can learn autonomously from the data and adapt to the nonlinear and multidimensional information found in the football player market<sup>24</sup>.

One innovative approach that has gained popularity is neural network-based modelling<sup>25</sup>. Neural networks can be trained to recognize patterns in player characteristics and generate transaction price predictions based on those patterns. The use of deep neural networks makes it possible to model complex relationships between variables and capture hidden information in the data.

---

<sup>22</sup> Pricing Football Players using Neural Networks – Dey, University of Southern California (2017)

<sup>23</sup> <https://github.com/sanjitva/Predicting-Football-Player-Transfer-Values>

<sup>24</sup> PlayeRank: Data-driven Performance Evaluation and Player Ranking in Soccer via a Machine Learning Approach - Luca Pappalardo, Paolo Cintia, Paolo Ferragina, Emanuele Massucco, Dino Pedreschi, and Fosca Giannotti. 2019.

<sup>25</sup> Pricing Football Players using Neural Networks – Dey, University of Southern California (2017)



It is important to note that traditional and innovative approaches in transaction price modelling do not necessarily compete with each other, but can be combined to achieve better results. For example, using econometric models to identify the most relevant variables and then applying machine learning techniques to improve the accuracy of predictions. The ideal approach might be to combine the strengths of both approaches. Integrating traditional approaches with machine learning techniques can result in a completer and more robust model. To obtain the best results one way could be starting with a linear regression model to identify the most significant variables and then use a machine learning algorithm to refine and improve the predictions. In addition, using model ensembles, which combine different machine learning techniques, can provide even better results.

Despite the significant progress made in the application of machine learning techniques for transaction price prediction in the football player market, there are still some limitations and challenges that need to be addressed. It is important to carefully consider these issues in order to achieve accurate and reliable predictions.

One of the main limitations is the availability and quality of data. While there are many sources of data in the football industry, such as [wyscout.com](https://www.wyscout.com), [soccerment.com](https://www.soccerment.com), and [driblab.com](https://www.driblab.com), the quality and coverage of data can vary widely. It is critical to have access to complete, accurate, and up-to-date data to feed machine learning models. In addition, some crucial information may not be readily available or available, such as contract details and exact transaction figures. The lack of sufficient and reliable data may affect the accuracy of predictions.

Another challenge concerns the dynamic nature of the football player market. Market conditions, club strategies, and agent behaviours can change rapidly over time. This requires proper management of the "stationary station" concept in the prediction process. Market fluctuations, transfer trends and changes in teams must be taken into account to achieve more accurate and up-to-date predictions.

In addition, the complexity and interconnectedness of the variables involved in determining transaction prices can be a challenge. For example, transaction prices can be influenced by factors beyond statistics and player characteristics, such as market demand, club interest, and negotiation between the parties involved. Effectively integrating these

complex variables into machine learning models requires careful consideration and in-depth analysis.

Finally, another limitation concerns the interpretation of the results obtained from machine learning models. Some algorithms, such as neural networks, can be considered "black boxes" in that it is not always clear how they arrive at a given prediction. This can raise concerns in terms of transparency and interpretability of the models. It is important to develop methods to understand and explain the decision-making processes of machine learning models to ensure greater confidence in the predictions obtained.

Addressing these limitations and challenges requires a holistic approach that integrates domain knowledge, statistical methods, and machine learning techniques. It is critical to develop robust models that are able to adapt to market changes and account for the complexity of factors involved.

### **3. Methodology**

#### **3.1 Overview of data used in the research: sources, characteristics**

The research is based on analysing the performance of football teams using different datasets. A detailed description of the main datasets used to conduct the study is provided below.

##### **3.1.1. UEFA Team Rankings (uefarankteam3.csv):**

The dataset "uefarankteam3.csv" is a reliable source for evaluating the ranking of teams according to the UEFA system. UEFA scores are an important indicator of teams' performance in European competitions and contains the results achieved over the past 10 years.

UEFA scores reflect the strength and consistency of teams in the context of UEFA competitions<sup>26</sup>. Each team receives points based on its performance in matches played in tournaments such as the UEFA Champions League, UEFA Europa League, and UEFA Super Cup.

The UEFA points are counted according to a complex formula that takes various factors into consideration. Scores are awarded not only according to the final result of matches (win, draw or loss), but also according to the round of the competition in which it was achieved, the coefficient of the opposing team and the coefficient of the team's home country.

For example, a win in the group stage of the UEFA Champions League may be worth more points than a win in a later stage such as the quarterfinals. Similarly, defeating a team with a high UEFA coefficient leads to a greater gain in points than a win against a team with a lower coefficient.

UEFA points are accumulated over the years and are used to calculate the ranking of teams in the UEFA system. This ranking represents a hierarchical scale of European teams based on their performance at the continental level.

The dataset "uefarankteam3.csv" was obtained through a web scraping process directly from UEFA's official website<sup>27</sup> using the Python library BeautifulSoup. This process made it possible to collect data regarding UEFA team scores in an automated way and make it available for analysis and evaluation of the performance of football teams in the context of European competitions.

### **3.1.2. SPI Global Rankings (spi\_global\_rankings.csv):**

The "spi\_global\_rankings.csv" dataset contains the SPI (Soccer Power Index)<sup>28</sup> scores of teams. SPI ratings represent an overall measure of a team's strength in football, taking into account both offensive and defensive aspects of the game.

---

<sup>26</sup> How association club coefficients are calculated - UEFA

<sup>27</sup> <https://www.uefa.com/nationalassociations/uefarankings/tenyears/>

<sup>28</sup> <https://projects.fivethirtyeight.com/soccer-predictions/>

The SPI rating is a rating system that uses a number of factors to determine the relative strength of one team compared to others. These factors include the team's past performance, match results, quality of opponents faced, the number of goals scored and conceded, position in the standings, the level of competition the team is in, and other relevant parameters.

Regarding the composition of the dataset, the columns contained are named "team name", "off", "def" and "spi".

The "off" column represents a team's offensive strength index and reflects the team's ability to score goals and create effective offensive actions. This score takes into account various factors, such as the number of goals scored, passing accuracy, attacking actions, and other metrics related to the offensive aspect of the game.

The "def" column indicates a team's defensive strength index, which reflects a team's ability to defend and prevent goals from opponents. This score considers factors such as the number of goals conceded, goalkeeper save rate, defensive actions, and other metrics related to the defensive aspect of the game.

Finally, the "spi" column represents a team's overall strength index, which takes into account both the offensive and defensive aspects. This aggregate score provides an overall assessment of the team and represents an estimate of its relative strength compared to other teams in the dataset.

### **3.1.3. Football Data from Transfermarkt (games.csv):**

The dataset "games.csv" provides a source of information about football games scraped from Transfermarkt website<sup>29</sup>. It contains a large set of data related to games, including team names, results, goals scored and conceded, and other related information. This dataset covers a significant period since 2012 and includes more than 60,000 of the major national and international football competitions.

---

<sup>29</sup> <https://data.world/dcereijo/player-scores>

Transfermarkt is a well-known online portal dedicated to football that provides a vast amount of data and information on teams, players, games, and competitions. The "games.csv" dataset was obtained from Transfermarkt and contains data collected and organized in a structured manner to facilitate the analysis of team performance and evaluation of strength balance.

The competitions contained in the dataset may vary over the years, but generally include the major national leagues, such as the English Premier League, Italian Serie A, Spanish La Liga, German Bundesliga, and French Ligue 1. In addition, there may also be international competitions such as the UEFA Champions League, Europa League, and other national and continental competitions.

### **3.1.4 Transfers.csv**

The "transfers" dataset contains detailed information on player transactions within European football clubs, as reported on the Transfermarkt website<sup>30</sup>, since the 1992/93 season. This dataset is of fundamental importance for making the predictions regarding player transaction prices.

Included within the dataset is data concerning the major European leagues, such as the English Premier League, French Ligue 1, German Bundesliga, Italian Serie A, Spanish La Liga, Portuguese Liga NOS, Dutch Eredivisie, and Russian Premier Liga. Each league has a separate file in .csv format, allowing for competition-specific analysis.

Common variables in the dataset provide detailed information on players' transactions. Some of these variables include:

- club\_name: the name of the football club involved in the transaction. This variable identifies the club that bought or sold the football player.
- player\_name: the name of the football player involved in the transaction. This variable identifies the soccer player involved in the transaction.

---

<sup>30</sup> <https://github.com/ewenme/transfers>

- position: the position on the field of the kicker involved in the transaction. This variable specifies the role or position of the kicker within the team.
- club\_involved\_name: the name of the second club involved in the transaction. This variable identifies the club that was involved in the transaction along with the main club.
- fee: raw information about the amount of money in the transaction. This variable provides details about the amount of money involved in the player's transaction.
- transfer\_movement: indicates whether the transaction represents an entry or exit from the club. This variable specifies whether the football player was bought or sold by the club in question.
- transfer\_period: indicates whether the transaction occurred during the summer or winter window. This variable identifies the time period in which the transaction took place.
- fee\_cleaned: the transaction figure transformed into a numerical format, expressed in millions of euros. This variable represents the transaction figure in a standardized form to facilitate analysis.
- league\_name: the name of the league to which the club belongs. This variable specifies the competition league to which the club involved in the transaction belongs.
- year: the year in which the transaction occurred. This variable specifies the year in which the player's transaction occurred.
- season: the season of the transaction, obtained by interpolation from year. This variable represents the football season corresponding to the year of the transaction.
- country: the country in which the league is played. This variable identifies the country in which the competition league in which the club is active is located.

### **3.2 Data cleaning and pre-processing**

As part of this research, an in-depth study was carried out using the "uefarankteam3" and "spi\_ratings" datasets in order to calculate the "Overall" score for each football team. The 'Overall' scores made it possible to assess the overall strength and capabilities of the

teams, considering several variables such as UEFA scores, SPI ratings and other performance measures such as Offence and Defence.

To begin the analysis, the necessary libraries such as pandas, fuzzywuzzy and sklearn.preprocessing were imported, which allowed the manipulation and comparison of the data, as well as the application of value scaling techniques.

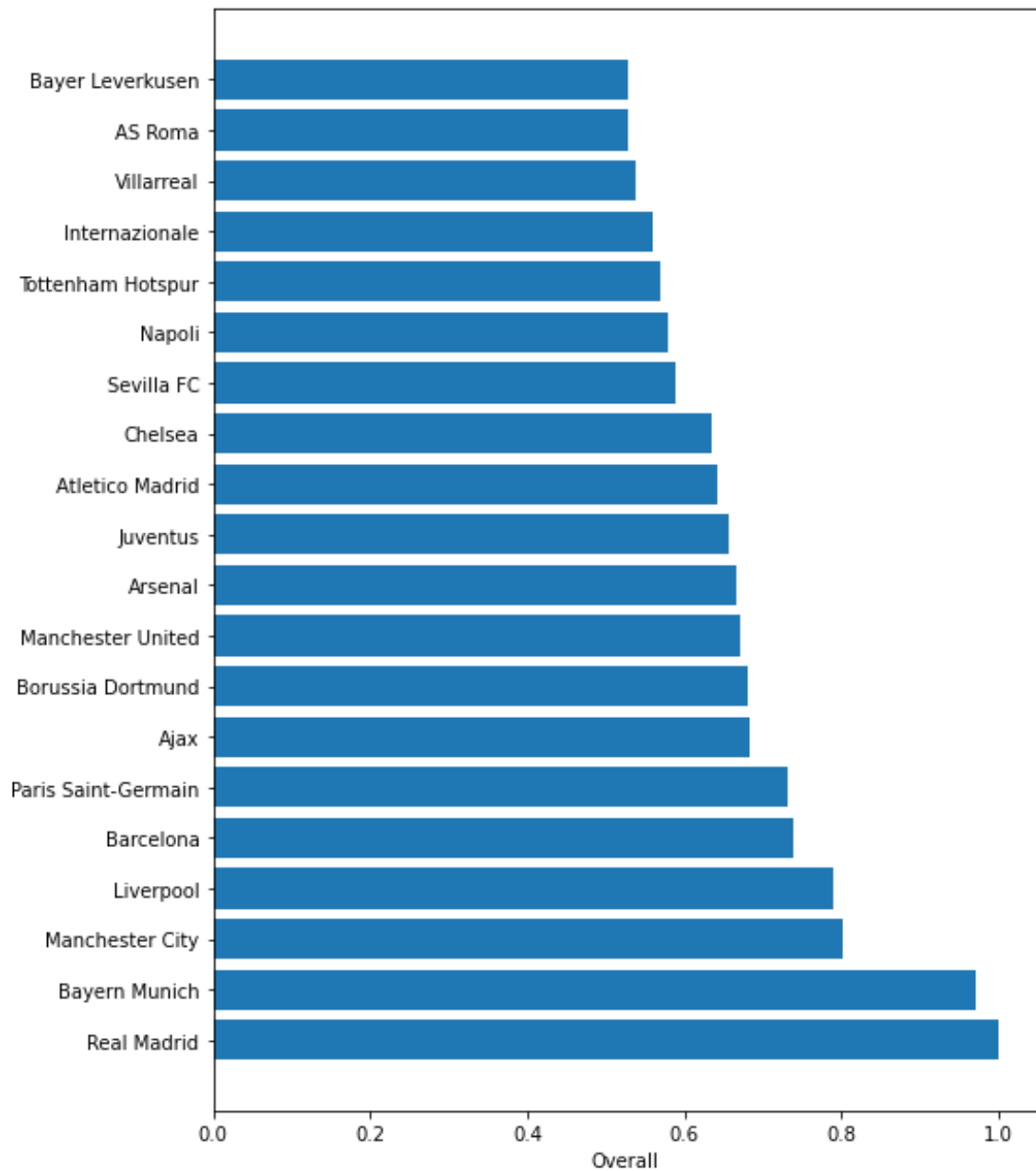
Next, a matching between the team names in the different datasets was carried out, using the 'extractOne' function of the fuzzywuzzy module. This process made it possible to compare the team names and find matches with a degree of similarity greater than or equal to 95%. The results were saved in a new dataframe.

After matching the teams, a filter was applied to the previous dataframe to select only rows related to the countries of interest, defined within a list called 'desired\_countries'. This filter made it possible to focus the analysis on the teams of the leagues of interest, excluding leagues not present in the transfer dataset.

Next, the normalisation of the scores of each variable in the dataframe was performed, using sklearn's StandardScaler tool. This operation was necessary to make the data comparable and reduce the impact of scale differences between the variables considered.

To assign a different weight to each variable, a dictionary called 'weights' was used. The weights were applied when calculating the 'Overall' score for each team in the dataframe. This score represents an overall measure of team performance, obtained by combining the scores of the variables with the corresponding weights, the formula used was  $(0.2 \text{ Att} + 0.2 \text{ Def} + 0.2 \text{ SPI} + 0.4 \text{ UefaPoints})$ .

Then, the 'Overall' score was normalised to vary between 0 and 1 in order to facilitate the comparison between teams.

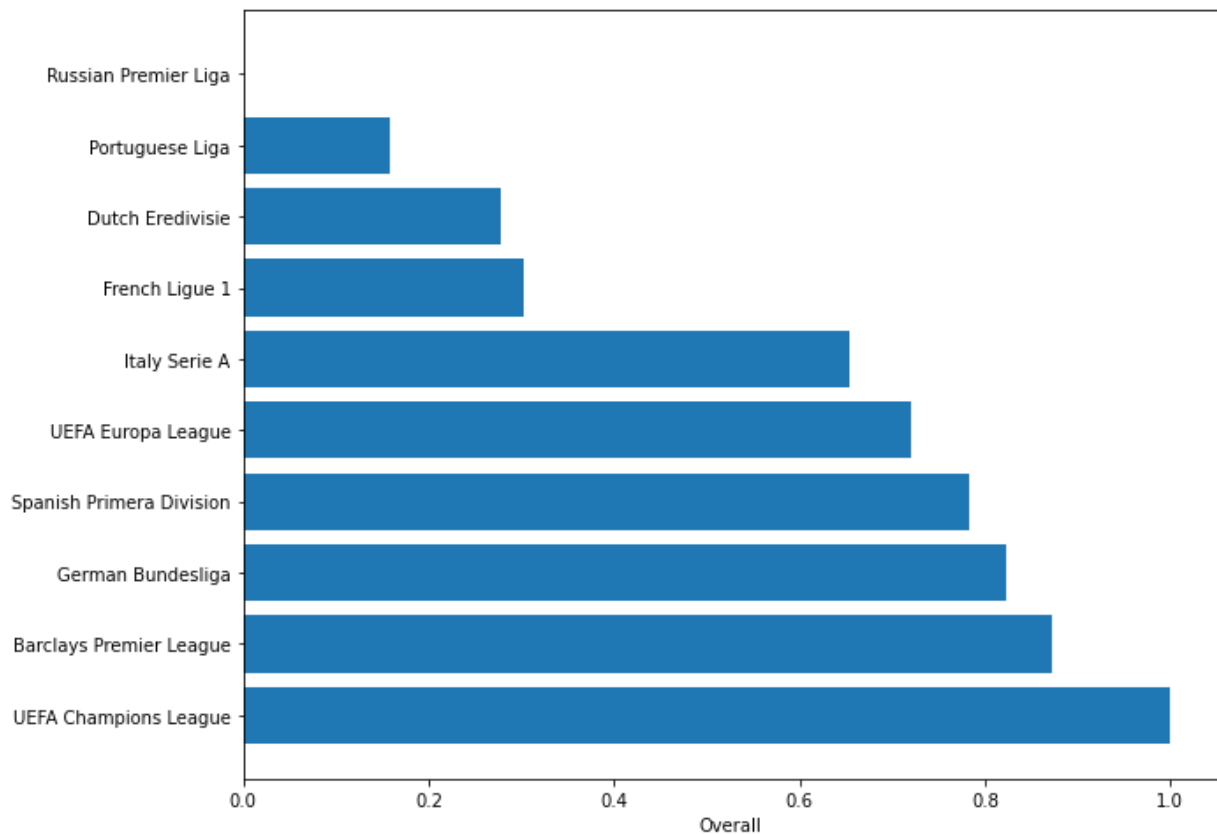


*Figure 1 – Teams sorted by their Overall*

The dataframe was then grouped by the 'league' column, and the sum of the 'Overall' scores was calculated for each league. Then, the average of the 'Overall' scores for each league was calculated. This made it possible to assess the average performance of the teams within each league.

The league scores were then standardised according to their average 'Overall' scores.





*Figure 2 – Leagues sorted by their Overall*

In order to further integrate the data, a new dataset was created through a merge with the sheet games.csv, which contains lot of scores and matches. By operating in that dataset there was obtained the average 'Overall' scores for each team in each year by summing and standardizing every “Overall Match” value of a team for every season. This makes it possible to track the performance of the teams over time and provide a clear view of performance. There is a list of the best 20 teams with their best seasonal result, according to the calculations.

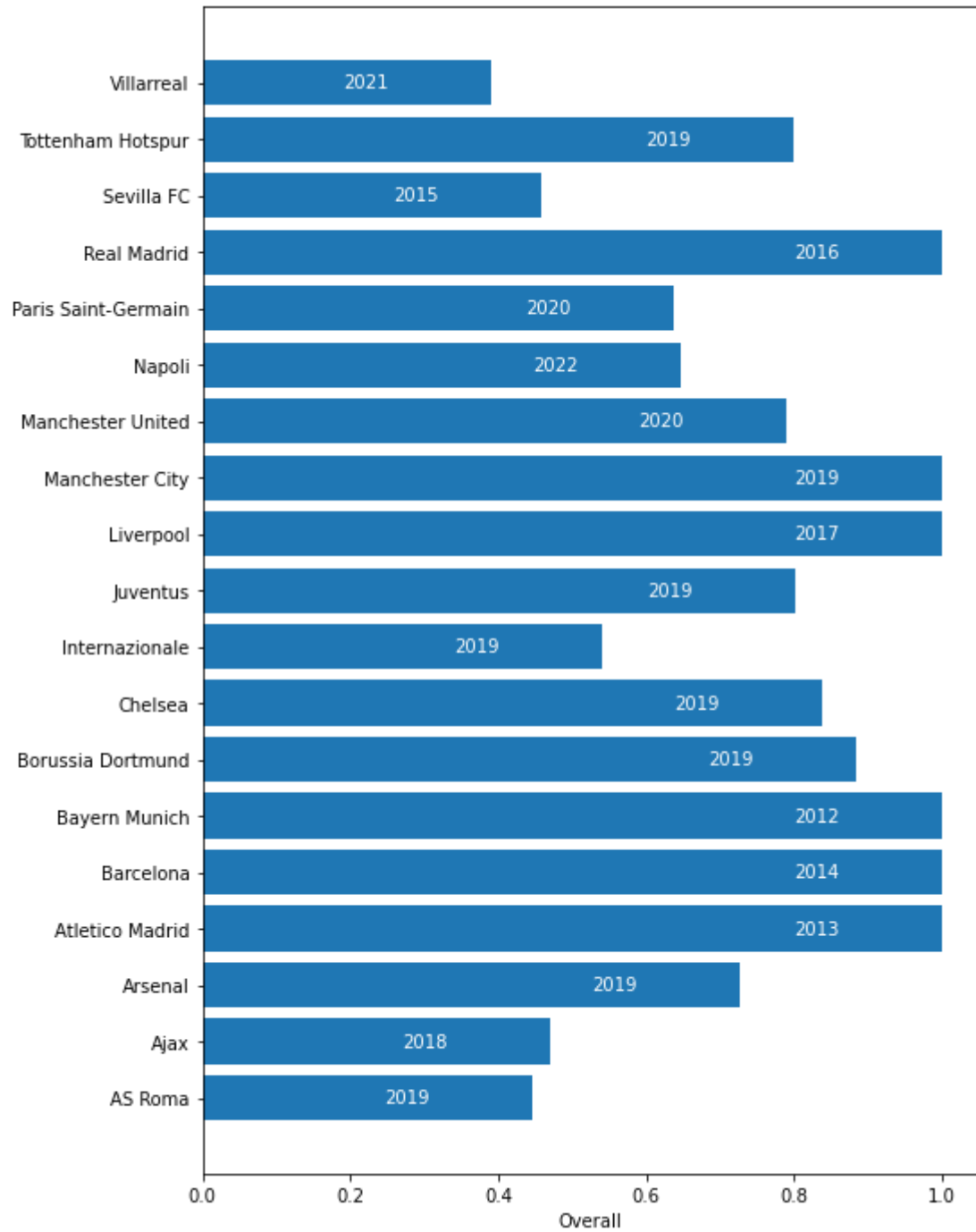


Figure 3 - Best seasonal result in the last 10 years

Next, a merge was performed between the transfers dataset and the previous dataframe. This merge made it possible to add match results to the transfer dataset, creating a single

data source that combines information on player transactions with team performance.

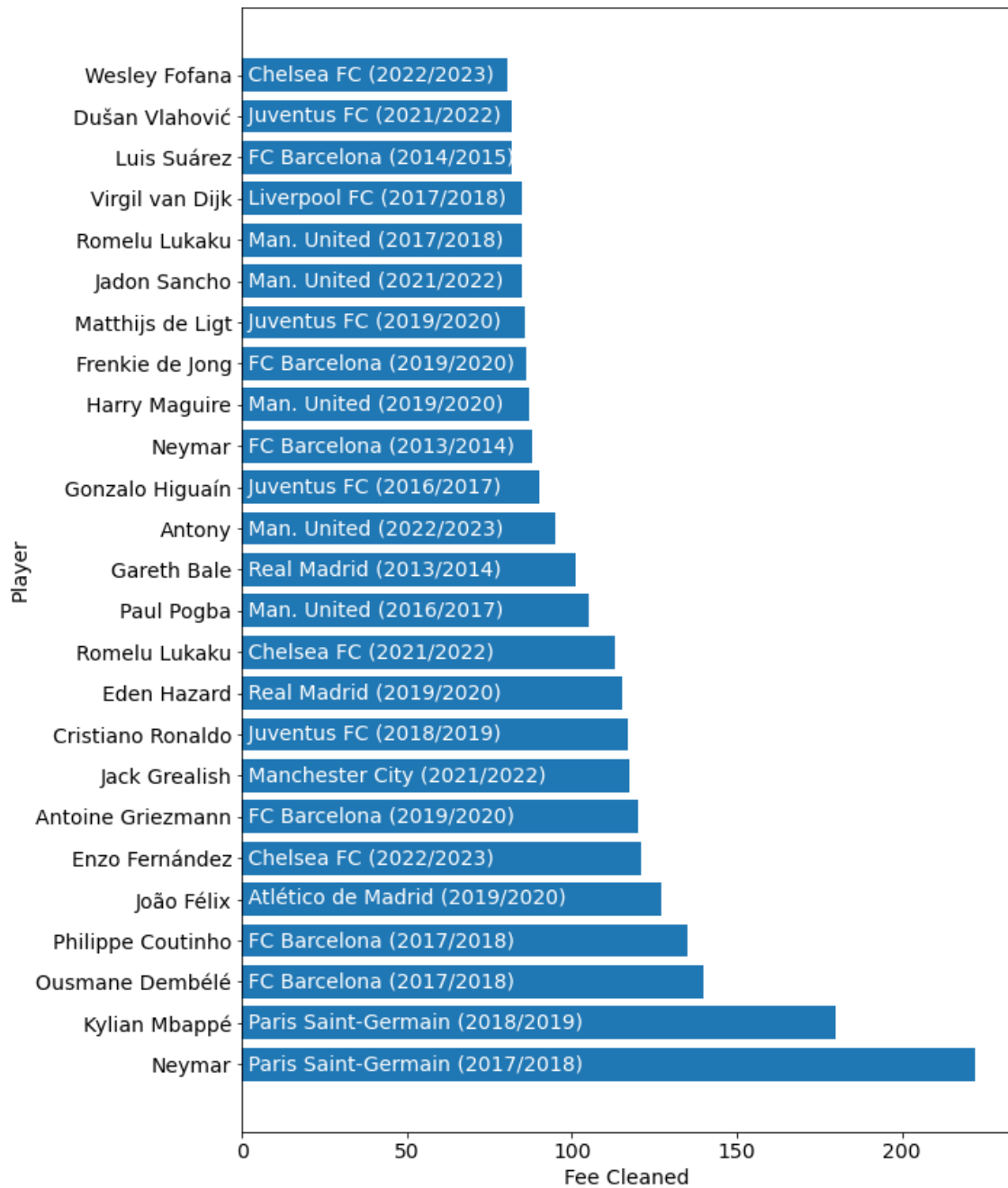


Figure 4 – Top 25 expensivest transfers in yhe last 10 years

In conclusion, the methodology adopted in the research involved the use of several datasets to analyse the performance of football teams. Well-structured steps were applied, such as matching team names, normalising the data, assigning weights to variables and calculating 'Overall' scores to assess the overall performance of the teams.

### 3.3 Excluded variables

Some potential variables, easily available on the internet such as FIFA overall, Football Manager potential ability, games played, goals, assists and contrasts, were excluded from the study for specific reasons.

FIFA's overall and Football Manager's potential ability are two metrics used in the respective football simulation video games to evaluate footballers' abilities<sup>3132</sup>.

FIFA's overall<sup>33</sup> represents an overall numerical evaluation of a footballer's abilities based on various characteristics, such as speed, shooting, passing, dribbling and defence. It is a score that indicates the overall skill level of a footballer within the FIFA game.

Football Manager's potential ability<sup>34</sup>, on the other hand, indicates the growth potential of a footballer over the course of his career. It represents an estimate of the maximum skills a player could achieve over time, considering his technical, physical and mental attributes.

Both metrics are used to assess the performance of players within their respective games, but it is important to underline that they have no direct impact on valuations or transfer prices in the reality of professional football. The decision to exclude these variables was based on several considerations. Firstly, the relevance of the data was carefully assessed and it was considered that these variables might not be directly significant for the specific objective of predicting transaction prices.

Within the academic context, several studies have been conducted to analyse the performance of football players and the dynamics of the transaction market. However, some of these studies may have introduced biases into their analyses. For instance, one study focused on the subjective overall ratings of footballers according to the FIFA game. However, these scores are influenced by factors not strictly related to on-field

---

<sup>31</sup> Ian G. McHale, Benjamin Holmes. Estimating transfer fees of professional footballers using advanced performance metrics and machine learning, - European Journal of Operational Research

<sup>32</sup> Using machine learning to identify high-value football transfer targets, Tattersal. – Towards data Science

<sup>33</sup> FIFA player ratings explained: How are the card number & stats decided? – Goal.com

<sup>34</sup> Current & Potential Ability – FMINSIDE.net

performance, potentially creating a bias in the overall assessment of the footballers. Furthermore, some studies have focused exclusively on specific factors such as goals scored or number of matches played, neglecting other relevant aspects such as defensive performance or the specific role of the players in the team. The introduction of such biases could limit the completeness and accuracy of the analyses conducted, thus compromising the validity of the predictions regarding the players' transaction prices.

The choice made was to focus on variables more closely related to the dynamics of the transfer market and the factors influencing prices. Metrics such as SPI scores, UEFA scores and other player performance measures that are directly related to the financial and valuation aspect of transactions were considered. In addition, other aspects of players' performance or market dynamics were considered, as the excluded variables were not central to the analysis conducted. Attention was paid, for example, to team evaluation scores, financial aspects or market dynamics, which required the use of variables more relevant to these issues. The careful selection of variables was based on the specific objective of the thesis and their relevance in the context of players' transactions.

### **3.4 Final dataset explanation**

The cleaning, merging and maintenance of past datasets offers a wide range of information related to players' transactions, covering aspects such as the clubs involved, the players themselves, financial details and the performance of the teams.

Explanation of the variables and the information they represent<sup>35</sup>:

- club\_name: Represents the club that carried out the transaction. This information is crucial to identify which club acquired the player and influenced the value of the transaction.

- player\_name: Indicates the name of the player involved in the transaction. This variable uniquely identifies the player in question and allows his transactions to be monitored over time.

---

<sup>35</sup> <https://github.com/marcovita33/FBFeesCalculator>

- age: represents the age of the player at the time of the transaction. Age can influence the value of a player, as younger, more promising players tend to have higher transfer values.
- position: Indicates the player's playing position. The position of the player can have a significant impact on the market value, as some positions are more desirable and require specific skills.
- club\_involved\_name: Identifies the other club involved in the transaction, who is buying the player. This information is useful for tracking interactions between clubs and understanding the context of the transaction.
- transfer\_period: Indicates the period in which the transaction took place, summer or winter. The transfer period can be related to market dynamics and the clubs' financial availability.
- fee\_cleaned: Represents the amount of the transaction in millions of Euros, without considering additional details. This value gives an idea of the economic value associated with the transfer of the player.
- league\_name: Specifies the league to which the club that made the transaction belongs. The league can influence the market value of a player, as some leagues are considered more competitive or prestigious than others.
- season: Indicates the year in which the transaction occurred. This information is useful for keeping track of transactions over time and analysing market trends over different seasons.
- OverallSeasonClub: Represents a measure of the team's overall performance during the season for the club mentioned in 'club\_name'. This aggregate value of the team's performance can influence the value of a player within the club.
- OverallSeasonClub2: Represents an alternative measure of the team's overall performance during the season for the club mentioned in "club\_involved\_name". This aggregate value of the team's performance can influence the value of a player within the club involved in the transaction.
- league\_destination: Indicates the league the player was transferred to, if different from the league of origin. This variable is relevant for understanding international market dynamics and the competitive environment in which the player transferred.

- Overall\_club\_name: Represents a measure of the team's overall performance, aggregated for all past years, for the club that owns the player. This value can reflect the reputation and historical performance of the club, which influence the market value of the player.
- Overall\_club\_involved: Represents a measure of the team's overall performance, aggregated for all past years, for the club purchasing the player. This value may reflect the reputation and historical performance of the purchasing club, which influence the market value of the player.

### **3.5 Selection of the machine learning model and performance evaluation techniques**

During the selection process of the final model, several challenges and iterations were carried out in order to obtain a reliable model for the prediction of players' transaction prices. Initially, the final dataset did not contain the 'OverallSeason' columns, which were later added after negative results were found in the first analyses.

Because the dataset included both numerical and categorical variables, several machine learning model options were explored, and were evaluated using two indicators:  $R^2$  and Mean Squared Error<sup>36</sup>.

The coefficient  $R^2$ , also known as the coefficient of determination, is a statistical measure used to assess the goodness of fit of a regression model. It provides an indication of the percentage of variation of the dependent variable that can be explained by the independent variables in the model.

In the context of predicting player transaction prices, the use of the coefficient  $R^2$  allows us to assess the accuracy of the machine learning model in predicting transaction prices.

The value of  $R^2$  can range from 0 to 1, where 0 indicates that the model is unable to explain any variation in the dependent variable, while 1 indicates that the model can fully

---

<sup>36</sup> James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). *An introduction to statistical learning* - Springer.

explain the observed variation. An intermediate value of  $R^2$  indicates the proportion of variation explained by the model relative to the total variation in the data.

Mean Squared Error (MSE) is a statistical measure used to assess the quality of a regression or prediction model. It represents the mean squares of the differences between the values predicted by the model and the actual values of the dependent variable in the dataset.

In the context of the prediction of players' transaction prices, the use of the MSE allows us to assess how much the model deviates from the actual transaction prices on average.

A lower MSE value indicates a higher accuracy of the model, as the differences between the forecast and the actual values are small. Conversely, a higher MSE value indicates a larger deviation between forecasts and actual values, indicating a lower accuracy of the model.

During the initial research phase, with a 20% test size, several models were evaluated, including Linear Regression, Polynomial Regression, SVM, Decision Tree and Decision Trees. However, poor results were obtained with negative  $R^2$  for the first two methods, and accuracy below 20% for the other models.

Next, neural network-based approaches, such as Artificial Neural Networks and Neural Networks, were explored, which showed significant improvement with accuracy around 30%. Gradient Boosting was then examined, which showed promising results with an accuracy of 30%.

The best results were achieved using the XGBoost algorithm, for which parameter optimisation via a grid was applied. This resulted in a significant increase in the  $R^2$  coefficient to 48%. Other models such as Random Forest and Gradient Boosting Machine achieved reasonable results with an accuracy of 40%.

However, in order to exceed the 50 per cent threshold of  $R^2$ , it was decided to include the columns 'OverallSeason' and 'OverallSeason2', already calculated, representing the club's static overall grade based on the UEFA results of the last 10 years, known as the SPI index.

With the new dataset, the best machine learning models were identified, namely Gradient Boosting Machine with 51% accuracy and XGBoost with 50% accuracy.



Some previously evaluated models showed an increase in accuracy (Random Forest at 44%, Gradient Boosting at 40% and Decision Tree at 27%), while others showed a decrease in performance.

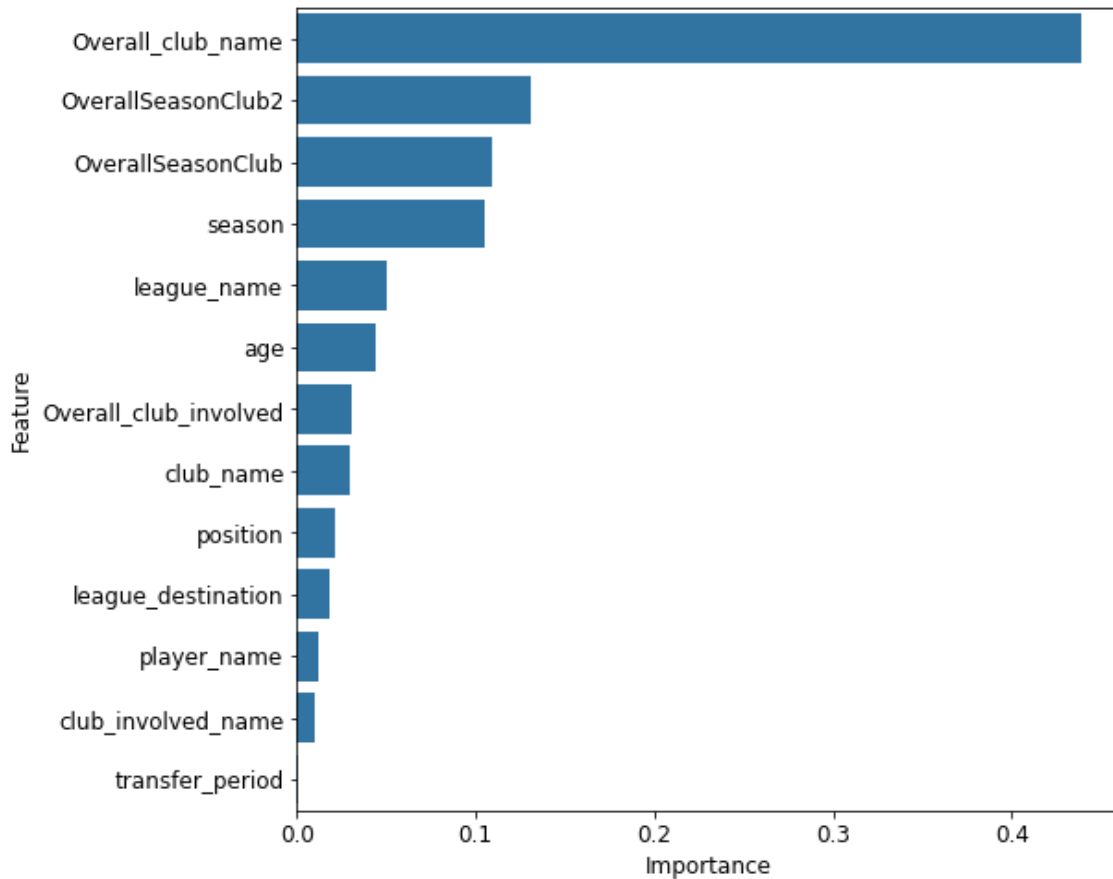


Figure 5 – Weight of the single variables

The table shown shows the importance of the different variables according to the results obtained using the Gradient Boosting Machine model. The analysis of the feature importances is a fundamental aspect in order to understand the influence of each variable in the process of predicting players' transaction prices.

From the results obtained, it is observed that the variable "Overall\_club\_name" is the most significant with an importance of 44%. The variables 'OverallSeasonClub2' and 'OverallSeasonClub' follow with importance of 13% and 11% respectively. Other relevant variables include 'season' with 10% importance and 'league\_name' with 5%. The variables 'age', 'overall\_club\_involved', 'club\_name' and 'position' present between 2% and 4% importance, while the remaining variables contribute less than 1%.

The implementation of this iterative model selection process resulted in a reliable machine learning model for the prediction of player transaction prices.

#### **4. Analysis of Juventus transfers fees case**

In the context of the research conducted, an audit of some of the exchange prices during the selected period of the prosecution in the Juventus FC capital gains case was carried out. Trading data from the team during this specific period were collected and analysed.

Then, a prediction algorithm was applied to estimate transaction prices of the selected players, using the available variables.

The predicted prices were then compared to the actual transaction prices recorded during the analysis window. The objective of this comparison was to assess the effectiveness of the prediction model in accurately predicting transaction prices during this specific period.

During the discussion of the results obtained, any discrepancies between the predicted prices and the actual transaction prices were analysed. Such discrepancies may provide important indications regarding variables or aspects that could significantly influence the value of players during the considered transfers.

Through this analysis and the subsequent discussion of the results, it was possible to assess the accuracy of the developed prediction algorithm and identify possible critical factors that may have a significant impact on transaction prices during the period in question. These results contribute to the understanding of the context of capital gains at Juventus FC and provide valuable information for future research and analysis in the field of football transactions.

##### **4.1 Application of the model to estimate transaction prices**

In order to specifically assess the players involved in Juventus' capital gains, it was necessary to isolate the rows of the dataset related to specific transfers. Among these, one

of the cases considered was the exchange between Arthur and Miralem Pjanic with Barcelona. Another case analysed was the exchange between Danilo and Joao Cancelo with Manchester City. In addition, the transfers made with Genoa were examined, which included the purchase of the young talent Nicolò Rovella by Juventus and the sale of two young players, Manolo Portanova and Elia Petrelli, to Genoa.

These trades contributed to Juventus' capital gains management strategy according to the prosecutors<sup>37</sup>.

In order to predict the transaction prices of the selected transfers various machine learning models were applied.

The techniques that had provided the best results up to that point were applied and the best performance was obtained with a Random Forest, which achieved an  $R^2$  of 51%.

The entire dataset was updated by standardizing the "fee\_cleaned" column in order to comprehend the most accurate way ever used to determine a player transaction fee. Then, were performed various predictions, using several machine learning approaches on a smaller test size (0,002), also utilizing many different seeds.

Again, Random Forest and Gradient Boosting models produced the greatest results, with both of them achieving the best results in a broad range of examples, keeping every Mean Squared Error very close to zero (0 followed by at least two zeros), and achieving  $R^2$  scores of 66% and 79%, respectively.

---

<sup>37</sup> Inchiesta Juventus: Le 22 Plusvalenze nel mirino della procura – Redazione DAZN (2022)

## 4.2 Comparing predicted prices with actual transaction prices

Below are the data obtained with the prediction model for the selected football transactions, comparing the actual transaction values with the predicted values. Let's analyse the results on a case-by-case basis:

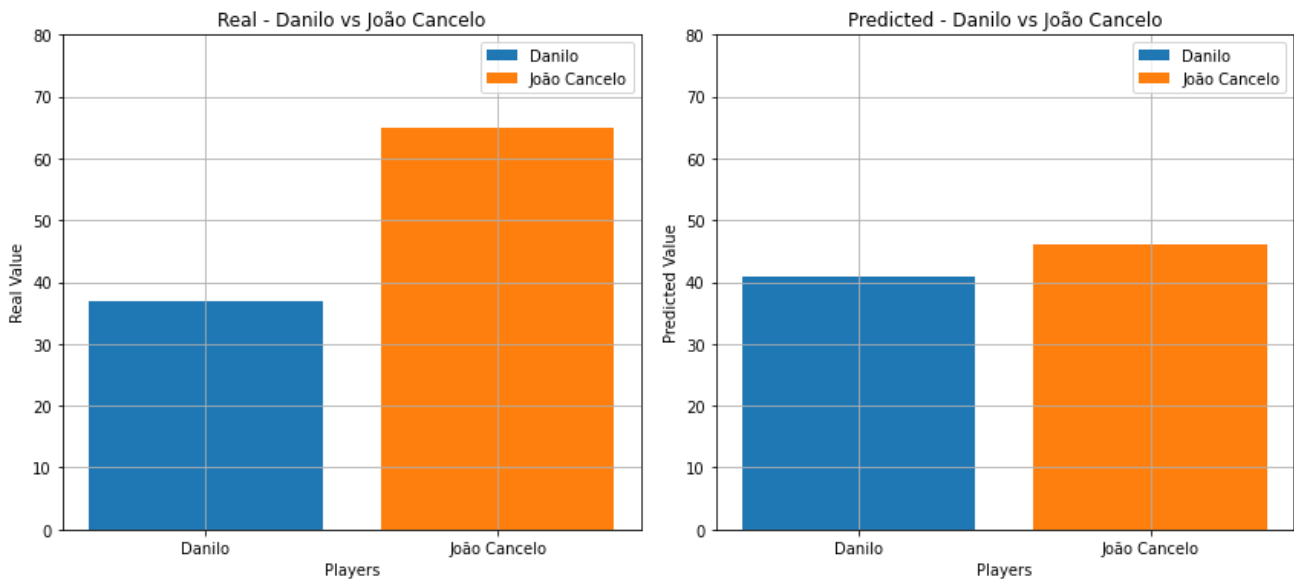


Figure 6 – Danilo and Cancelo, real vs predicted trade

João Cancelo<sup>38</sup>: The actual value of the transaction with Manchester City is EUR 65.0 million, while the predicted value is EUR 46 million. The difference between the values is -€18.9988 million. In this case, the predicted value is lower than the actual value of the transaction.

Danilo: The actual value of the transaction with Manchester City is EUR 37.0 million, while the predicted value is EUR 41 million. The difference is EUR 4 million. Here, the predicted value is slightly higher than the actual value of the transaction.

---

<sup>38</sup> Man City and Juventus confirm Cancelo-Danilo swap in €65m deal - Fotmob

The deal for João Cancelo indicates that the value predicted by the analysis turned out to be lower than the actual value of the transaction with Manchester City. This suggests that the player may have been valued higher by the market than predictions. In contrast, for Danilo, the prediction was closer to the actual transaction value. This could indicate that the player was more accurately valued or that the market more consistently reflected his value.

The discrepancy between the predicted and the real data may be due to negotiations: Cancelo was bought by Juventus a year prior for EUR 40 million, and in order for the Manchester City to buy 25 years old Cancelo they must had to offer a similar player, but older (both right-back, in national team and playing big competition), plus money to persuade Juventus.

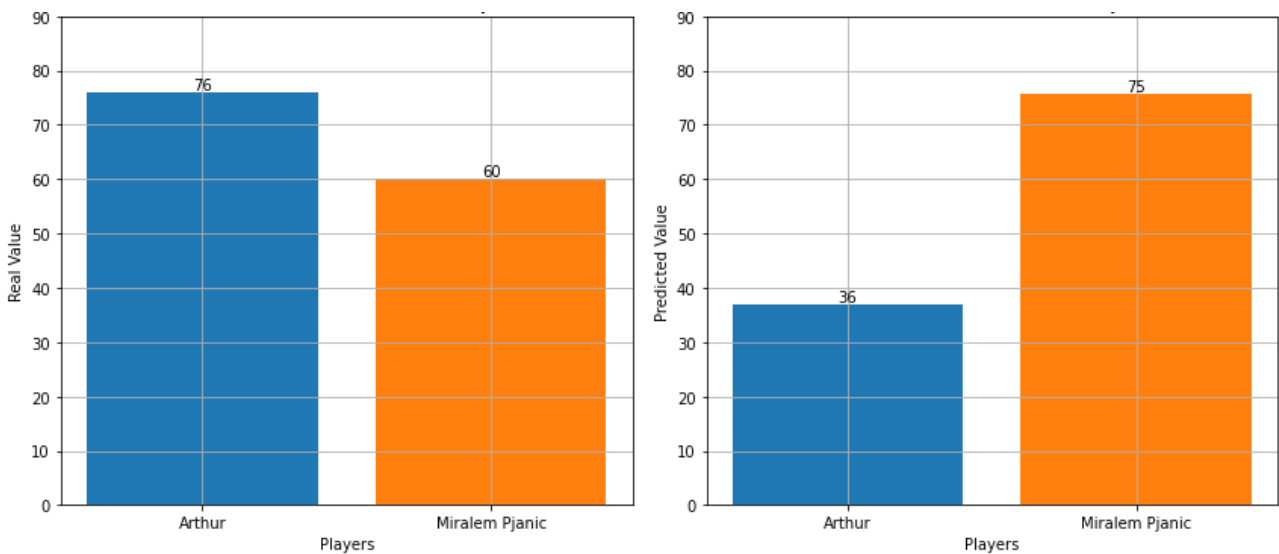


Figure 7 – Arthur and Pjanic, real vs predicted trade

Miralem Pjanic<sup>39</sup>: The real value of the transaction with FC Barcelona is €60.0 million, while the value predicted by the model is €75.7 million. The difference between the two

<sup>39</sup> Barcelona's proposed Arthur-Pjanic swap points to future full of NBA-like trades – Goal.com

values is €15.7 million. This indicates that the predicted value is higher than the actual value of the transaction.

Arthur: The actual value of the transaction with FC Barcelona is EUR 76.0 million, while the predicted value is EUR 37 million. The difference between the two values is -€39 million. This indicates that the predicted value is significantly lower than the actual value of the transaction.

This exchange of players between Arthur and Miralem Pjanic with FC Barcelona showed a significant discrepancy between the predicted and the actual transaction values. In the case of Arthur, the predicted value turned out to be much lower than the actual transaction value. This difference could indicate that the player was valued higher by the market than the prediction. On the other hand, in the case of Miralem Pjanic, the predicted value was found to be higher than the actual transaction value. This suggests that the player may have been valued lower by the market than the prediction.

One interpretation could be that in real negotiations the age parameter is a fundamental booster of the price (Arthur was 23 years old and Pjanic was 30 at the time of the trade), and that the method implied to predict the prices just rely on team performances and overalls without giving enough weight to the age.

These discrepancies highlight the importance of considering a number of complex factors, such as player reputation, sporting performance and market factors, that can justify this transaction pricing in football.

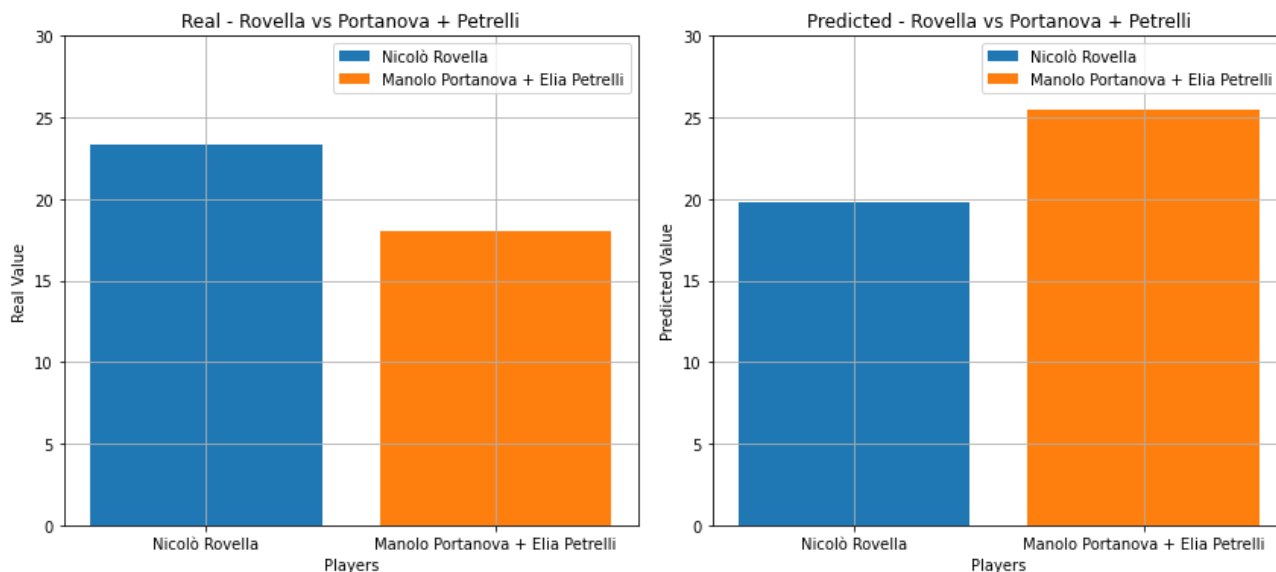


Figure 8 – Rovella and Portanova with Petrelli, real vs predicted trade

Manolo Portanova<sup>40</sup>: The actual value of the transaction with Genoa CFC is €10.0 million, while the predicted value is €13.3 million. The difference is EUR 3.3 million. Here, the predicted value is slightly higher than the actual value of the transaction.

Elia Petrelli: The actual value of the transaction with Genoa CFC is €8.0 million, while the predicted value is €12 million. The difference between the two values is EUR 4 million. Again, the predicted value is higher than the actual value of the transaction.

Nicolò Rovella: The actual value of the transaction with Genoa is €23.3 million, while the predicted value is €20 million. The difference is EUR -3.3 million. Here, the predicted value is lower than the actual value of the transaction.

The analysis of the cases of Manolo Portanova, Elia Petrelli and Nicolò Rovella, concerning transactions with Genoa CFC, revealed some interesting discrepancies between the predicted values and the actual transaction values. In the case of Manolo Portanova, the predicted value was found to be slightly higher than the actual transaction

<sup>40</sup> Juve & Genoa set to complete Rovella & Portanova swap - Lucrative deal for the accounts - transfermarkt.com

value. This could indicate that the player was valued higher by the model than the actual market value. Similarly, in the case of Elia Petrelli, the predicted value turned out to be higher than the actual transaction value. This suggests that the model may have overestimated the value of the player compared to what he was actually paid.

On the other hand, in the case of Nicolò Rovella, the predicted value turned out to be lower than the actual transaction value. This discrepancy could be attributed to the fact that the model underestimated the player's value or that the player was valued higher by the market than prediction.

In general, the results show that the prediction model obtained forecasts that deviated both upwards and downwards from the actual transaction values. These discrepancies can be attributed to various factors, such as market fluctuations, player performance or other elements that influence transaction pricing in football.

The discrepancy between data are approximately the same and the prediction appear quite realistic. In order to adjust differences, as they are mainly youth players, under 21 teams would have to be included as a separate value from their principal team.

The analysis of such discrepancies highlights the complexity and uncertainty associated with transaction pricing in football. Many factors, such as player potential, age, sporting performance and market expectations, can influence transfer values. Analysing these discrepancies can provide important insights to better understand football market dynamics and improve the accuracy of future predictions.

## **5. Summary of the Findings**

In this Master's thesis, the objective was to address the challenge of predicting football player transaction prices by using a machine learning model. We analysed the context of football player transactions, their importance in the football industry, and the specific case of the capital gains investigation against Juventus FC.

Then was developed a machine learning model capable to estimate transaction prices of selected player transfer in order to compare the predicted prices with the actual transaction



prices. Through the analysis, we found that the prediction algorithm, based on machine learning techniques, showed promising results in estimating transaction prices. We considered different machine learning techniques, from which the Random Forest and Gradient Boosting algorithms achieved the best performance, with  $R^2$  scores of 66% and 79%, respectively. These results indicate that the methods developed can provide valuable insights into player valuation and transaction prices in the football market.

When comparing the predicted prices with the actual transaction prices, some discrepancies were observed. These differences highlight the complex nature of player valuation in football and the various factors that can influence transaction prices. In some cases, the predictions were lower or higher than the actual values, indicating that market dynamics and negotiation factors play a significant role in determining player prices. Furthermore, the analysis of specific player exchanges, as the swap between Arthur and Pjanic and the trade between Danilo and Cancelo revealed interesting insights, while trades between Juventus and Genoa were in line with actual prediction.

The differences between the predicted and actual transaction values emphasized the importance of considering multiple factors, including player reputation, age, sporting performance, and market conditions, in determining player prices.

## **5.1 Application of the tool for evaluation of players and teams**

The tool can be of great value to clubs, player agents and observers, who can rely on the valuations provided by the model to make informed decisions regarding player valuations and investment potential. This approach offers an objective methodology not only to determine the market values of players but also to evaluate the overall market behaviour of the teams.

One of the main applications of the machine learning model concerns the identification of potential signings. As market strategies play a key role in football, by using the data and evaluations generated by the analysis, teams can identify footballers with a high market value but a lower purchase price than their real capabilities, as well as better

analyse proposal to buy or sell players. Allowing them to acquire talented player at a low cost, negotiate fair prices or maximizing the income of a transaction.

By analysing the estimated values of players in different positions and roles, the model can be used to find potential signings to strengthen the team or identify players to be sold in order to free up financial resources and improve the overall budget.

The model can be used also to monitor trends in the market value of players over time. Teams can analyse trends and value fluctuations to make timely decisions regarding the purchase, sale or renewal of players' contracts. This allows them to adapt to market dynamics more effectively and make strategic decisions based on available opportunities.

## **5.2 Suggestions for future developments and possible improvements of the algorithm**

There are numerous opportunities for expansion and improvement to enhance the effectiveness and adaptability of the model to meet the evolving needs of football clubs. Presented below are key suggestions for future developments and improvements of the model:

- Physical conditions can have a significant impact on his performance and value. The model could be enhanced to consider information on a player's past injuries, medical history and performance after recovery.
- Acknowledge the challenges players face when adapting to different playing styles or new countries, the model should consider past player experiences to address potential success in varying context and cultural environments.
- Individual performances and awards won impacts on the marketability of a player, so include them in the model could enhance the overall precision.

-Inclusion of financial and marketing data to explore the impact of external factors such as macroeconomic conditions, regulatory changes, and transfer market dynamics on player transaction prices. For example, the value of a player could be influenced by his ability to generate commercial revenue for the club through sponsorship or merchandise sales.

-Integrating a real-time data analysis system would allow for constant monitoring of players' latest transactions to better understand market dynamics because the analysis focused on a limited number of player exchanges.

-Work closely with coaches, team managers, and player agents to incorporate additional variables and factors that can contribute to a more accurate evaluation. Their expertise and insights can enhance the model's predictive capabilities.

The implementation of these suggestions and the pursuit of further developments and improvements will help make the football transaction value predicting model increasingly sophisticated and effective in supporting the market decisions of teams.

## 6. Bibliography

1. Transfer (association\_football) – Wikipedia  
[https://en.wikipedia.org/wiki/Transfer\\_\(association\\_football\)](https://en.wikipedia.org/wiki/Transfer_(association_football))
2. Repository with full code: <https://github.com/marcovita33/FBFeesCalculator>
3. Plusvalenze fittizie e manipolazione del mercato: il caso Juventus F.C.- F. Maraschin  
<https://www.iusinitinere.it/plusvalenze-fittizie-e-manipolazione-del-mercato-il-caso-juventus-f-c-44005>
4. Plusvalenze Juventus: quali sono quelle sospette. I giocatori (e le squadre) coinvolti – Il Giorno  
<https://www.ilgiorno.it/sport/calcio/juventus-plusvalenze-1.7084632>
5. Explaining Why Juventus Are So Willing To Swap Miralem Pjanic For Barcelona's Arthur Melo - Forbes  
<https://www.forbes.com/sites/chloeberesford/2020/06/28/transfers-barcelona-juventus-pjanic-arthur-sarri-midfield7>
6. Il senso dello scambio Danilo-Cancelo - Ultimouomo.com  
<https://www.ultimouomo.com/cancelo-danilo-senso-scambio/>
7. Juve, Rovella è la madre di tutte le plusvalenze: le cifre e i valori 'scandalosi' di Portanova e Petrelli – calciomercato.com  
<https://www.calciomercato.com/news/juve-rovella-e-la-madre-di-tutte-le-plusvalenze-tutte-le-cifre-e-52640>

8. Bhatia, Adil. Do Football Transfers Actually Improve Performance? Bachelor's thesis, Harvard College. (2020)  
<https://dash.harvard.edu/bitstream/handle/1/37364729/BHATIA-SENIORTHESIS-2020.pdf?sequence=1>
9. Determinants of football players' valuation: A systematic review M. Franceschi, Wiley Online Library (2023)  
<https://onlinelibrary.wiley.com/doi/full/10.1111/joes.12552>
10. How do you value a player – Forbes  
<https://theathletic.com/3085749/2022/01/27/premier-league-how-do-you-value-a-player/>
11. Transfer (association\_football) – Wikipedia  
[https://en.wikipedia.org/wiki/Transfer\\_\(association\\_football\)](https://en.wikipedia.org/wiki/Transfer_(association_football))
12. The art of negotiation – Sky Sports  
<https://www.skysports.com/football/news/11096/7339197/the-art-of-negotiation>
- Role of agents and intermediaries in sports - iPleaders
13. <https://blog.ipleaders.in/role-agents-intermediaries-sports/>
14. The valuation of human capital in the football player transfer market – van den Berg, Erasmus University (2011)  
<https://thesis.eur.nl/pub/9763>
15. Introduction to Machine Learning, fourth edition. Adaptive Computation and Machine Learning series - Ethem Alpaydin - MIT Press (2020)  
<https://dl.matlabyar.com/siavash/ML/Book/Ethem%20Alpaydin-Introduction%20to%20Machine%20Learning->

[The%20MIT%20Press%20\(2014\).pdf](#)

16. Shalev-Shwartz, S., & Ben-David, S. (2014). Understanding Machine Learning: From Theory to Algorithms. Cambridge: Cambridge University Press  
<https://www.cs.huji.ac.il/~shais/UnderstandingMachineLearning/understanding-machine-learning-theory-algorithms.pdf>
  
17. Poli R, Besson R, Ravenel L. Econometric Approach to Assessing the Transfer Fees and Values of Professional Football Players. *Economies*. (2022)  
<https://www.mdpi.com/2227-7099/10/1/4>
  
18. Užík M, Warias R, Glova J. Management of Transfer Prices in Professional Football as a Function of Fan Numbers. *Mathematics*. (2022)  
<https://www.mdpi.com/2227-7390/10/16/2982>
  
19. <https://www.footballtransfers.com/en/values/players>
  
20. [https://www.footballbenchmark.com/methodology/player\\_valuation](https://www.footballbenchmark.com/methodology/player_valuation)
  
21. A robust method for clustering football players with mixed attributes - D'Urso, De Giovanni, Vitale – Springer (2022)  
[https://iris.luiss.it/retrieve/7963c5d4-304d-43c3-a638-f98497498689/ANOR\\_2022\\_football.pdf](https://iris.luiss.it/retrieve/7963c5d4-304d-43c3-a638-f98497498689/ANOR_2022_football.pdf)
  
22. Pricing Football Players using Neural Networks – Dey, University of Southern California (2017)  
<https://arxiv.org/ftp/arxiv/papers/1711/1711.05865.pdf>

23. <https://github.com/sanjitva/Predicting-Football-Player-Transfer-Values>
24. PlayeRank: Data-driven Performance Evaluation and Player Ranking in Soccer via a Machine Learning Approach - Luca Pappalardo, Paolo Cintia, Paolo Ferragina, Emanuele Massucco, Dino Pedreschi, and Fosca Giannotti. (2019)  
<https://dl.acm.org/doi/10.1145/3343172>
25. Pricing Football Players using Neural Networks – Dey, University of Southern California (2017)  
<https://arxiv.org/ftp/arxiv/papers/1711/1711.05865.pdf>
26. How association club coefficients are calculated - UEFA  
<https://www.uefa.com/nationalassociations/uefarankings/country/about/>
27. <https://www.uefa.com/nationalassociations/uefarankings/tenyears/>
28. <https://projects.fivethirtyeight.com/soccer-predictions/>
29. <https://data.world/dcereijo/player-scores>
30. <https://github.com/ewenme/transfers>
31. Ian G. McHale, Benjamin Holmes. Estimating transfer fees of professional footballers using advanced performance metrics and machine learning, - European Journal of Operational Research (2022)  
<https://www.sciencedirect.com/science/article/pii/S0377221722005082>
32. Using machine learning to identify high-value football transfer targets, Tattersal. – Towards data Science (2021)  
<https://towardsdatascience.com/using-machine-learning-to-identify-high-value-football-transfer-targets-d4151a7ffcac>

33. FIFA player ratings explained: How are the card number & stats decided? – Goal.com  
<https://www.goal.com/en/news/fifa-player-ratings-explained-how-are-the-card-number--stats-decided/1hszd2fgr7wgf1n2b2yjdpgynu>
34. Current & Potential Ability – FMINSIDE.net  
<https://fminside.net/guides/basic-guides/76-current-potential-ability>
35. <https://github.com/marcovita33/FBFeesCalculator>
36. James, G., Witten, D., Hastie, T., & Tibshirani, R. An introduction to statistical learning - Springer. (2021)  
<https://towardsdatascience.com/using-machine-learning-to-identify-high-value-football-transfer-targets-d4151a7ffcac>
37. Inchiesta Juventus: Le 22 Plusvalenze nel mirino della procura – Redazione DAZN (2022)  
<https://www.dazn.com/it-IT/news/serie-a/inchiesta-juventus-le-22-plusvalenze-nel-mirino-della-procura/f5cyc6xmdwqq1sz5xgpq6zq57>
38. Man City and Juventus confirm Cancelo-Danilo swap in €65m deal – Fotmob  
<https://www.fotmob.com/news/1b15u170gzuvy1c8i3egl116hi/man-city-and-juventus-confirm-cancelo-danilo-swap-in-%E2%82%AC65m-deal>
39. Barcelona's proposed Arthur-Pjanic swap points to future full of NBA-like trades – Goal.com  
<https://www.goal.com/en-us/news/barcelonas-proposed-arthur-pjanic-swap-points-to-future-full-of-nba-like-trades/8x67hbsadf0c1n2dt1im60qmt>
40. Juve & Genoa set to complete Rovella & Portanova swap - Lucrative deal for the accounts - transfermarkt.com



<https://www.transfermarkt.com/juve-amp-genoa-set-to-complete-rovella-amp-portanova-swap-lucrative-deal-for-the-accounts/view/news/377155>

Repository with full code: <https://github.com/marcovita33/FBFeesCalculator>

## 7. Summary

This Master's thesis in Data Science and Management aims to develop a machine learning algorithm for accurately predicting football players' transaction prices. The project recognizes the significance of player transactions in modern football, as they play a crucial role in clubs' performance improvement and achievement of sporting and financial goals. The thesis acknowledges the complex network involved in player transactions, including clubs, agents, intermediaries, and sports organizations, which creates a dynamic and competitive market.

The ability to accurately predict players' transaction prices has become vital for clubs and insiders in the transfer decision-making process. It provides a competitive advantage in the transfer market, enabling clubs to optimize financial resources and build competitive teams. Leveraging data analysis techniques and advanced predictive models, the thesis aims to identify the key factors influencing players' prices and create a predictive system capable of accurately estimating a player's value in the market.

To achieve this goal, the thesis will utilize data from various sources on football, considering player characteristics, team performance, and other relevant variables. Statistical methods and supervised learning algorithms, such as multiple linear regression or xgboost, will be employed to train the machine learning algorithm. The selection of relevant variables, data quality management, and cross-validation techniques will be crucial for evaluating the effectiveness of the predictive model.

The research aims to provide an innovative and accurate method for evaluating player transaction prices using machine learning. The results obtained can be applied in professional football and offer valuable insights for market decisions. Additionally, the methodology and approach used in this thesis can be extended to other sports sectors or similar economic evaluation contexts.

In the context of the Juventus FC capital gains case, the competent authorities have closely scrutinized several player exchanges. Specifically, two exchanges have received significant attention. The exchange between Arthur and Pjanic in 2020 raised questions about player valuation and the fairness of the transaction. Similarly, the exchange between Cancelo and Danilo in the previous year underwent close scrutiny. The authorities aimed

to assess the accuracy of player valuations and compliance with financial and regulatory rules.

The investigation into capital gains at Juventus FC involved a comprehensive evaluation of the club's player exchange transactions. The objective was to ensure fairness, equity, and compliance with football's financial and regulatory rules. The analysis of these transactions in the context of the capital gains case sheds light on the need for financial transparency and integrity in market transactions, generating discussions about the regulation of financial transactions in football.

The second chapter of this thesis explores the theoretical foundations and state of the art in football player transactions and price prediction. It analyzes the economic significance of these transactions and their impact on teams, players, and the football industry. Factors influencing transaction prices, including age, experience, skill level, and role, are examined. The chapter also investigates machine learning techniques used for price prediction, highlighting effective approaches and methodologies.

This theoretical foundation sets the stage for innovative research in football player transaction price prediction using advanced machine learning. The football player transaction market is economically important, impacting teams, players, and the industry as a whole. Successful acquisitions can enhance performance, attract sponsors, increase sales, and improve the club's image. Poor transaction management can lead to negative consequences. Understanding and predicting transaction prices through machine learning can provide a strategic advantage.

Factors influencing transaction prices include age, experience, current performance, popularity, role, league, demand, and contract length. These factors interact and affect the negotiation process. Agents and intermediaries play a crucial role in facilitating negotiations, while pricing mechanisms involve supply, demand, and bargaining. Understanding these concepts is vital for analyzing the football player transaction market.

The application of machine learning in the sports industry, particularly in predicting transaction prices in the football player market, has gained significant interest. Machine learning allows for the analysis of vast amounts of data to build predictive models that estimate future transaction prices. Various techniques, both supervised and unsupervised, are employed in this domain. Supervised approaches use labeled training data to train

models, while unsupervised approaches discover patterns and relationships in data without labels. Common algorithms used include linear regression, decision trees, random forest, support vector machines, and neural networks.

The use of machine learning techniques in the football player market enables clubs to make more accurate predictions about transaction prices, aiding in strategic decision-making. It allows for the estimation of a player's value, identification of investment opportunities, and assessment of player acquisitions and disposals. Real-world examples include regression algorithms for estimating player values, unsupervised clustering algorithms for identifying player groupings, and neural networks for predicting transaction prices.

The choice of algorithm for transaction price prediction depends on data availability, problem complexity, and prediction objectives. Linear regression, decision trees, random forest, support vector machines, and neural networks are among the common algorithms used. Each algorithm has its strengths and limitations, and a combined approach often provides more accurate results. Factors like data availability, problem complexity, and prediction goals influence algorithm selection.

Literature review plays a vital role in understanding football player price prediction. Previous research has explored traditional and innovative approaches, including econometric models and machine learning techniques. Key factors influencing prices, such as player age, past performance, and league affiliation, have been investigated. Machine learning approaches have allowed for the analysis of large datasets, including statistical, market, and performance data, to develop accurate predictive models.

Both traditional and innovative approaches have been used in modeling transaction prices in the football player market. Traditional approaches rely on econometric models like linear regression and time series models, while innovative approaches utilize machine learning algorithms such as artificial neural networks, decision trees, and support vector machines. Combining both approaches can yield better results by leveraging the strengths of each.

Challenges in transaction price prediction include data availability and quality, market dynamics, complexity of variables, and interpretability of machine learning models. Data quality and availability can impact prediction accuracy, while market dynamics require

adapting predictions to changing conditions. The complexity of variables involved, beyond player characteristics, poses challenges in model development. Interpretability of machine learning models is also important for transparency. Combining domain knowledge, statistical methods, and machine learning techniques can address these challenges and improve predictions.

The research focuses on analyzing the performance of football teams using various datasets. The datasets used in this study are "uefarankteam3.csv," "spi\_global\_rankings.csv," "games.csv," and "transfers."

The "uefarankteam3.csv" dataset provides reliable rankings of teams based on the UEFA system. It includes UEFA scores accumulated over the past 10 years, reflecting teams' strength and consistency in UEFA competitions. The dataset was obtained through web scraping from UEFA's official website, allowing for automated data collection and analysis.

The "spi\_global\_rankings.csv" dataset contains SPI (Soccer Power Index) scores, which represent an overall measure of a team's strength, considering offensive and defensive aspects. It uses advanced algorithms and factors such as past performance, match results, and opponent quality to calculate scores.

The "games.csv" dataset provides a wealth of information on football games, including team names, results, goals, and more. It covers major national and international competitions since 2012, allowing for detailed analysis of team performance and match outcomes.

The "transfers" dataset contains detailed information on player transactions within European football clubs since the 1992/93 season. It includes data on major European leagues and variables such as club names, player names, transfer fees, and more. This dataset is valuable for understanding market dynamics and developing predictive models for player transaction prices.

Data cleaning and pre-processing steps were applied to ensure data quality and comparability. Team name matching, data filtering, score normalization, and weighted score calculations were performed to assess team performance. The "Overall" scores were used to evaluate teams' overall strength.

The research integrated data from different datasets, such as merging match results with the transfers dataset, providing a comprehensive view of player transactions and team performance.

Variables like FIFA overall and Football Manager potential ability were excluded from the study due to their limited relevance to predicting transaction prices in professional football. Other variables, such as SPI scores and UEFA scores, were considered more relevant to understanding market dynamics and factors influencing prices.

The chosen methodology ensured a systematic analysis of team performance and provided a solid basis for future predictions in the football sector. The research aimed to avoid biases and limitations observed in previous studies and focused on variables closely related to transfer market dynamics.

Overall, this research provides valuable insights into team performance analysis and predictions regarding player transaction prices in professional football.

The research study conducted on the exchange prices of some transfers in the Juventus FC capital gains case. The study involved collecting and analyzing trading data from the team to assess the effectiveness of a prediction algorithm in estimating transaction prices. The algorithm, based on a machine learning model, used available variables to make predictions on player valuation.

The results revealed that the prediction algorithm achieved the best performance using Random Forest and Gradient Boosting algorithms, with  $R^2$  scores of 66% and 79% respectively. The algorithm was able to predict transaction prices for various football transfers, and a comparison was made between the predicted values and the actual transaction values.

The analysis of specific cases, such as the exchanges between Danilo and João Cancelo with Manchester City, and Arthur and Miralem Pjanic with FC Barcelona, revealed discrepancies between the predicted and actual transaction values. In some cases, the predicted values were lower or higher than the actual values, indicating possible overvaluation or undervaluation by the market. The age of the players was identified as a potential factor influencing the transaction prices.

Furthermore, transactions involving Genoa CFC, including the purchases of Nicolò Rovella and the sales of Manolo Portanova and Elia Petrelli, also showed discrepancies between the predicted and actual values. The algorithm tended to slightly overestimate or underestimate the player's values, highlighting the need to consider additional factors such as market fluctuations and player performance.

Overall, the results demonstrate the complexity and uncertainty associated with transaction pricing in football. Factors such as player potential, age, sporting performance, and market expectations all contribute to the determination of transfer values. The analysis of these discrepancies provides valuable insights into football market dynamics and offers a basis for improving the accuracy of future predictions.

The algorithm developed for predicting transaction values in football has demonstrated its power and versatility, extending its potential applications beyond its initial purpose. The chapter explores the algorithm's various uses, with a focus on player and team valuation, market strategies of football clubs, and recommendations for future enhancements.

Regarding player and team valuation, the algorithm provides an innovative and objective perspective by leveraging historical transaction data and relevant variables. This enables accurate and reliable estimates of players' market values. The algorithm's valuations support clubs, player agents, and observers in making informed decisions related to player valuations, investment potential, and team formation. Moreover, the algorithm extends its evaluation capabilities to teams as a whole, considering the interactions between players to assess overall performance and market value. This analysis assists coaches and managers in identifying team strengths and weaknesses, devising strategies, and managing financial resources efficiently.

The application of the algorithm for player and team evaluation offers several advantages. It mitigates the risk of subjective evaluations and personal biases, facilitating decisions based on objective data. Additionally, it identifies under- or over-valued players, uncovering investment opportunities and optimizing transfer strategies to maximize market value. However, it is crucial to acknowledge that the algorithm serves as a decision support tool and should not replace human expertise, as non-quantifiable aspects like personality and future growth potential require additional analysis.

Market strategies significantly impact a team's success, and the algorithm enhances the evaluation of players and teams to optimize these strategies. Firstly, it aids in identifying potential signings by evaluating data and estimations to uncover players with high market value and relatively lower purchase prices. This enables smart investments, granting teams a competitive advantage. Secondly, the algorithm helps evaluate incoming and outgoing transfer offers, allowing for objective decision-making, fair price negotiations, and maximization of transaction values. Furthermore, the algorithm assists in optimizing team composition by identifying areas for improvement and suggesting potential signings or player sales to enhance the overall budget. It also aids in refining bargaining strategies by assessing the contractual value of players, ensuring optimal balance over time. Lastly, the algorithm monitors market value trends, empowering teams to make timely decisions regarding player contracts, purchases, sales, or renewals, thus capitalizing on available opportunities.

While the algorithm has already demonstrated its usefulness, there are opportunities for expansion and improvement. Suggestions for future developments include incorporating additional variables to increase complexity, implementing artificial intelligence and machine learning for dynamic adaptations, introducing real-time data analysis for timely decision-making, ensuring reliability of data sources, involving industry experts for valuable insights, analyzing long-term performance, considering tactical context, evaluating the impact of injuries and physical conditions, including financial and marketing factors, adapting to regional and cultural contexts, and constantly verifying and updating models to reflect market dynamics.

By implementing these suggestions and pursuing further advancements, the football transaction value assessment algorithm will evolve into a sophisticated and effective tool for supporting market decisions in the football industry.

This Master's thesis aimed to predict football player transaction prices by developing a machine learning algorithm. The study analyzed the context of football player transactions, focusing on the capital gains investigation against Juventus FC. The prediction algorithm was applied to estimate transaction prices and compared with actual prices. The results showed promising performance, with Random Forest and Gradient Boosting algorithms achieving  $R^2$  scores of 66% and 79% respectively. However, some



discrepancies were observed between predicted and actual prices, highlighting the complex nature of player valuation and the influence of market dynamics and negotiation factors. The analysis of specific player exchanges provided interesting insights, emphasizing the importance of considering multiple factors in determining player prices.

The development of the machine learning algorithm has implications for the football industry and its stakeholders. It provides a valuable tool for clubs, agents, and intermediaries involved in player transactions, enabling informed decisions and operational optimization. Accurately estimating player values allows clubs to strategize their transfer activities, optimize financial resources, and build competitive teams. Agents and intermediaries can benefit from reliable estimates to enhance negotiation strategies. The research contributes to the understanding of market dynamics in the football industry and offers insights into asset valuation and transactions in other industries.

despite the promising results, there are limitations to address. The accuracy of the prediction algorithm depends on data availability and quality, necessitating improvements in data collection and accuracy. Expanding the scope of analysis to include more transactions and a longer timeframe would provide a broader understanding of player valuation and transaction prices. Incorporating additional variables and more sophisticated machine learning models could enhance prediction accuracy. Exploring the impact of external factors such as macroeconomic conditions and regulatory changes would contribute to a comprehensive prediction model. Qualitative research, including case studies and interviews, could provide valuable insights into decision-making processes.

Overall, this research contributes to the field of player valuation and transaction pricing and offers a foundation for further enhancements in prediction accuracy and understanding.