**Logo, company name

Description automatically generated**

**Valuing Benfica Football Players**

**Programming for Data Science**

**Prof. Carlos J. Costa**

Areska, T. | l58521 | [l58521@aln.iseg.ulisboa.pt](mailto:l58521@aln.iseg.ulisboa.pt)

De Besi, E. | l59033 | [e.debesi@aln.iseg.ulisboa.pt](mailto:e.debesi@aln.iseg.ulisboa.pt)

Junca, D. | l58520 | [l58520@aln.iseg.ulisboa.pt](mailto:l58520@aln.iseg.ulisboa.pt)

# Introduction

The transfer market in football is a multi-billion dollar industry, with clubs around the world investing heavily in player acquisitions. Determining the market value of a player can be a complex and challenging task. In this paper, we present a statistical model that aims to solve the problem of accurately determining the prices of S.L. Benfica football players based on their age, goals scored, positions, preferred foot, and other relevant factors.

**Objectives:**

The main objective of this paper is to develop a statistical model that can accurately predict the prices of S.L. Benfica football players based on their performance metrics and other relevant factors. To achieve this objective, we have trained our model on historical sales data starting in 2010, using performance metrics such as goals scored, assists, and cards, as well as player age, position, and preferred foot, to predict the prices of current players.

In addition, we have secondary objectives that include understanding how player prices are determined and identifying which factors are most important in determining the prices of football players.

By achieving these objectives, we hope to provide valuable insights into the transfer market and help football clubs make more informed decisions when buying and selling players.

**Contributions:**

This model can help clubs make informed decisions when buying and selling players in the transfer market. Secondly, our study provides insights into which factors are most important in determining the prices of football players, which can be useful for future research in this field.

In conclusion, our paper presents a novel approach to predicting the prices of football players using a statistical model based on their performance metrics and other relevant factors. We believe that our model has the potential to revolutionize the way football clubs evaluate player values and make transfer decisions, not only for S.L. Benfica but also for other football clubs around the world.

# Literature review

**[WIP]**

Words record player transfer fee and SL Benfica quite often used in one sentence. For SL Benfica, one of Portugal's most successful football clubs, player transfers represent a key strategy for maintaining their position as one of the countries and top teams and help remain among the elite clubs in Europe. In this study, we aim to predict the value of player transfers for SL Benfica using various machine learning techniques. Our objective is to develop a model that can accurately estimate the market value of a player, based on a range of factors such as age, position, and performance metrics. By doing so, we hope to provide valuable insights for public and SL Benfica fan base, enabling them to understand transfer process better and forecast club’s possible transfer values more accurately.

Football is a multi-billion-dollar market that continues to grow, especially in terms of transfer fees paid for players. While certain variables that influence player’s market value have been studied empirically, such as their demographics, characteristics, and performance, other factors such as the influence of advisors, outfitters, and social media popularity have not been extensively analysed. Lennart Ante analyse the data from the summer transfer window of the 2018/19 season into the five major European leagues, namely the Premier League (England), Primera Division (Spain), Bundesliga (Germany), Serie A (Italy) and Ligue 1 (France). Author uses multiple regressions to identify effects for sub-populations in terms of transfer fee size, continent of birth and playing position. (Ante, 2019). Ante suggests that existence of external advisors has a positive effect on transfer fees above €5 million and the existence of outfitters has an explanatory value. Therefore, arguing that generalized models across playing positions and heritage may only provide basic information but do not possess practical use.

In football, where salary information is generally unavailable, crowd-source estimated values are common when predicting possible values of transfer. Dennis Coates argues that one of the best know crowd-sourced database in football, Trasnfermarkt crowd-sourced "market value" underestimates the worth of players with national team experience and presents a biased estimate of true value. The findings of his study suggest caution should be exercised when using transfer fees as a proxy for actual fees or player salaries in regression analysis. The accuracy of Transfermarkt values varies across the distribution of fees and between top and lesser leagues (Coates, 2022). As highlighted by Coates the crowd-sourced market value should be taken with a grain of salt therefore in our research sustain from using such estimates and try to investigate alternative sources for price estimation.

# Methodology

# Results

# Discussion

Based on the results obtained from our study, we have developed three different models to determine player prices: linear regression, random forest, and KNN. Among them, the random forest model provided the best results with a mean square error of 122298172626628,92. However, it is important to note that the data set used for the study was relatively small, consisting of only 49 historical sales from Benfica. This limited data set could have contributed to the large mean square errors observed in the models.

Regarding the limitations of the study, the small data set size may have impacted the accuracy of the models. With only 49 historical sales to train and test the models, the results may not be fully representative of the broader transfer market. Additionally, the decision to only include players who played the season before being sold in Benfica may have limited the scope of the analysis. Players who were sold straight after being loaned to other teams were removed from the data set because for such cases, we would have to collect their most recent stats from the team data they were loaned to before being sold. It would be very time-consuming, so we decided to only go with players who played a full season in Benfica before being sold, resulting in a smaller dataset.

The small data set size also means that the models may not be robust enough to capture all the factors that determine player prices in the market. The inclusion of more data from different leagues, teams, and player positions could potentially improve the accuracy of the models. Furthermore, the absence of important factors, such as injury history, could have also affected the accuracy of the models. Incorporating such data could lead to more accurate predictions of player prices.

The age of the players sold by Benfica ranged from 18 to 32, with most players being sold at ages 23 and 25. In terms of positions, 21 players in the Attack group were sold, followed by 13 players in the Midfield group, 9 in the Defense group, and 3 goalkeepers. Additionally, the Premier League and La Liga were the top two leagues that bought players from Benfica, with an average transfer price of $37,246,154 and $27,916,667, respectively.

Despite these limitations, the study provides valuable insights into the determinants of player prices in the football transfer market. The use of machine learning models can help football clubs make more informed decisions when it comes to player transfers. By identifying the key factors that determine player prices, clubs can better allocate their resources and maximize their returns on investment.

In terms of suggestions for future research, one area that could be explored is the inclusion of more variables in the model. For example, data on player injuries and performance in big games could be incorporated to further refine the model's accuracy. Additionally, expanding the data set to include more historical sales from other football clubs could provide a more comprehensive analysis of player transfer prices. Additionally, incorporating additional data sources such as social media metrics or fan sentiment could potentially improve the accuracy of the models. Lastly, it would also be worthwhile to explore the applicability of the models to other sports and industries beyond football.

Another suggestion for future research could be to investigate the impact of the nature of National Team experience on player value. For instance, the effect of representing a top-ranked team in international competitions versus a lower-ranked team or participating in major tournaments versus friendly matches could be explored. Understanding these nuances could help football clubs make more informed decisions when it comes to player transfers and potentially lead to more accurate predictions of player prices.

# Conclusion

# Reference list

1. Ante, L. (2019). Determinants of transfers fees: Evidence from the five major European football leagues. University of Hamburg: Hamburg, Germany.
2. Coates, D., & Parshakov, P. (2022). The wisdom of crowds and transfer market values. European Journal of Operational Research, 301(2), 523-534.