ERASMUS UNIVERSITY
ROTTERDAM
Erasmus School of Economics
Master Thesis [Data Science and Marketing Analytics]

Market of Markets: Do European football transfer details influence teams' stock prices? - A machine learning calculation

Student:	Gábor Soós
Student ID number:	698623
Supervisor:	prof.dr. Richard Paap.
Second assessor:	Dr. Clément S. Bellet
Date final version:	11th August 2024

Abstract

This study investigates the impact of European football transfer details on the stock prices of publicly traded football clubs. The primary objective is to determine the significance of transfer-related factors compared to financial and team form-related variables in predicting stock price changes. While previous research has primarily focused on linear models (e.g., OLS, fixed effects panel model), this study introduces two non-linear models. XGBoost and Feedforward Neural Networks (FFNN), thus aiming to capture more complex relationships within the data. Transfermarkt, the market-leading information collector, has provided the core dataset containing match, transfer, and performance data. Auxiliary information has been obtained from various sources. 21 publicly listed football clubs comprise the study sample, covering a decade from 2013 to 2023.

Key findings reveal that team form and financial ratios are the most influential factors in predicting stock price changes, surpassing transfer-related variables, which have a less significant impact. The study employs SHAP (SHapley Additive exPlanations) values to interpret model predictions and assess feature importance, offering a robust, model-agnostic explanation of the results. The results indicate that while transfers play a role, the market is more responsive to the overall financial health and recent team performance. This suggests that investors place more weight on stable, long-term indicators rather than the immediate effects of player transfers.

In conclusion, this research contributes to the existing literature by applying cutting-edge machine learning techniques to a traditionally linear domain. It enhances predictive accuracy and offers more profound insights into the factors driving football stock prices. The findings have practical implications for investors, club management, and policymakers in the sports finance sector.

Contents

Chapter 1 Introduction

The transfer history of modern football dates back to 1893, when Aston Villa secured a player named Willie Groves for £100 (Chalabi, 2013). Since then, countless aspects have changed, and the current transfer system has profoundly shaped the game in the long run. The modern system – including the summer and winter transfer window, as we know it today – was introduced in the 2002/2003 season as part of a compromise agreement with the European Commission. The alternate solution was creating a structure similar to how most labor contracts are constructed, where notice periods need to be served, and players move freely. Generalized transfer windows, however, not only help teams plan for a set period, but they also ensure the opportunity for younger players to get the chance due to the lack of options for replacement (Premier League, 2017). However, exceptional cases (e.g., severe goalkeeping shortage) are worth mentioning where clubs can make an emergency signing (Bailey, 2023). It is also essential to note that free agents (players without an active contract) can be signed at any point of the year. This complicates the transferring & signing process because the current teams often only cancel their leaving players' contracts so that they can be signed off-window.

To put the relevance of this strategy in perspective, another critical aspect should be observed: the differing opening and closing dates for windows throughout countries. Besides the fact that many windows last 12 weeks, several shaping factors impact deadlines, such as national holidays or fairness reasons. Many leagues, for example, align their transfer window deadlines to the first or last league matches. This has also been voted in the Premier League for the 2018-19 season (Bailey, 2023). For example, as a broad overview of the leagues under the UEFA (Union of European Football Associations) Conferedation, summer windows open between the 11th of June and the 17th of July and close between the 12th of July and the 18th of September (Transfermarkt, Transfermarkt, 2024). This means that there are leagues with unmatching transfer windows. To overcome this issue, contract cancellation is the most used solution.

The list of public football teams has changed broadly in the past decades. An extensive boom can be observed in the late 1990s, from 1997 until the turn of the millennium. The number peaked at 37 and has decreased ever since (Aglietta, Andreff, & Drut, 2010). Currently, the most extensive list contains 21 active teams (Maci, Pacelli, & D'Apolito, 2020), which will serve as the sample of this study.

Even though the topic of transfers and publicly traded soccer teams' financial aspects have been widely researched in the past decade, studies are more likely to focus on certain influencing aspects: cultural differences (Bruin, 2023), Financial metrics and details (Maci, Pacelli, & D'Apolito, 2020), the relation between transfers and performance (Bhatia, 2020), differentiating player moves (sales, acquisitions, loans in and out) (de Bakker, 2016), match results (Galloppo & Boido, 2020), and investor attention (Kirchner, 2022). These aspects will be individually introduced in Chapter 2.1

The primary goal of this research is to determine the shaping power of football transfer details regarding stock price changes of football clubs. To do so, an empirical financial event study will be conducted. The key motivation is to introduce and test the performance of advanced machine learning methods regarding stock price prediction. It is vital to understand that prediction is a necessary step to extract and correctly understand variable importance. Thus, the more accurate the model's predictions are, the more precise results will be obtained to compare the influence of certain factors on football clubs' stock price change. As already discovered in the stock price prediction event studies, non-linear models operate with significantly more minor prediction errors than the most widely used linear regression (Agrawal, 2020). However, de Bakker from 2016 and Kirchner from 2022, the authors of the two most recent and most detailed studies that specialized on football transfers, solely relied on linear methods, OLS, and panel regression. Besides Agrawal – who compared the performance of linear regression with various nonlinear machine learning techniques –Pengfei and Xuesong arrived at a similar conclusion in 2020:

"Understanding the pattern of financial activities and predicting their development and changes are research hotspots in academic and financial circles. Because financial data contain complex, incomplete and fuzzy information, predicting their development trends is an extremely difficult challenge. Fluctuations in financial data depend on a myriad of correlated constantly changing factors. Therefore, predicting and analysing financial data are a nonlinear, time-dependent problem." (Pengfei & Xuesong, 2020)

As a result, what this study is expected to add to the existing study space are the following:

- increasing predictive power by including as many potential shaping factors as possible
- capturing nonlinear relationships using state-of-the-art machine learning methods

- determining the importance of transfer-related details compared to other pillars using an independent and uniform metric, stepping over linear coefficients.

To achieve this, the study is going to be based on the following main research question:

RQ: "How determinative transfers are in the case of publicly traded football teams' stock price change, compared to other shaping factors based on extensive transfer data and different machine learning methodologies?"

Besides contributing to the existing literature, this research has practical relevance in multiple ways. While opening up new directions in further investigating the football stock market, it also aims to exploit a yet unused method to deepen the comparative interpretability of the initial details of football transfers and other attributes of public football teams.

This paper is structured around five main chapters. After the introduction (1), section (2) reviews the existing literature, detailing used frameworks and results. It then introduces a conceptual model after breaking the research question into sub-questions. This is followed by the research methodology (3), which presents the base data, its processing, and the models used. The outcomes can then be found in the Results chapter (4), which breaks down the relevant findings and reflects on the implementation provided. Lastly, a general discussion and conclusion wrap the paper up (5) by offering recommendations and addressing limitations.

Chapter 2 Theoretical Framework

Chapter 2.1 Literature Review

Chapter 2.1.1 The Early Event Studies in Sports

Analyzing and measuring the impact of specific effects on sports teams dates back to the 1990s. One of the earliest papers on this topic studied the impact of prior success on current success (Mizruchi, 1991). The paper used a psychological approach, official NBA data, and logistic models to determine past results' impact on the upcoming match's outcome. The study found no significant relation between past and future results by relying on Pearson and Yates Chi-square values to test the hypotheses. Another exciting event study quantified the effect of artificial pitch surfaces on home team performance. Although results were insignificant, correlations were displayed regarding obtained points after incorporating artificial turf (Barnett & Hilditch, 1993). Later in the decade, a model-based project was published that served as a basis for sports teams' ratings for years. Even though investigating different rating methods was a topic going back to the 1930s, there was still significant room for improvement with the constant development of mathematical techniques and new ways of calculations coming to the surface. By introducing broken-down calculation flows for relevant linear models and several rating approaches based on linear regression, it was cited and referred to in many later papers (Massey, 1997).

Chapter 2.1.2 The Early Pioneers of Sports Event Studies on Economic Performance

One of the early pioneers of stock prediction was also published in parallel with the different event studies before the millennium. To put the psychological approach of past wins mentioned above into perspective, this paper also pinpoints the phenomenon called "hot hand", a common term describing a streak of good hits in various sports, which can be extended to describing consecutive good results (Wood, 1992). The author used a different common terminology to note a cause independent of chance, "luck". He states that gamblers tend to use both luck and chance as explanatory concepts. Based on these two phenomena, the paper consists of three different analyses to determine whether recent performance can be used to increase prediction. The first is a conditional probability analysis, where the two components are the following:

the probability of a win for game $i + 1$, given a win in game i,

(1)

$$P(W_{i+1}|W_i)$$

and the likelihood of a win for game n+1, given a loss in game n.

(2)

$$P(W_{i+1}|L_i)$$

Based on this, the ideal scenario is where

(3)

$$P(W_{i+1}|W_i) > P(W_{i+1}|L_i)$$

the probability of winning the game $i+1$ is more significant if game i has also been won (Wood, 1992).

The Wald-Wolfowitz Runs Test was performed for the second analysis, which creates a chain of symbols based on the pattern of losses and wins during the analyzed period (the 1988-89 regular season for multiple NBA, American League, and National League teams). Here, the more runs a team has, the fewer streaks it performs during the season. The third analysis was more elaborate, containing three variables of interest: outcome of the previous game, record against the opponent for the earlier games in the season, and home-court advantage.

After conducting the three analyses for all the selected teams, Wood performed the same conditional probability analysis on Dow Jones daily closing industrial averages between 1980 and 1989 obtained from Standard and Poor's Daily price record. While the structure remained the same, the conditions changed to the following:

the probability of an increase in closing price for day $i+1$ compared to day i, given a rise in closing price for day n compared to day $i-1$ in year y

(4)

$$P(U_{i+1}^{y}|U_i^{y})$$

and the probability of an increase in closing price for day $i+1$ compared to day i, given a decrease in closing price for day n compared to day $i-1$ in year y.

(5)

$$P(U_{i+1}^{y}|D_{i}^{y})$$

Based on this, Wood compared which probability is bigger. For example, in 1986

(6)

$$P\left(U_{i+1}^{1986}\middle|U_{i}^{1986}\right) = 0.515$$

and

(7)

$$P\left(U_{i+1}^{1986}\middle|D_{i}^{1986}\right) = 0.574$$

which means that

(8)

$$P\left(U_{i+1}^{1986}\middle|U_{i}^{1986}\right) < P\left(U_{i+1}^{1986}\middle|D_{i}^{1986}\right)$$

the conditional probability of day $i+1$ being an “up-day” after day i was a “down-day” is slightly higher than that of day i being an “up-day” as well (Wood, 1992).

Unfortunately, the research ends after these results and insights without deeply comparing the trends between sports and stock returns. Regardless, Wood’s publication played a crucial role in laying the foundation of trend-based stock-price analyses in the sports industry.

A similar study was conducted in 1993, focusing on the Boston Celtics (Frederick, Abbott, & Thompson, 1993). Instead of simply relying on independent trend analysis, the researchers sought causal relationships. However, with limited significance, they found a connection between match outcomes and the change in stock prices (Frederick, Abbott, & Thompson, 1993). In any case, it is essential to highlight that investment culture and the general sports business culture fundamentally differ in the two continents; thus, drawing conclusions based on this would not be thoughtful (Stelmach, 2023). In 2011, Maarten van Bottenburg also studied the path-dependence of European and American sports history and pointed to specific self-reinforcing mechanisms (e.g., how embedded sports are in the American culture compared to the European) that also spill over to the stock market (van Bottenburg, 2011). Regardless, Boțoc et al. researched the worth of a win expressed in stock returns using the GARCH-M (Generalized Autoregressive Conditional Heteroscedasticity in Mean) econometric model on three Italian teams. Even though they achieved significance at the 0.01 P-value (Boțoc, Mihancea, & Molcuț, 2019), they highlighted the heterogeneity of the EUROSTOXX market index, which was used as the index of interest. This tells us that even after considering all the differences, even single-match results might induce emotion-based investment in both continents.

Chapter 2.1.3 Modern Event Studies in Football Teams' Stock Price Prediction

Present-day event study methodologies regarding stock prices cannot be compared with the early ones that provided the initial momentum for this field of research. New predictor factors of stock price change have been discovered, identified, and measured, and the statistical model portfolio has also been expanded. Moreover, due to the change in the nature of predictive methodology, the form of expression for the variable of interest (e.g., daily closing stock prices) has also been fine-tuned.

In 2016, de Bakker introduced his Master's Thesis, which focused on the effect of transfers on stock prices for European-listed football clubs. In addition to being extensive and elaborative, this paper thoroughly concluded all the best practices and the underlying theory known thus far, using 15 listed teams.

Firstly, he addresses the question of mood proxying (de Bakker, 2016). Based on a study from 2007 by Boido et al., a significant relationship could be observed between Italy's FIFA World Cup win in 2006 and the increase in investor mood in the Italian stock market (Boido & Fasano, 2007). However, contrary results have been published in a study by the Oslo Business School in 2020 (Wilhelmsen, 2020). This paper uses Google search volume as a metric for investor mood proxy on the Nasdaq Copenhagen market (OMXC20 Index). Even though there is a somewhat significant relationship between abnormal search volume index (ASVI) and trading activity, the number of searches on company ticker and/or company name does not indicate the direction of change in stock price change. Similar results have been discovered by Kim et al. on the Norwegian stock market, with even more minor significance and robustness (Kim, Molnár, Lučivjanská, & Villa, 2019). In conclusion, while the Google search index might be more or less predictive regarding trading volume, it is not confirmed or supported that it affects the amplitude of change in daily prices.

De Bakker also considers the announcement effect. Financial announcements in other industries and the transfer news in football have at least one fundamental similarity. Just as "insider trading" is present in all sectors, since specific traders get to financial results earlier than the majority, transfers are also often leaked in some ways, which shrinks the effect of the official announcements (Stadtmann, 2006). The most prominent example of this phenomenon is Fabrizio Romano, who uses his private contact network to obtain transfer news as early as possible from the outside world (Romano, 2024). His catchphrase “Here We Go” is well known in the football community, and the effect of his tweets would deserve a separate study. Another study focuses solely on information leakage, where an apparent reaction was discovered between the early information about a transfer and the increased trading volume, which is assumed to happen mainly due to the predominance of emotional investors compared to rational ones (Fűrész & Rappai, Information Leakage in the Football Transfer Market, 2022). However, as most studies on the topic do, different types of cumulation of stock price change can control for the announcement effect. Using abnormal returns as a dependent variable became a standard approach in the study space. Abnormal returns are the performance of an instrument that exceeds what is expected based on certain conditions (Corporate Finance Institute Team, 2024). It was used by de Bakker in 2016, Galloppo et al. in 2020, and Navest, who analyzed the effect of match results on stock returns using abnormal returns (Navest, 2023). This study will elaborate on the computation and exact way of using abnormal returns.

Match performance is the next important factor when analyzing football stock prices and has also been widely studied. Stadtmann also found that while general results affect the market price, unexpected losses or wins have a much more significant effect (Stadtmann, 2006). He also used the results of the rival team(s) to be even more detailed; however, a large amount of data is required (Stadtmann, 2006). A critical study was conducted by Bell and Brooks in 2012, who found that match results have a shaping effect on a club’s share price and proved that the results of games played in mid-season are less relevant from the market’s point of view. They say it is because matches played in the middle part of the season are considered less important, and general investor attention might decrease (Bell & Brooks, 2012).

Beyond sports performance, financial metrics – just as in other industries – play a significant role in shaping stock prices for football teams. A significant positive link was found between total revenue, return on equity, net income, and stock prices (Maci, Pacelli, & D'Apolito, 2020).

The models used and their outcomes should also be investigated to provide a final look at the results of the two most recent and most complex papers closest to this study’s topic. De Bakker used OLS regressions on different samples from his overall dataset to investigate whether results would differ based on certain factors (e.g., age, transfer value, etc.). With this method, he could differentiate between specific details of the transfers and decide which ones were more dominant based on coefficients and significance (de Bakker, 2016). The adjusted R-squared values of his models varied between 0.02% and 5%. On the contrary, Kirchner based his calculations on a fixed effects panel model, which – unlike OLS methodologies – accounts for cross-sectional correlations (Kirchner, 2022). Kirchner's results, while also showing significance for the absolute fee or the Champions League control variable, R-squared values also move around the 1% mark (Kirchner, 2022), which means that the models can only capture a minor part of the total variability.

Chapter 2.2 Conceptual Overview

Chapter 2.2.1 Breakdown of the research question

To fully understand the purpose of this study and why the thesis question is relevant, it is best to define subquestions that all refer back to the main thesis question and the three main goals introduced in Chapter 1.

SQ1: "What are the main shaping factors for football teams' stock price change?

The initial step is to create an overall, complex model containing as many influencing factors as possible to determine how vital transfers are in comparatively determining the change in stock price. Its results give us the answer to the first subquestion.

SQ2: "What details of transfers are the most important in impacting stock price change?

After the overall nominal importance values are computed, the individual details of transfers can be ordered based on shaping power.

SQ3: "In stock performance prediction, to what extent have modern machine learning methodologies increased predictive power compared to linear models?

After executing machine learning methods, their performances will be compared based on an accuracy-related metric to see which type of model best captures the relationships in the dataset.

Chapter 2.2.2 Conceptual Model

Based on the in-depth analysis of the studies mentioned in Chapter 2.1, the following structure has been assembled for this research, displayed by a conceptual model in Figure 1. To help further studies on the topic and better understand what model type is preferred in the process, two models will be used and compared that incorporate different approaches in terms of calculation. Since mood proxying has controversial results and findings, it will not be part of this study, and the announcement effect will be left out due to the form to which stock price change is expected to be transformed. Moreover, as this is only the high-level visualization of the project, it is essential to note that specific, more operative tasks (e.g., creating intercepts and controls for years or teams) are not highlighted in the conceptual model but will be incorporated in the final model. As shown below in Figure 1, the same data structure is planned for both models (determined later), which is essential for a meaningful and methodologically appropriate comparison.

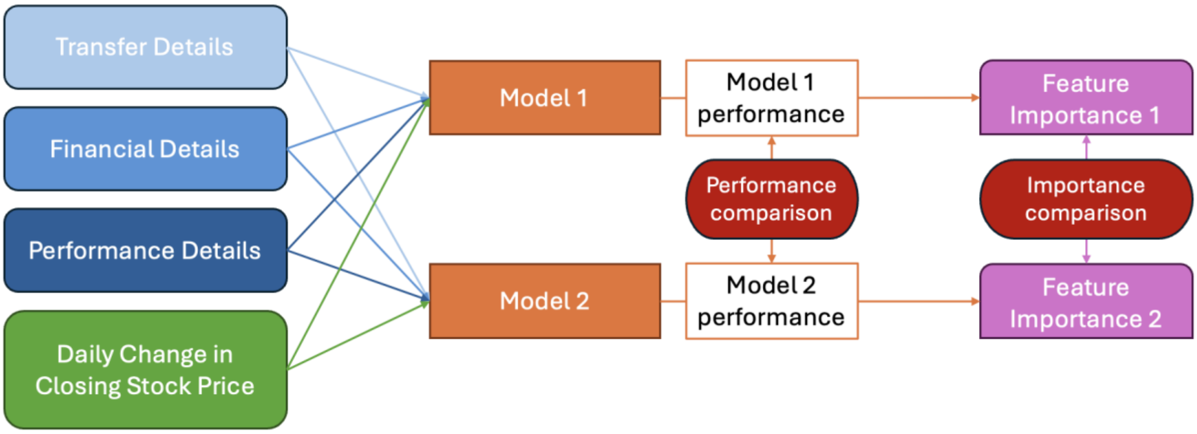

Figure 1: The conceptual model explains the high-level relationships and processes to get to the final results, which are the model performances, importance, and comparisons

To conclude the mapping process of the current research space and Chapter 2, although recent research raised the progress of the topic to the next level, the lack of involvement of advanced machine learning methodologies to capture nonlinearities and more metrics to measure shaping power, which creates a clear position for this study in the research space.

Chapter 3 Data & Methodology

In this section, the study's fundamental parts will be discussed. First, the different pillars of the dataset will be detailed, elaborating on their collection method, characteristics, and place in the study. After this, the machine learning models will be selected and reviewed along with the result interpretation strategy.

Chapter 3.1 Data Collection

Data collection is the most crucial part of the study. Validity is a pivotal aspect to remember, and excessiveness is also a fundamental part once the presence of nonlinear relationships is expected in the data. Even though machine learning methods are, by definition, better at capturing nonlinearities, they also need more data to do so (Brownlee, How Much Training Data is Required for Machine Learning?, 2019). Thus, the best alternative is to obtain all corresponding information from one source to ensure the data's quality and extensiveness. In this study, Transfermarkt provided the transfer-related datasets. According to The New York Times's boiled-down wording, Transfermarkt is a website "where people go to find and discuss information about soccer players." (Smith, 2021). As a more elaborate description from the same article, it provides a source of knowledge, a point of reference, and, through its humming chat boards, a place for a community of like-minded individuals to gather (Smith, 2021). However, this is not what Transfermarkt was initially known for. At first, the website tried to determine the worth of as many player's value as possible. To do so, Transfermarkt built out a network of thousands of volunteers worldwide, thus gathering the required information (Smith, 2021). By 2020, however, the website counted 1.5 million football matches in its database, 760,000 player profiles, and 75,000 clubs (Transfermarkt, News, 2020).

As mentioned before, 21 current public teams will create this study's sample based on a past study's own processing (Maci, Pacelli, & D'Apolito, 2020). Table 1 shows these teams.

Club name	**Short name (as referred to)**	**Country**
Aalborg Boldspilklub	Aalborg	Denmark
Aarhus Gymnastikforening	Aarhus	Denmark
Allmänna Idrottsklubben	AIK	Sweden
Amsterdamsche Football Club Ajax	Ajax	Netherlands
Beşiktaş Jimnastik Kulübü	Besiktas	Turkey
Ballspielverein Borussia 09 e.V. Dortmund	Dortmund	Germany
Brøndbyernes Idrætsforening	Bröndby	Denmark
Celtic Football Club	Celtic	Scotland
Futebol Clube do Porto	Porto	Portugal
Fenerbahçe Spor Kulübü	Fenerbahce	Turkey
Galatasaray Spor Kulübü	Galatasaray	Turkey
Juventus Football Club	Juventus	Italy
Manchester United Football Club	Manchester United	England
Olympique Lyonnais	Lyon	France
Ruch Chorzów	Chorzow	Poland
Sporting Clube de Braga	Braga	Portugal
Sport Lisboa e Benfica	Benfica	Portugal
Società Sportiva Lazio	Lazio	Italy
Silkeborg Idrætsforening	Silkeborg	Denmark
Sporting Clube de Portugal	Sporting	Portugal
Trabzonspor Kulübü	Trabzonspor	Turkey

Table 1: The list of teams considered for the study's sample and the teams' respective countries.

Chapter 3.1.1 Transfer Data

The first core dataset provided by Transfermarkt includes the necessary transfer-related information. The study's timeframe is between 04-09-2013 and 18-09-2023. The specific days are chosen based on the trading day schedule and the availability of transfer data. The transfer dataset is responsible for displaying all incoming and outgoing loans and transfers regarding all target teams in Table 1. In the given timeframe, there are 9,827 transfers for the 21 clubs. The balance of in and outgoing transfers and loans can be observed in Table 7. To mention the most critical variables, the dataset contains the incoming & outgoing clubs' names and Transfermarkt IDs (to match in other datasets), the player's age, position, market value, transfer fee, the date of transfer, and a marker to see if it was a transfer or a loan.

The balance analysis in Figure 8 shows the differences between each team's incoming and outgoing transactions per type (transfer, loan). Interestingly, there is a solid inverse proportionality between the transfer and loan balances.

However, this phenomenon is relatively standard in the football market and is caused by several reasons. As per Günter and Vischer, it is essential to understand that both parties' (the loaned player and the owner club) agreement is required for a loan to happen since it changes the entire concept (Günter & Vischer, 2024). Firstly, loaning is the primary alternative if the upfront purchase of a player exceeds the loaning team's financial capabilities. Secondly, bigger teams often loan young talents to save a portion of their salary and other costs without letting them go, so the development made by the loaning squad can still benefit them (Günter & Vischer, 2024). These two factors lead to a trend where bigger teams loan players to smaller ones coming from their academy or freshly transferred and loan back from other teams to fill occasional gaps in their squad. This is important because, in the dataset, the transfer fee is 0 for all the loans and the special transfers (e.g., promoting from academies), which might introduce some bias, but since there are more than 2,000 loan entries, leaving them out would be an immense sacrifice from a potential model performance point of view.

Chapter 3.1.2 Match Data

Match details and results have also been requested from Transfermarkt. This is required to create further variables based on past match results and the details of these matches. An overview table for each team can be observed in Table 8, detailing win percentages and the number of matches. The third column (Distribution of Total Games) aims to display the distribution of teams in the dataset since each observation represents a transfer regarding one of the teams of interest. To further examine the distribution of games, Pearson's moment-based formula has been used to calculate skewness, which is the following:

(9)

$$Skewness = \frac{n}{(n-1)(n-2)}\sum_{i=1}^{n}\left(\frac{x_i - \bar{x}}{s}\right)^3$$

Where n is the number of instances, x_i is each individual data point, $\bar{x}$ is the mean of the data points, and s is the standard deviation of the data points (Doane & Seward, 2011). Based on this formula, the skewness of team distribution regarding played games in the dataset is -0.48. There is no standard interpretation of Pearson’s skewness, but based on the widely used rule of thumb, between -0.5 and 0.5, the distribution can be considered slightly skewed or symmetric (Tuychiev, 2023).

In addition to the competing teams and the results, the critical variables are the type of competition, the stage of the competition (if not a league match), the goals scored, and the actual position of both clubs. The engineered variables using this dataset will be further elaborated.

Chapter 3.1.3 Team Financial Data

For financial data, all available balance sheets for each team in the study’s time window have been obtained from Orbis. Orbis is a leading platform offering comparable data resources for private and listed companies owned by Moody’s (Moody's, 2024). It is crucial to mention that Besiktas has been removed from this study's team pool due to the lack of these data.

Maci et al. analyzed the financial determinants of football stock prices. Their study included revenue, total assets, financial debts, return on equity, index of revenue diversification, and net income (Maci, Pacelli, & D'Apolito, 2020). To further enhance their approach, this study aims to pick the financial variables so that they account for different types of financial aspects. The main groups of financial ratios are profitability ratios, which measure the generated return; financial risk ratios, accounting for measuring exposure to risks; turnover ratios, which are responsible for supplementing the first two ratios; valuation ratios, which estimate the fair value of any business and cash-flow based ratios that examine company performance based on cash-flow statement (Welc, 2022). After carefully analyzing the ratios available from Orbis, all teams have enough data to account for the first three types of ratios. Thus, the chosen financial ratios are represented in Table 2.

Financial variable	Ratio type
Return on Assets	Profitability ratio
Profit Margin	Profitability ratio
Current Ratio	Financial Risk Ratio
Solvency Ratio	Profitability ratio
Net Assets Turnover	Turnover Ratio

Table 2: The financial variables used and their respective ratio types.

Chapter 3.1.4 Stock Data

Yahoo Finance is used to obtain the daily closing prices in the target time window (04-09-2013 and 18-09-2023). Yahoo Finance is a convenient tool due to its dynamic range setting option. To perform the currency conversion, daily exchange rates are obtained from the European Central Bank Data Portal for the following exchanges: US dollar/Euro, UK pound sterling/Euro, Polish zloty/Euro, Danish krone/Euro, Turkish lira/Euro, and Swedish krona/Euro.

Furthermore, to perform the calculation of abnormal returns, a reference index is required to be paired for each team since the concept of abnormal returns is based on comparing the initial stock's performance to a reference index that is closely related (The Economic Times, 2024). While de Bakker (2016) used the MSCI all-country midcap index for all the teams, this can be improved. This study implements a three-step strategy to find the suitable reference index for each team. First, it is searched for in the Orbis database. If it is not displayed there, then Yahoo Finance's internal performance measuring system is used to identify the reference index. There were two cases, however, where information could not be deducted from either of the two websites. In this case, the Infront Analytics website was used. Table 4 shows the reference indices for all teams, the source of identification, and the data source.

Chapter 3.2 Method

After data collection, everything is set for the method elaboration. As discussed in Chapter 1, one of this study's key points is introducing machine learning methodologies to increase model complexity, predictive power, and interpretability. To completely understand the described procedures below, Table 9 introduces the variable set beforehand.

Chapter 3.2.1 Models

As detailed in Chapter 2.1.3, all the recent studies use different linear regression models. De Bakker (2016) uses OLS, while Kirchner (2022) uses a fixed effects panel model to account for cross-sectional correlation. While linear regression is a sound alternative for conducting statistical analyses, some limitations can be overcome by switching to more complex methods. First, as was discussed in Chapter 1, the event study of football transfers on stock returns includes nonlinear relationships, which linear models do not initially capture (Pengfei & Xuesong, 2020). To test if this is the case, a test dataset has been assembled without adding any extra variables (financial ratios, team-based variables, or performance-based variables) and two higher-order terms for the sake of testing to run an OLS regression and a Regression Specification Error Test (RESET), which assesses the significance of a regression of residuals on a linear function of vectors (Ramsey, 1969).

To understand how the test works, take a linear regression with the following form:

(10)

$$Y = X\beta + \varepsilon$$

Where Y is the dependent variable (outcome), X is a matrix of independent variables (predictors), β is a vector of coefficients and ε is the vector of residuals (errors). After this linear regression has been run, in the prediction step the predicted outcomes ($\hat{Y}$) and errors ($\hat{\varepsilon}$) are calculated using the fitted coefficients ($\hat{\beta}$):

(11)

$$\hat{Y} = X\hat{\beta}$$

(12)

$$\hat{\varepsilon} = Y - \hat{Y}$$

After making predictions, the RESET test introduces higher-order terms of the model's predicted outcomes (or fitted values) as additional regressors. These terms are squared values ($\hat{Y}^2$) and cubed values ($\hat{Y}^3$) in this study. After introducing these terms, the original base model is augmented with the higher-order terms, and the model becomes the following:

(13)

$$Y = X\beta + \theta_1\hat{Y}^2 + \theta_2\hat{Y}^3 + \upsilon$$

In this model θ_1 and θ_2 are the coefficients of the higher-order terms and υ is the new error term. The next step is hypothesis testing using the newly created regression model (Ramsey, 1969). The test states two hypotheses:

H0: The original model is correctly specified, meaning that the predicted errors of the augmented model do not differ from the predicted errors of the original model (higher-order terms are not omitted in the original model, appearing as part of the error term).

H1: The original model is incorrectly specified, so the errors of the augmented model are significantly lower than those of the original model (due to including significant higher-order terms). To test the hypotheses, RESET applies F-statistic, which is calculated in the following way :

(14)

$$F = \frac{(RSS_{original} - RSS_{augmented})/q}{RSS_{augmented}/(n-k-q)}$$

Where $RSS_{original}$ is the residual sum of squares from the original model, $RSS_{augmented}$ is the residual sum of squares from the augmented model, q is the number of new terms added, n is the number of observations, and k is the number of independent variables in the original model. The RESET results are shown in Table 5. The F-statistic measures the overall significance of the additional higher-order terms (Ramsey, 1969). A high F-statistic, such as 35.909, means these terms significantly improve the model. The p-value measures the probability that the observed data would occur if the null hypothesis were true. In this case, the p-value is extremely low, which means that H0 can be rejected. The degrees of freedom mark that two additional terms have been added, and the number of observations minus the number of estimated parameters in the original model is 7,310. As H0 can be rejected, the residual vectors of the two models significantly differ, which means – combined with the interpretation of the F-statistic – that the higher-order variables have a serious added value to the model. This can also be interpreted as there are indeed nonlinear relationships in the dataset, indicating the use of more advanced models to capture these higher-order relationships.

Chapter 3.2.1.1 Model Selection

To choose which models to use, firstly, the research of Yisheng Li is considered. Li studied the critical drivers for soccer player valuation in 2021 and incorporated predictions using the following models: multiple linear regression (MLR), single decision tree (DT), random forest (RF), support vector regression (SVR), and extreme gradient boosting

(XGBoost) (Yisheng, 2021). The results of Li's predictions can be found in Table 10. Based on his comparison, XGBoost performed the best with the lowest RMSE and the highest adjusted R^2.

RMSE stands for Root Mean Square Error, with the following formula:

(15)

$$RMSE = \sqrt{\sum_{i=1}^{n} \frac{(\hat{y}_i - y_i)^2}{n}}$$

Where $\hat{y}_i$ is the predicted value of the i -th observation, y_i is the actual value for the -th observation, and n is the total number of observations. Per definition, RMSE is the square root of the average squared differences between the predicted values (Hodson, 2022).

At the same time, R^2 is amongst the most widely used metrics for model evaluation as well. R^2 is an estimate of the proportion of variance in the outcome variable explained by the predictors in the sample. However, it tends to be biased upwards compared to the true proportion of variance explained in the population (Miles, 2005). The mathematical formula is the following:

(16)

$$R^2 = \frac{SS_{explained}}{SS_{total}} = 1 - \frac{SS_{residual}}{SS_{total}}$$

SS_{total} is the total sum of squares that measures the total variability in the dependent variable around its mean (Miles, 2005):

(17)

$$SS_{total} = \sum (y_i - \bar{y})^2$$

Where y_i is an observed value of the dependent variable, and $\bar{y}$ is the mean of the dependent variable. Residual sum of squares ($SS_{residual}$) measures the variation that is not explained by the regression model (Pennsylvania State University, 2018):

(18)

$$SS_{residual} = \sum (y_i - \hat{y}_i)^2$$

Where $\widehat{y_i}$ is a predicted value of the dependent variable. To calculate $SS_{explained}$ (the variation in the dependent variable that the model explains), the following formula should be incorporated (Pennsylvania State University, 2018):

(19)

$$SS_{explained} = SS_{total} - SS_{residual}$$

Moreover, the added value of this study could be even further increased by comparing multiple state-of-the-art methods that have already been tested for event studies or even stock price prediction but not in this specific field. Another comparative research was done by Orimoloye et al. (2020), which analyzed the performance of feedforward neural networks and other shallow architectures. It turned out that a support vector machine (SVM) performs better with specific time windows than different types of neural networks (LSTM, GRU, DNN) (Orimoloye, Sung, Ma, & Johnson, 2020). The results also display that LSTM neural networks that are initially more suited for event studies cannot perform better than feedforward neural networks. As a conclusion of this study, general machine learning methods and neural networks are worth comparing. Thus, the two models used in this study will be an XGBoost and a feedforward neural network.

Chapter 3.2.1.2 XGBoost

XGBoost is an end-to-end tree-boosting method developed by Chen & Guestrin in 2016. It builds on tree-boosting models known from the past, with the primary goal of developing those models' shortcomings. XGBoost's derivation follows the same structure as other gradient tree boosting. Moreover, the second-order method traces back to Friedman et al. (2003), based on Friedman's initial study from 2001 (Chen & Guestrin, 2016). Jerome Friedman invented gradient boosting in his seminar paper in 1999, which he updated in 2001 (Friedman, 2001). Friedman introduced the concept of gradient boosting as a stagewise additive model that builds the final model by iteratively adding base learners (e.g., decision trees) to minimize a loss function. This process uses gradient descent in function space rather than parameter space. In the first-order method, only the loss function's first derivative (gradient) is used to update the model. However, as described by Friedman, the second-order method incorporates both the gradient and the second derivative (Hessian) of the loss function (Friedman, 2001).

The main goal of tree-boosting methodologies is to minimize the regularized learning objective (Chen & Guestrin, 2016). A tree ensemble model uses the following formula for prediction:

(20)

$$\hat{y}_i = \phi(x_i) = \sum_{k=1}^{K} f_k(x_i), \quad f_k \in \mathcal{F}$$

Where $\mathcal{F}$ is the regression space (CART), each f_k corresponds to an individual tree system and K is the number of additive functions used to predict the outcome. XGBoost uses a gradient objective function to reach to an optimum. This objective function is as follows:

(21)

$$\tilde{\mathcal{L}}^{(t)} = \sum_{i=1}^{n} \left[g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i) \right] + \Omega(f_t)$$

Where g_i represents the first-order gradient (derivative) of the loss function concerning the prediction from the previous iteration, h_i represents the second-order gradient (second derivative) of the loss function for the prediction of the prior iteration, $\Omega(f_t)$ is the regularization term to control the complexity of the model, and $f_t(x_i)$ is the prediction from the new tree added at iteration t. Besides the efficient additive manner of the training process, XGBoost operates with cache-aware access, which enables it to work more fluently with large datasets. It also includes a regularization term in the objective function, which prevents overfitting (Chen & Guestrin, 2016).

Chapter 3.2.1.3 Feedforward Neural Network

Feedforward neural networks (FFNN), or multilayer perceptrons, are networks in which the directed graph establishing the interconnections has no closed paths or loops as there are for recurrent neural networks (RNN). The model can be traced back to McCulloch and Pitts in 1943, who found that the behavior of every net can be described based on the human nervous system's characteristics and attempted to translate and describe this phenomenon to the language of logical expressions (McCulloch & Pitts, 1943). Not only is it a widely used method in image recognition, natural language processing, and predictive analytics, but it also shifted away from black-box predictions toward statistical analyses due to the ability to perform statistical inference and covariance-effect visualizations (McInerney & Burke, 2023). FFNNs comprise an input layer, one or more hidden layers, and an output layer. Each node in one layer is connected to every node in the next layer, with no cycles or loops. The visualization can be observed in Figure 10 (De Mulder, Moens, & Bethard, 2014). FFNNs are typically trained using backpropagation, a supervised learning algorithm. This involves adjusting the network's weights to minimize the error between predicted and actual output. The optimization process employs gradient descent or its variants.

FFNNs can be incorporated and assembled based on multiple objective functions, depending on the goal and the type of task (regression or classification). The two most used alternatives for regression tasks are the mean squared error (MSE) and the root mean squared error (RMSE), and for classification, the cross-entropy loss (Ebert-Uphoff, et al., 2021). The formula of RMSE can be seen in Equation (15). Since this study focuses on a regression task, MSE and RMSE will be compared and decided between. The differences between the two approaches are marginal, and there is no detailed literature about them. By some, RMSE is an extension of MSE since root calculation means that RMSE units are the same as the dependent variable's units (Schneider & Xhafa, 2022). If this line of thought is continued, RMSE seems to be an ideal choice in this case since the study aims to compare the two models' performance on the same dataset since both models require and can handle the same structure. If the dependent variables for the two models were different, comparability was more complex, and relative errors obtained by MSE were more sufficient for comparison purposes. Thus, RMSE will be used as an objective function in this study since RMSE is often used as a training evaluation metric (Amjad, et al., 2022). So, while it serves different purposes for the two models, it is an ideal tool for this study.

The components of FFNNs are the following:

- Inputs: The input features (explanatory variables) for the neural network,
- Weights: The parameters that are learned during training,
- Biases: Additional parameters that are added to the weighted sum before applying the activation function,
- Activation Function: A nonlinear function applied to the weighted sum of inputs and biases (Tsoulos, Gavrilis, & Glavas, 2008).

Firstly, the model conducts the computation of the hidden layer's nodes:

(22)

$$z^{(l)} = W^{(l)}a^{(l-1)} + b^{(l)}$$

(23)

$$a^{(1)} = f(z^{(1)})$$

Where $W^{(l)}$ are the weights matrix for layer l, $b^{(l)}$ is the bias vector for layer l, $z^{(l)}$ is the linear combination of activations from the previous layer and weights for layer l, and $a^{(l)}$ is the activation of layer l (Nielsen, 2013).

As an addition, even though there are dozens of activation functions developed, nearly all of them are based on four main ones, which are the following (Dubey, Singh, & Chaudhuri, 2022):

(24)

$$Sigmoid\colon \sigma(z) = \frac{1}{1 + e^{-z}}$$

(25)

$$ReLU\ (Rectified\ Linear\ Unit)\colon ReLU(z) = \max\,(0, z)$$

(26)

$$Tanh\colon \tanh(z) = \frac{e^{z} - e^{-z}}{e^{z} + e^{-z}}$$

(27)

$$Softmax\colon \text{Softmax}(z_i) = \frac{e^{z_i}}{\sum_j e^{z_j}}$$

To conclude, Feedforward Neural Networks (FFNNs) and XGBoost are powerful machine learning models, each with unique strengths. While FFNNs excel in tasks requiring complex feature interactions and high-dimensional data, XGBoost is highly effective for structured data with its ability to handle missing values and prevent overfitting through regularization. These models provide a comprehensive toolkit for tackling diverse predictive analytics challenges, combining deep learning's flexibility with gradient boosting's robustness and efficiency.

Chapter 3.2.2 Variable Importance Interpretation

Since one of the study’s goals is to analyze how variable importances compare across the two chosen models, a generalized, uniform methodology is required. Multiple options exist when determining critical features based on model performance change, such as SHAP

(SHapley additive exPlanations) values, Gini impurity, F1-measure, recall, or AUC (Meng, Yang, Qian, & Zhang, 2020). However, the study of Meng et al. (2020) also states that the technology of SHAP ensures feature consistency and model stability, which makes it an even more ideal choice. Also, comparing the produced output of the listed alternatives, SHAP has the most customizable framework, resulting in a compact plot with extensive explanation and insights, especially for regression tasks. Considering all these aspects, SHAP values will be used in this study.

SHAP values trace back to Shapley values used in game theory. Shapley values provide a way to distribute a total gain (or payout) to individual players based on their contributions to the total gain (Kenton, 2023). In the case of SHAP values and machine learning methods, the "players" are the model's features, and the "payout" is the prediction made by the model. SHAP values attribute the change in the model output to each feature by considering all possible combinations of features. This means calculating the contribution of a feature by averaging over all possible subsets of features (Lundberg & Lee, 2017).

Regarding the calculation of the SHAP, the key idea is how the Shapley values are calculated for a given feature i:

(28)

$$\phi_i = \sum_{S \subseteq N \setminus \{\mathrm{i}\}} \frac{|S|!\,(|N| - |S| - 1)!}{|N|!} [f(S \cup \{i\}) - f(S)]$$

Where N is the set of all features, S is a subset of features not containing feature i, $f(S)$ is the model's prediction using the feature subset S, $f(S \cup \{i\})$ is the model's prediction using the feature subset S with the feature i added, and ϕ_i is the Shapley value for feature i, representing it's contribution to the prediction (Lundberg & Lee, 2017). SHAP extends this rationale for efficient implementation in machine learning methods. Since calculating the exact Shapley value in the case of big datasets is infeasible, SHAP's approximation algorithms help handle this issue (Lundberg & Lee, 2017).

Chapter 3.2.3 Dataset Assembly

After introducing and elaborating on all the initial data points and their collection method, the last step is to describe the final dataset's assembly procedure, which will serve as the input for the two models.

Chapter 3.2.3.1 Feature Engineering & Preparation

Firstly, the engineered features will be introduced in detail. In this study, the main goal when creating new variables is to synthesize or transform the original variables so that the fewest explanatory variables are included in the final dataset without losing information or introducing bias. This is required since the number of observations (9,202) is limited, and dataset complexity (regarding the number of explanatory features) should be minimized.

Chapter 3.2.3.1.1 Team Form

Assumably, the most crucial variable created is the current form of the teams of interest at the time of a transfer. This is a ratio between 3 and -3, thus ensuring comparable feedback on how each team performed in a window of eight matches at every point in time. Multiple preliminary calculations were required for this. First, the goal difference (GD) of every game was calculated and then weighted based on the difference and whether the game was played home or away from the team of interest's angle. Also, to compute match importance (MI), each game has been weighted based on the competition type and progress to order an importance value (IV) to each game. Table 11 and Table 12 show the weighting details.

To produce the actual form based on the last eight matches, a rolling average-based formula has been created:

(29)

$$RA_i = \frac{1}{\min\,(w_i + 1)} \sum_{j=\max\,(0,i-w+1)}^{i} IV_j$$

Where RA_i is the calculated rolling average of match i, j is the index of the matches within the rolling window used to calculate the average importance value (where j iterates over the matches within the window) from match $i - w + 1$ to match i, and w_i the window size for game i, so that in case the available window size is less than 8, it has been adjusted accordingly.

Chapter 3.2.3.1.2 Past Days Since Last Win

Based on the thorough analysis of the literature space in Chapter 2.1, the importance of a win is unquestionable. The number of days past since the last win has been calculated to control for this in the dataset. Thus, the temporary effect caused by a win and its diminishing effect can be caught. Given:

- d_i: the date of match i,

- T_i: the team playing in match i,
- R_i: the result of match i (either “win”, “loss” or “draw”),
- $LW\ D_{T_i}$: the date of the last win for team T_i

The past days since the last win for team T for match i is calculated as follows:

- If $R_i = "win"$:

(30)

  - $DS_i = (d_i - LW\ D_{T_i}).days$

(31)

  - $LW\ D_{T_i} = d_i,$
- If $R_i \neq "win"$:
  - $DS_i = (d_i - LW\ D_{T_i}).days$

Chapter 3.2.3.1.3 Transfer Window

Another key engineered variable from the transfer point of view is the transfer window. This study separates the calendar year into two categories: winter window (noted as w) and summer window (noted as s). This was required to control for the seasonality present in the transfer market. Usually, both the activity and the amplitude of transfers are significantly lower in the window transfer market, presumably due to the nature of contracts and best practices on the market (Hashi & Kroken, 2017). It is also crucial when computing the features above since (combined with the season marker) it serves as a flag for the season changes, where both team form and past days reset. This is why not only the transfer windows are market, but the entire year, to handle the events happening before the actual transfer window and the transfers included as one, thus pairing the effects and the transfers together.

The rules for the assignment are the following:

- if the month is between March (inclusive) and August (inclusive), the match is categorized as summer (“s”),
- if the month is February and the day is less than 15, the match is categorized as winter (“w”),
- if the month is February and the day is 15 or later, the match is categorized as summer (“s”),
- Otherwise, the match is categorized as winter (“w”).

However, there are big differences between the start and end date of transfer windows for each league relevant to this study, so the determined change points are only partially exact.

Chapter 3.2.3.1.4 Cumulative Abnormal Returns – Dependent variable

As highlighted in Chapter 2.1.3, abnormal returns will be the dependent variable in this study's modeling, which follows the standards of the recent, most advanced papers. Moreover, cumulative abnormal returns will be applied to capture the total abnormal returns over a set period. As already introduced in Table 4, a fitting reference index has been ordered for each team, which will serve as the reference indices for the abnormal return calculation, which stands below:

For each team T, a linear regression model was fitted to estimate the relationship between market returns (R_m) and the team's stock returns (R_T):

(32)

$$R_T = \alpha + \beta R_m + \epsilon$$

Where α is the intercept, β is the slope and ϵ is the error term.

After this, using the estimated parameters α and β, calculate the expected return (ER) for each observation:

(33)

$$ER = \alpha + \beta R_m$$

Finally, the abnormal return (AR) for each observation is the difference between the actual stock return and the expected return:

(34)

$$AR = R_T - (\alpha + \beta R_m)$$

After the actual abnormal returns for all transfers have been obtained and the window size has been determined (e.g., -2, +2), CAR can be computed by a cumulation:

(35)

$$CAR_{[t_1,t_2]} = \sum_{t=t_1}^{t_2} AR_t$$

Where $[t_1, t_2]$ is the set event window (Corporate Finance Institute Team, 2024).

Since the goal of this study is not to find the ideal CAR window or to compare how specific effects differ between different CAR windows, it will use only one CAR window as the independent variable, CAR(-2,2). Important to note that because the initial stock prices' frequency are daily, the window size also represents days.

Chapter 3.2.3.1.5 Interaction Term

Relying on the preliminary assumption that market value and transfer fee could significantly shape cumulative abnormal returns, an interaction term will be introduced to the final feature set, which is the product of the two variables. Involving this variable will provide a more nuanced picture of the two variables' combined effect on stock performance.

Chapter 3.2.3.1.6 Standardization

Normalization is a technique that is considered best practice when preparing datasets for any machine learning task. Even though tree-based models do not require it based on the theory (Chen & Guestrin, 2016), it is still often used. On the other hand, this is not the case for FFNNs, where it is a key step in the preparative phase (Bhanja & Das, 2019). Based on this, the Z-score standardization has been executed on the dataset, which is one of the most common normalization techniques, and is computed for the value x_i in the following way:

(36)

$$x_{i_{norm}} = \frac{x_i - \mu(X)}{\delta(X)}$$

Where $\mu(X)$ is the mean value of feature X, and $\delta(X)$ is the standard deviation of feature X.

Chapter 3.2.3.1.6.1 One-hot Encoding

Finally, the last preparative step is to convert all the categorical variables into a form both models can handle. Both XGBoost and FFNN require and benefit from one-hot encoded variables (Chen & Guestrin, 2016) (Seger, 2022). This means that each categorical variable gets divided into as many binary variables as the variable's categories. When performing regression-based analyses, leaving a reference category out is best practice, thus making coefficient interpretation more efficient. Even though it is not strictly required when working with SHAP values, it is still acceptable to avoid multicollinearity when others can perfectly predict a variable.

Chapter 3.2.4 Summary Statistics & Data Map

Before diving into the results in detail, it is essential to see the nature of all the numeric variables in the final dataset. One-hot encoded variables (team flag, season flag, window flag, transfer type flag, position flag, and transfer direction flag) have not been included.

	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
age	9202	24,377	4,196	16	21	27	43
market_value	9202	3381199,739	7454491,037	0	250000	3000000	150000000
transfer_fee	9202	1771408,739	7390046,709	0	0	0	135000000
team_form	9202	0,755	0,889	-3	0,25	1,375	3
days_since_last_win	9202	15,395	20,517	2	6	15	74
assets_turnover	9202	2,000	3,109	0,124	0,781	1,884	23,27
current_ratio	9202	0,752	0,546	0,040	0,378	1,027	3,745
ROA	9202	-4,735	18,959	-87,509	-11,776	5,054	91,309
profit_margin	9202	-8,857	29,614	-99,196	-23,116	9,694	95,932
solvency_ratio	9202	7,803	42,698	-99,793	-15,385	41,781	81,053
fee_market_interaction	9202	4,59443E+13	4,00673E+14	0	0	0	1,236E+16
CAR(-2,2)	9202	0,000	0,067	-0,605	-0,022	0,022	0,413

Table 3: The summary statistics of the final dataset's numeric variables, displaying the number of instances, mean, standard deviation, minimum value, 25% percentile, 75% percentile, and maximum value

An interesting observation regarding the transfer fee-market value interaction term is that the value is 0, even at the 75% percentile. This tells that no or a relatively small transaction fee has been recorded for most transfers. This is no surprise since in each transfer window, only a few signings account for most of the money moving on the market. In other words, this phenomenon shows us the presence of "blockbuster signings". However, in the background, dozens of players move in and out every window for each team, who might never get the chance in the first team. Regarding the team_form variable, it is good to see that the entire range defined by the methodology (-3,3) is represented in the dataset, even if the mean value is low. This leads to the conclusion that teams struggle to get long positive streaks.

In Figure 9, the data map of this study can be observed. It visually represents how all individual datasets are used and the steps incorporated to get to the final dataset. It is essential to highlight that all structure pillars are created initially, and an overall merge is performed afterward to minimize the risk of mismatching specific data points.

Chapter 3.2.5 Training

The model training process aims to find the most optimal model structure and thus perform the most accurate predictions. This is crucial to getting a valid picture of how feature importances compare.

Chapter 3.2.5.1.1 Train-Test Split

A fixed and static train-test split is one of the most used preparative procedures in machine learning (Tan, Yang, Wu, Chen, & Zhao, 2021). It helps evaluate a model's performance and can help prevent overfitting. After training on a training set, the testing procedure is done on a separate dataset (test set), unseen for the model before (Brownlee, Train-Test Split for Evaluating Machine Learning Algorithms , 2020). Based on a recent study analyzing the relationship between the train-test ratio and dataset size, 80%-20%, or 70%-30% split, is ideal for more advanced machine learning methods (Rácz, Bajusz, & Héberger, 2021). In this study, a split of 80%-20% will be used.

Chapter 3.2.5.1.2 Cross-validation

In addition to the train-test split, implementing cross-validation (CV) is a powerful technique for preventing overfitting (Berrar, 2024). In the case of this study, CV is going to be used within the hyperparameter tuning procedure. Since the dataset is not overly complex, k-fold cross-validation will be applied, which is the simplest and the most widely used cross-validation alternative (Berrar, 2024). Figure 11 introduces how k-fold cross-validation is executed. The dataset of interest is equally and randomly split into k folds, and in each split, $k-1$ folds represent the training sets, and one fold (V_i) represents the validation set. This split is repeated until all folds serve as a validation set. The goal of the procedure is to calculate the cross-validated accuracy and compare the iterations based on it. Since $k=10$ is the most suggested setting to use (Berrar, 2024), it will be incorporated in this study as well.

Chapter 3.2.5.1.3 Hyperparameter Tuning

After the preparations to avoid overfitting in the model training process, the tuning grid has been defined for both models. Since the models have been trained on a GPU, the opportunity to cover a more extensive hyperparameter space and use more computational resource-friendly alternatives (e.g., Bayesian optimization) is optional. Regarding the FFNN assembly, the model structure should also be determined. When building an FFNN, choosing the correct number of hidden layers and their components is vital. Based on Svozil et al. from 1997, there is no reason to use more than two hidden layers. In fact, the best practice is to start with one hidden layer containing a high number of units and only increase to two if it cannot provide enough accuracy (Svozil, Kvasnicka, & Pospichal, 1997). This process is also used in this study, which results in a model with two hidden layers, as presented in Figure 12. Table 13 and Table 14 introduce the parameter grid for both models. It is important to mention for the

FFNN model that early stopping is introduced to reduce further the risk of overfitting with a value of 15 epochs, and the size of the tuning is three executions per trial with 30 trials and 150 maximum epochs each. Due to the three separate preventional strategies against overfitting, a more profound space can be left for the model to run and learn.

Chapter 4 Results

Chapter 4.1 Best Models and Evaluation

It is key to mention that no further external robustness checks have been incorporated due to the 10-fold CV and grid search-based hyperparameter tuning in both models. Bischl et al. (2021) wrote extensively about the challenges and best practices of hyperparameter optimization, and they stated that cross-validation is essential for estimating the generalization error of models, and grid search is a standard method to explore hyperparameter spaces to find the optimal configurations systematically (Bischl, et al., 2023). Thus, besides these methods being key elements to prevent overfitting, they also provide internal robustness checks on the go. The parameter grids for both models and the evaluation metrics have already been introduced. Table 6 now shows the final parameters for the two best-performing models.

Figure 13 and Figure 14 show how the evaluation metric (loss metric) decreased over the optimization process of the hyperparameters and iterations. Even though neither of the models can perform the same accuracy on the validation set as on the training set, this is normal. In the early stages of the training, the loss values decrease more or less equally in the case of both models and validation loss starts to flatten out, where the model cannot reproduce what it learned based on the test set's relationships. Overfitting should not be present based on the different preventive tools implemented and the analysis of the loss plots; however, some fluctuations can be observed in the FFNN model's loss plot. This means that the model is unsure about the validation set. Still, the flattening of the curve is aligned with the training loss's flattening, which tells that it is happening due to the data structure's complexity and not overfitting. Also, it is clear that the minimum RMSE value is not at the last epoch, which is due to the presence of early stopping.

Table 15 holds the two models' exact values on the test and validation sets, augmented with a reference OLS executed on the same dataset. Regarding R^2, XGBoost performs the best by far, with 28.3% on the validation set, followed by FFNN with 16.3% and the OLS model with 2.4%. The biggest difference between the performance on the training set and the validation/test set is in the case of the OLS model, which shows that the linear model struggles the most with the conversion of the discovered patterns to the prediction phase, which is also backed by the fact that the test RMSE is smaller than the training RMSE, which goes against the theory. This relative difference is 69.47% for the OLS, 46.1% for the XGBoost, and 42.32%

for the FFNN. Moreover, the exciting phenomenon between the R^2 and RMSE instances has to be elaborated on. The comparison of the differences between models displays that while FFNN performs significantly better in terms of R^2, this is not true when it comes to RMSE. Even though intuition says that the two metrics are supposed to move hand-in-hand, this phenomenon is not unrealistic at all, especially when comparing models with different complexities. The roots of this occurrence are the difference in the distribution and the nature of errors. Based on equations (15), (16), (17), and (18), the different roles of individual errors cause this discrepancy. Suppose the individual errors are more volatile and of a larger magnitude. In that case, the model may have the same RMSE – since the mean of the errors can be the same with larger volatility, but it won't capture the underlying variance in the data well. This leads to a lower R^2. Knowing this, incorporating the best practice of evaluating model performances based on two metrics that are responsible for different aspects seems to be an even better decision. These results answer *SQ3*; advanced machine learning methods are not only more favored because of the more nuanced interpretation techniques, but they also predict with significantly larger accuracy than linear models. This is most likely due to their ability to handle nonlinear relationships and the option to tune hyperparameters extensively.

Chapter 4.2 SHAP Values

To start interpreting SHAP results, one "bee swarm" plot, the most used type of SHAP plot, has been conducted for both models. These can be observed in Figure 2. The color coding of the plots represents the values of the feature; the warmer the color, the higher the given value is. The X-axis shows the effect a given value has on the model's prediction. In this case, the bee swarm plot is the ideal selection to see the overall importance composition and to compare variable effects on the independent variable, CAR(-2, 2). The features' order represents the order based on the overall importance of the feature. This order-based importance ranking can be utilized to create a comparison metric by calculating the mean of the position of each feature in the given feature group. For the sake of simplicity, call it the Mean Feature Importance Index (MFIP). As we group the features based on what aspect they represent (other than the season and team flags), the three primary groups are left: team form feature group, financial feature group, and transfer feature group. After the calculation of the MFIP for all three of these groups, the team form is clearly the most definitive aspect with an MFIP of 2.5, followed by the financial feature group with an MFIP of 6, and lastly, the transfer-related features with an MFIP of 22.5. This high-level metric is a perfect way of sensing the difference between feature group importances before diving deeper into the result analysis and seemingly provides eye-opening insights. Proving that team-form-related information has such a key role in stock performance brings back and redefines the approach of “hot-hand” Wood used in his study in 1992, who tried to research this phenomenon from the players’ side (Wood, 1992). Now, this study shows that while the hot-hand effect is still a key concept for sportspersons, it is a key concept for football investors as well, who are likely to predict the performance of teams that are in their portfolio based on their recent results.

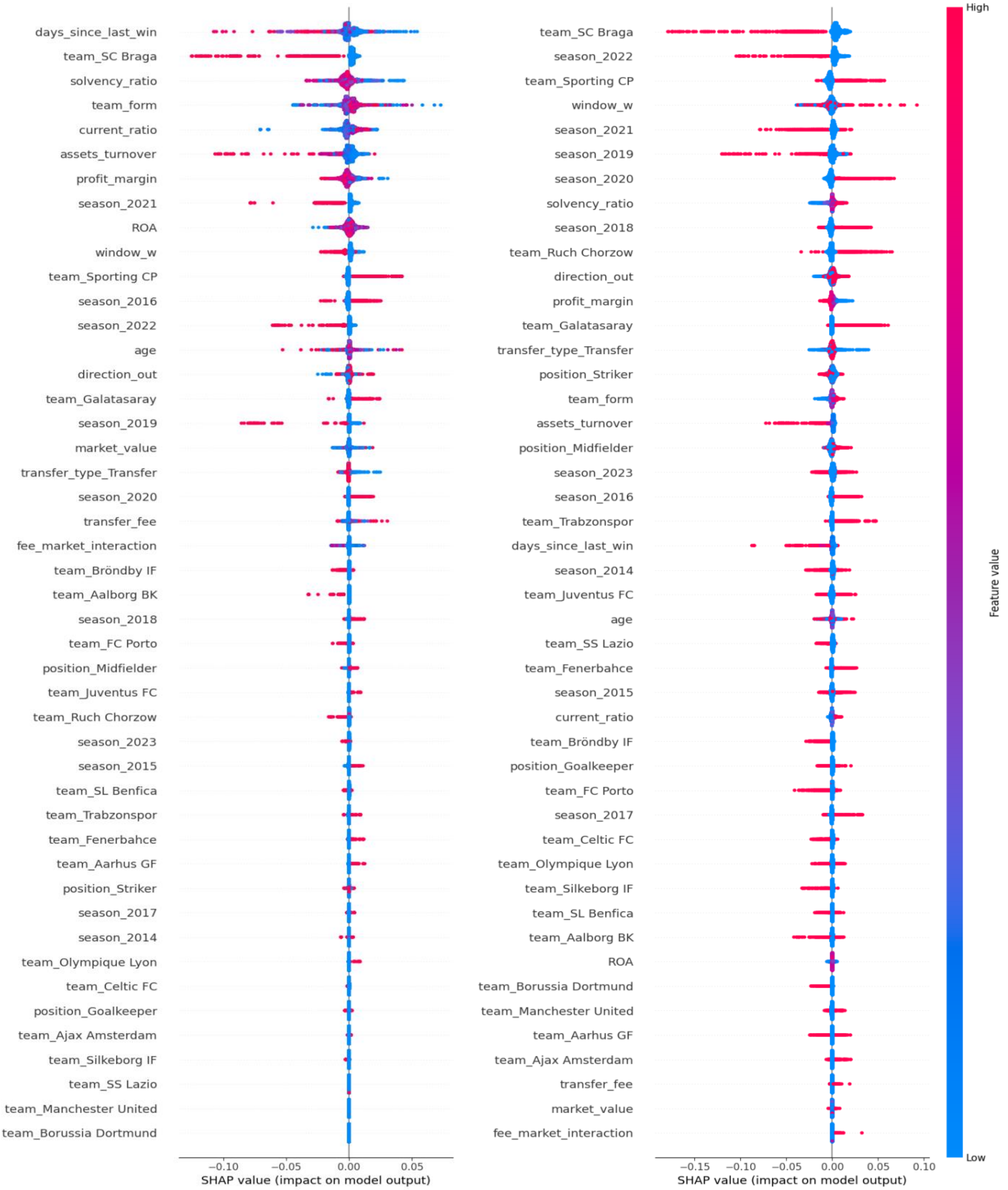

Figure 2: The SHAP bee swarm plots of the two models: XGBoost on the left, FFNN on the right

Chapter 4.3 Transfer-related Variables

Primarily, the transfer-related variables (age, market_value, transfer_fee, positions, direction_out, window_w, fee_market_interaction) should be analyzed to answer *SQ2*. Since XGBoost performed significantly better on the dataset, its results will be the ones in focus, compared to the FFNN model's results recurringly. Among these variables, the winter transfer window (window_w) flag is the most important variable in the case of both models (tenth for XGBoost and fourth for FFNN). However, while on the left plot, higher values harm the prediction – so when window_w=1, the CAR(-2, 2) predictions are smaller – on the right plot, the effect is the opposite. Based on Chapter 3.2.3.1.3, the behavior of the variable is more reasonable under XGBoost since the presence of flagship transfers is less likely there, which could rapidly increase closing prices. Interestingly, FFNN could not use the extra information provided by transfer-related variables. Likely, FFNN functions less efficiently with outliers since most of the transfer_fee values are 0 and the market_values are low, as shown in Table 3. On the right plot, transfer_fee, market_value, and their interaction term are the three least important variables, while for XGBoost, they are the eighteenth, twenty-first, and twenty-second, respectively, which means the middle of the feature importance space. To further inspect these variables, partial dependence plots (PDP) have been created to see the effect of certain variables on the model prediction in more detail. Partial dependence plots show the average effect of each value in a feature on the predicted outcome of a machine learning model. The y-axis represents the change in the prediction (marginal effect) as the feature value changes along the x-axis, holding all other features constant. The grey bars in the background represent the distribution of the feature values.

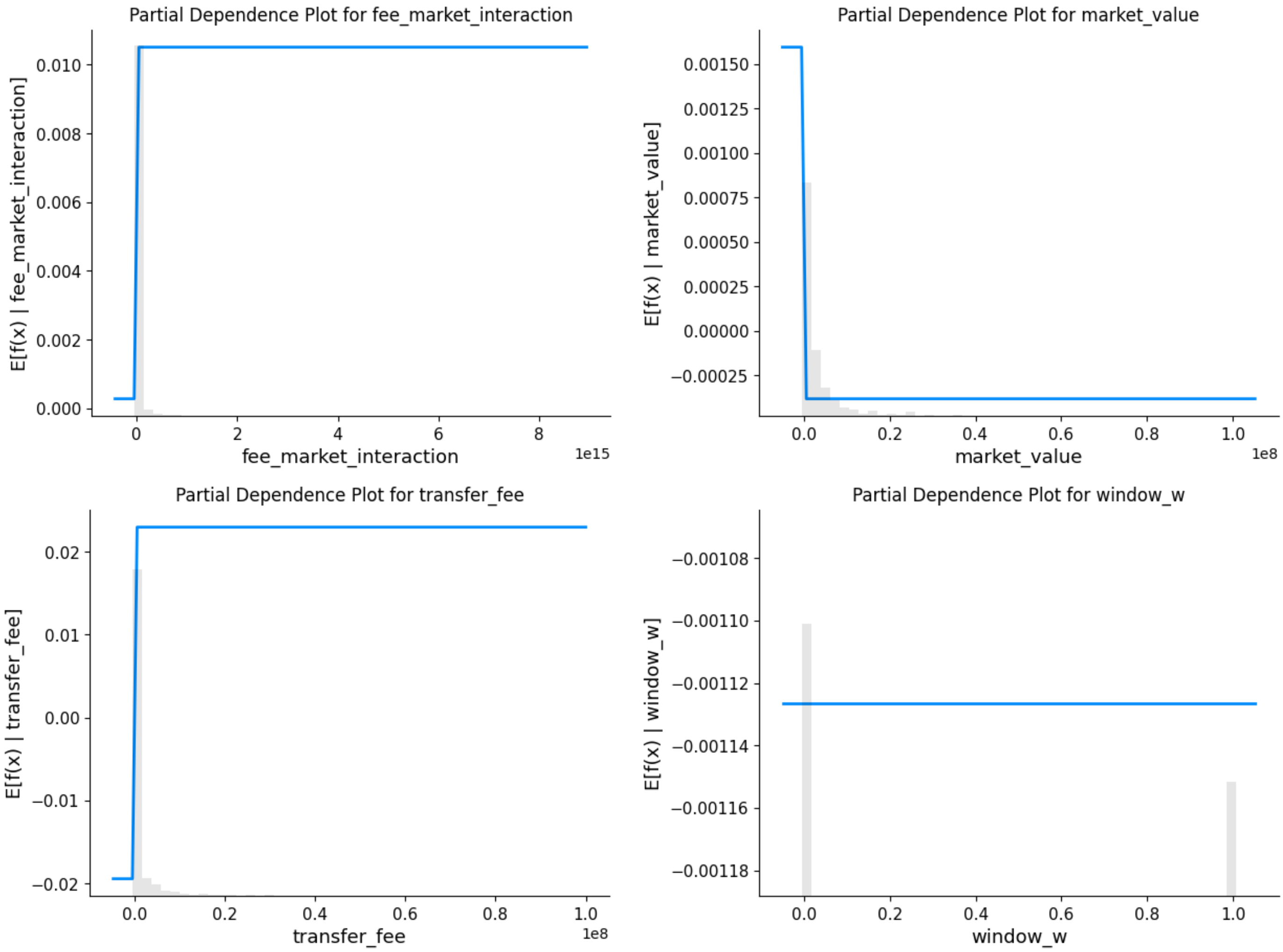

Figure 3: The partial dependence plots derived from the XGBoost model for the variables window_w, transfer_fee, market_value, and market_fee_interaction

The variables' effect based on the PDPs shows interesting and unintentional understanding from XGBoost's side. First, the third plot clearly shows how skewed the distribution of transfer fees is, which makes the scaling and the gradually increasing manner disappear, and every transfer fee value that is not 0 has the same effect on stock prices, slightly above 0.02. Market value is even more interesting, where skewness is also present, but oppositely, immense market value hurts stock prices. Either the model cannot capture the relationship sufficiently, or there is a phenomenon of investor skepticism, where they do not support spending a given amount on a given player. Due to the skewed distribution of values for the transfer fee, market value, and their interaction term, it seems like the model struggles with capturing the effect change where the number of observations with given values decreases. This also leads to the flatlining phenomenon on the PDP plots. The FFNN model in Figure 16 and Figure 17, however, struggled with capturing relationships that serve any intuition since the magnitude of the y-axis is often unlogical, and the predictions cannot leave the negative range.

Also, since there are barely any observations above 20,000,000, The PDP line is just influenced by a small number of outlier values. In the case of the winter transfer window flag, the PDP shows a constant effect on the prediction with a value slightly lower than 0.00112. Even though being the 10th most important variable overall and having a flat PDP line seems contradictory at first sight, it is a valid scenario. For example, for certain teams or during certain seasons, being in the transfer window might lead to significant stock price increases (positive SHAP values), but for other teams or seasons, it might have little impact or even a negative impact (negative SHAP values). In the bee swarm plot, this variability is apparent. However, when these effects are averaged in the PDP, they might cancel out, resulting in a flat line.

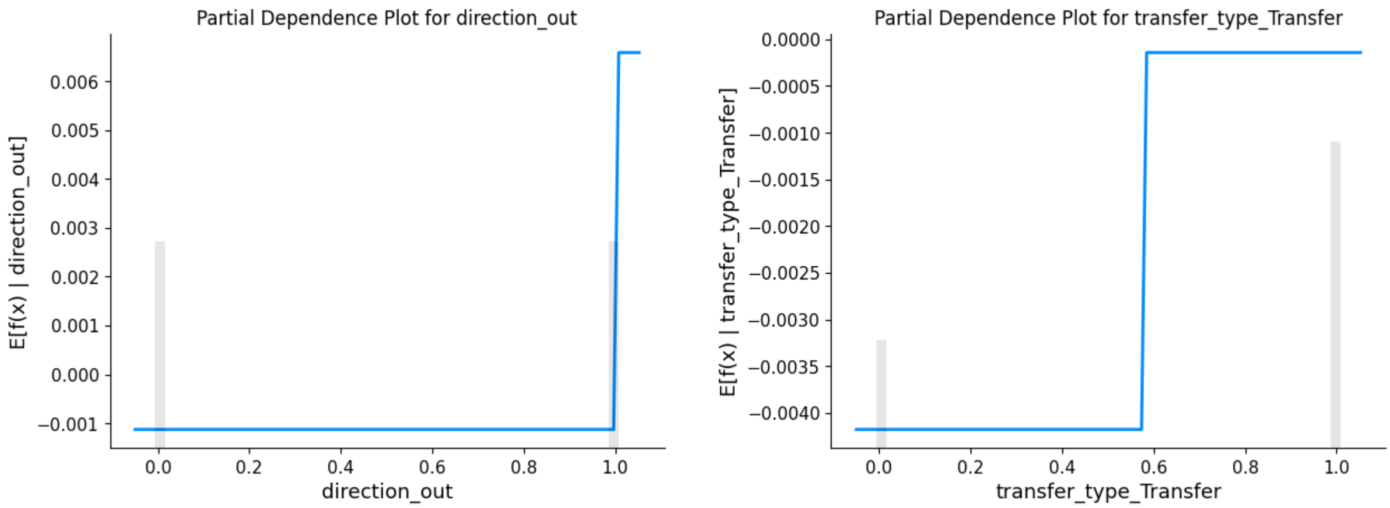

Figure 4: The partial dependence plots derived from the XGBoost model for the variables direction_out, and transfer_type_Transfer

The transfer direction and transfer_type flags differ since they are booleans, one-hot encoded variables. The transfer direction's PDP displays that outgoing transactions strongly and positively shape stock prices. Conversely, while loans have a negative effect, transfers are still in the negative range. This might point out conventional investor behavior, where money inflow is generally considered positive. Another key insight from the y-axis is that the direction of the transaction has a significantly larger impact on the prediction than the type of the transfer. However, the distribution of the transfer type flag is not the same as that of the direction flag, which challenges the model in the correct understanding.

Lastly, the remaining transfer-related variables are players' age and position flags. As the PDPs show, they all constantly affect the independent variable regardless of the value. This is understandable for the position flags, but it is more surprising in the case of age. The most likely underlying reasons are that these attributes are expected in other variables and are not generally as important. Even though age has a constant slightly negative value, it is the 14th most important variable. It is also worth mentioning that the impact on the prediction is the same for all four features, and it is for the transfer window flag. As it seemed logical for only one case, for six it does not. The most likely scenario is that the model is capturing interactions between features rather than the individual effects of each feature. As a result, the PDPs for

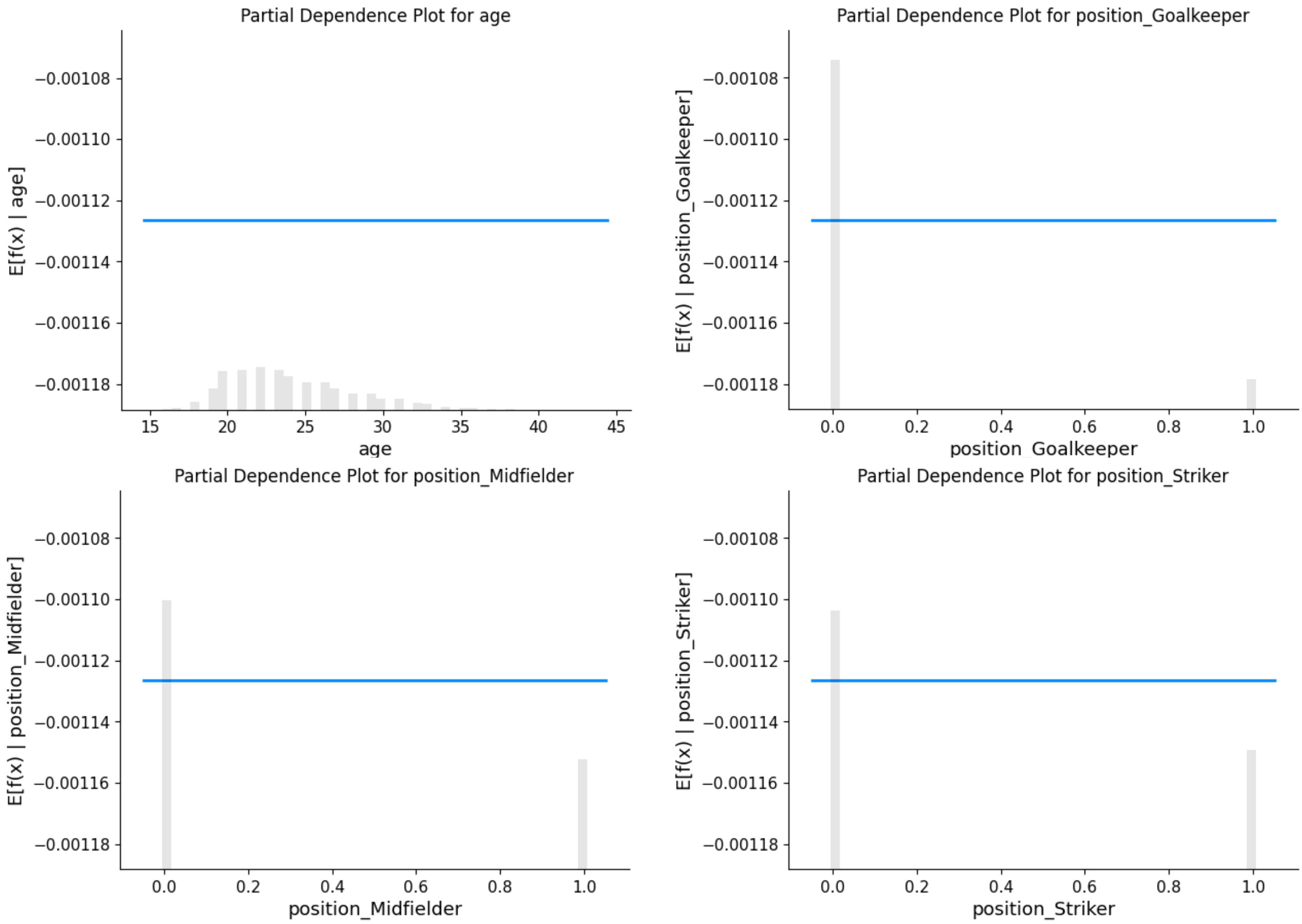

Figure 5: The partial dependence plots derived from the XGBoost model for the variables position flags and age

individual features might appear flat because the interaction effects dominate, and the individual effects are minimized. This requires observation and finetuning in the future. Contrarily, the PDP of the age variable for the FFNN model in Figure 15 is not constant; the inclining trend as age increases does not support the domain knowledge.

Generally, to answer *SQ2*, the three most critical transfer-related factors in shaping public football teams' stock price change are the transfer window flag, the player's age, and the transfer direction flag. Surprisingly, financial details of transfers only represent the second bunch in importance, closing with the player's position.

Chapter 4.4 Financial & Team Form-related Variables

Interestingly, XGBoost recognizes financial and team form variables as the most important categories. The variables days_since_last_win and team_form are the first and fourth, respectively, and solvency_ratio, current_ratio, assets_turnover, profit_margin, and ROA are the third, fifth, sixth, seventh, and ninth. For days_since_last_win, understandably, the higher the value, the lower the prediction will be due to the negative investor mood caused by the poor form. The noise is more considerable for team_form, but the overall picture shows the inverse, which is correct.

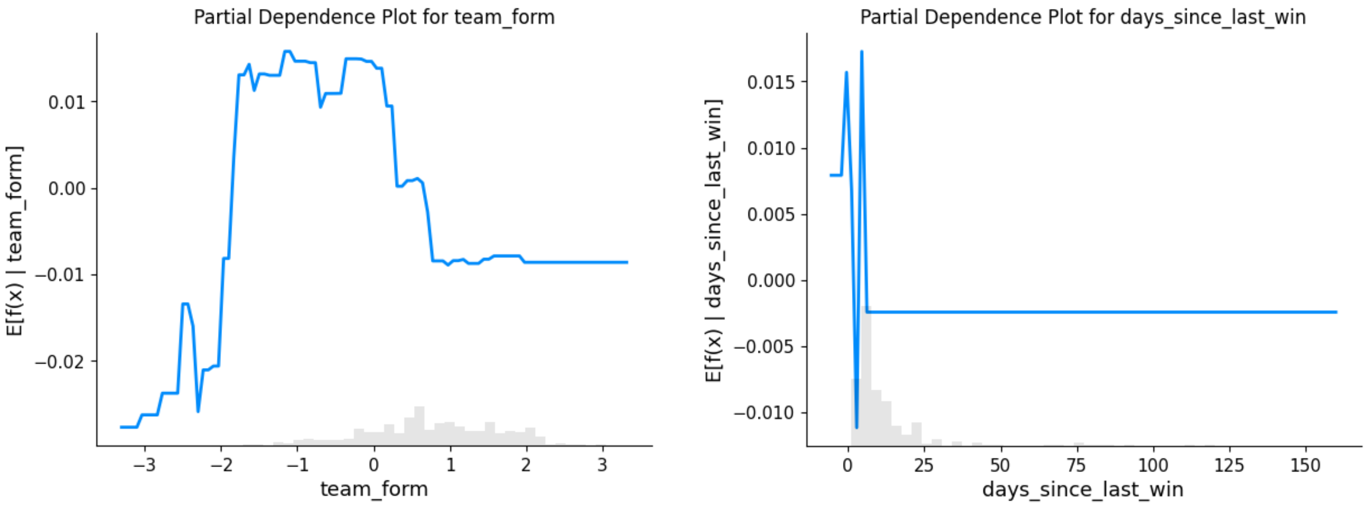

Figure 6: The partial dependence plots derived from the XGBoost model for the variables team_form and days_since_last_win

For team_form, the PDP shows intuitional results, but there are still some details to highlight. Around team_form=1, the effect line flattens below the 0.00 point, which means that excellent form does not impact the stock prices, but slightly lousy form strongly does. This might also be because of the small number of values in that range, which makes the model overestimate the effect. For days_since_last_win, the effect line is concentrated around the distribution of the values. It is hard to derive insights, but the first few days strongly negatively impact stock prices, which marks the instant decrease in investor mood after a loss. However, around 5-7 days, this effect turns and bounces to the positive side, and the effect line flattens out slightly below 0 after 8-10 days, which means the investor tolerance level is low. Interestingly, in the case of both features, FFNN seemed to capture the feature relationships better (Figure 23 and Figure 24). The impact of the past days since the last win increases in the negative direction as the number of days grows, and the team form feature shows a perfect incline. However, the FFNN model's predictions are still significantly more negative than they are for XGBoost, which might be one of the effectors of the accuracy difference.

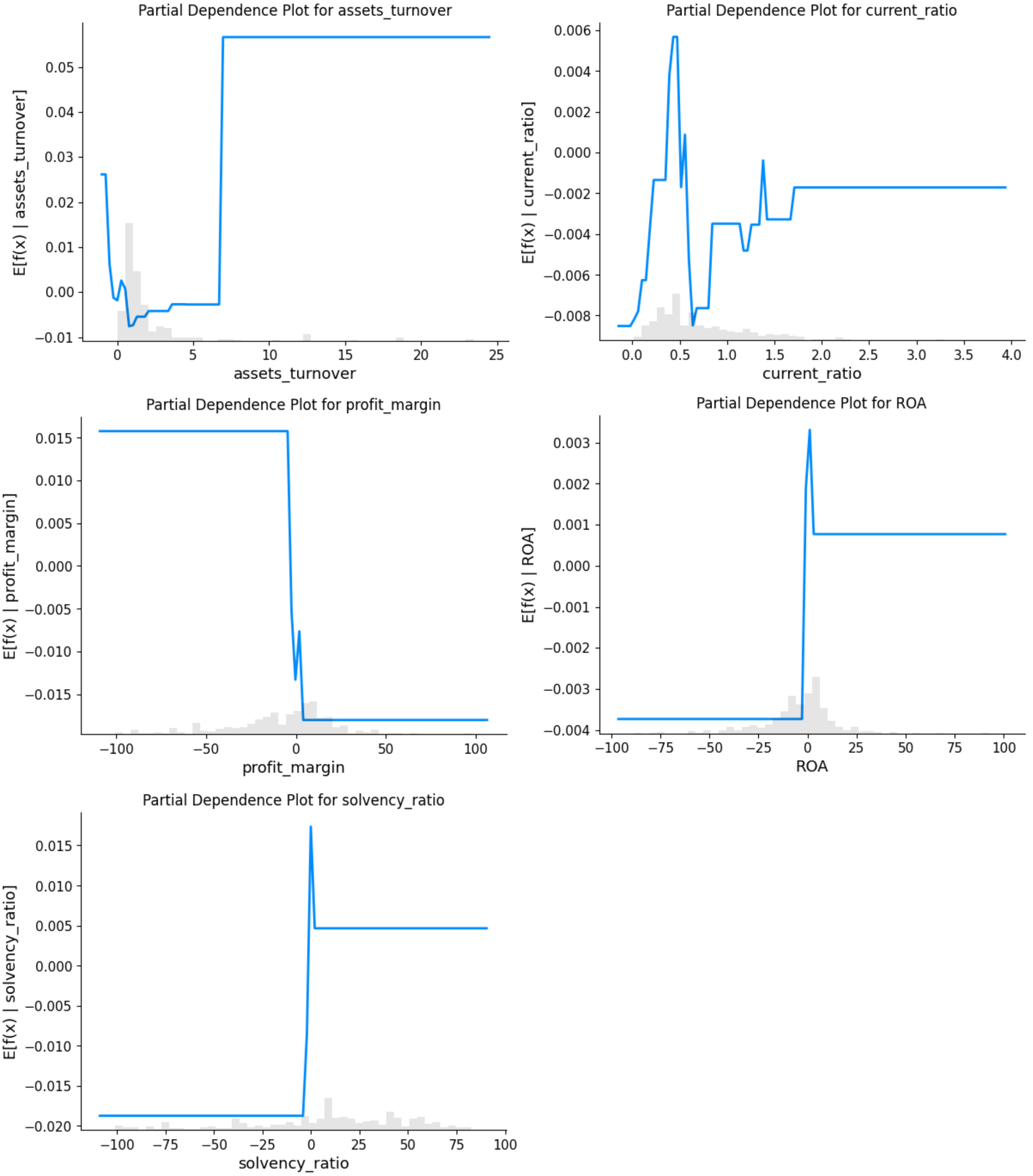

Figure 7: The partial dependence plots derived from the XGBoost model for the variables assets_turnover, current_ratio, profit_margin, ROA, and solvency_ratio

By analyzing the PDPs of the five financial variables, profit_margin shows the most exciting relationship between the predictions and the feature values. The prediction values increase from a very negative to a slightly negative margin, and there is a sharp decrease at the

0 value, turning negative. The model might be capturing complex interactions between profit_margin and other features. The sharp decline could be an artifact of how the model fits these interactions, primarily if profit_margin interacts strongly with other vital features. All four other plots show a trend aligned with intuition since the bigger these metrics are, the more significant the financial stability they mirror.

To answer *SQ1*, among the included variables in this study's analysis, the variable groups with the most prominent importance are team form-related and financial variables. These are the factors that shape closing price change the most. This means that even though individual transfers have significance in how closing prices change, investors are more likely to rely on economic components sourcing from financial reports and longer-term constants displaying how the entire team performs. This also means that football team investors like to be active and engaged in their investments and follow the components of their portfolio.

Chapter 5 Conclusions & Limitations

In the final chapter, the research process and findings are summarized, followed by limitations and suggestions for future studies based on this research.

Chapter 5.1 Summary and Implications

The subject of this study was to analyze the relationship between public football teams' player transfer details and their change in stock prices, compared to other factors (financial- and recent team form-related aspects.

Thus far, despite a limited number of studies on the topic, the study space has expanded rapidly due to the quality of these studies. The two most actual papers are from De Bakker (2016), who approached the problem from a strictly statistical point of view. In his paper, de Bakker focuses on using profound proxying and controlling techniques, thus increasing accuracy and reducing bias, while using an OLS model to deduct variable coefficients. De Bakker already used abnormal and cumulative abnormal returns to predict, which is the standard in financial event studies (de Bakker, 2016). In 2022, Kirchner followed up on de Bakker's work, spotlighting investor attention. Kirchner wanted to know the effect of football transfers on stock performance and investor attention separately. To conduct his research, he tried to use trading volume as a proxy for investor attention, but he did not find significant relationships. In his model, he used a fixed effects panel model.

This research was designed to extend the already existing research space, specifically based on an extensive analysis of past and recent studies connected to the topic. The first step is to leave the linear models and use state-of-the-art machine learning methodologies that can capture nonlinear relationships. XGBoost and feedforward neural networks were the final choices based on a selective process. The second comes from the first; since leaving linear models, the study design must also replace coefficient-based interpretation. To add to the interpretability, an independent, model-agnostic solution, SHAP values, is introduced, which is an agile and compact way to obtain various aspects of feature importance and effect on the dependent variable. Thus, now we know that the use of advanced models greatly increases accuracy (which leads to a more accurate prediction of SHAP values), and we have a detailed, further expandable overview of what the most important shaping factors are for public football teams' stock price change, and how they shape that change.

To elaborate on this, the study found that even though transfer details have some importance, it cannot be compared to the shaping factor of metrics that evaluate the teams' recent form (see Chapter 3.2.3.1.1 and Chapter 3.2.3.1.2), and the key financial ratios, deducted from the teams' balance sheets (see Chapter 3.1.3). We can create a comparative metric to compare the feature groups based on their mean importance (the mean ranking of the individual features). Based on this rationale, the team form group's mean feature importance index (MFIP) is 2.5, 6 for the financial feature group, and 22.5 for the transfer-related features. A more nuanced comparison can be obtained by expanding the number of groups used in the model. As we looked at the effect created by different variables more in detail, it became clear that skewness in the feature distribution dramatically impacts how a feature alters the prediction, which aligns with the theory since the fewer instances there are of a value, the fewer options the model has to capture the relationships. This phenomenon and the individual effect of features on the prediction can be observed in the partial dependence plots in Chapter 4.3.

Chapter 5.2 Limitations

The two main limitations of this study are data availability and the nature of financial event studies. Even though Transfermarkt is open to providing all the match-, transfer-, and player performance-related details, the number of these games and the connecting information recorded in an organized manner are limited (also due to the small number of public teams). The shallow record of player performance variables (goals, assists, appearances, yellow cards, red cards) is not nuanced enough to create in-depth player form evaluation metrics the way it has been done for the teams since players from multiple positions, especially defenders and goalkeepers could only be evaluated based on cards, which would be against rationale. This is also the case with the financial variables. Their reporting duties also differ because the teams are represented in different countries. This results in inconsistent reporting dates, reportable metrics, and ratios. This is why only three of the five leading financial metric categories could have been covered in the study (Welc, 2022).

Regarding the limitations caused by the nature of event studies, there are clear uncertainties that can strongly affect the direction of further research. The independent variable set, including flags, controls, and proxies, must be as nuanced as possible to get more profound results and answers on the topic. De Bakker started this study in 2016, and this study aimed to continue it by adding certain financial variables but excluding elements without unquestionable significance or effect (e.g., a proxy for investor mood). The academic space is still trying to determine what factors might play a key role in shaping the stock performance. We know this is a particular market, but we do not know precisely why. Thus, the key is to improve step by step. The most successful related studies always revolved around a specific phenomenon with a profound focus (Galloppo & Boido, 2020), (Kirchner, 2022), (Wilhelmsen, 2020).

Chapter 5.3 Suggestions For Further Studies

Based on the last sentences of the previous chapter, I suggest keeping the study designs as they turned out to be successful, narrow, and strictly aimed. Focus on 1-3 key ideas to be developed or changed, and maintain the scope throughout the entirety of the study. I do not see the need to perform complex, summarizing studies before the collective understanding of the topic reaches a level where information gain suddenly reduces. Based on my literature review and research space analysis, this topic is still immature. There are still key effectors waiting to be discovered, as well as the interaction between different effector groups, or within the groups. Hopefully, this procedure will be more straightforward based on the findings and development of this study.

The increased modeling accuracy due to the application of complex machine learning models and the more intuitive interpretation strategy will make evaluating new effectors significantly easier. Even though the primary focus of this study was on transfer-related details, I certainly think that the insights on the importance of recent team form are unavoidable. The following steps should be correctly determining the window size or even creating a more complex formula that captures a team's recent performance even better. Another field for development is the pick of financial variables. If one is able to find a deeper source database, maybe more aspects of the five financial pillars can be covered, which means a more detailed picture of a team’s economic stability. All considered, not only are there development opportunities based on the scope of this study, but there are still undiscovered areas waiting to be unveiled by creativity.

Chapter 6 Bibliography

Aglietta, M., Andreff, W., & Drut, B. (2010). Floating European football clubs in the stock market. *EconomiX*, 4.

Agrawal, S. C. (2020). Deep learning based non-linear regression for Stock Prediction. *IOP Conf. Series: Materials Science and Engineering* .

Amjad, M., Ahmad, I., Ahmad, M., Wróblewski, P., Kaminski, P., & Amjad, U. (2022). Prediction of Pile Bearing Capacity Using XGBoost Algorithm: Modeling and Performance Evaluation. *Applied Sciences 12 no. 4*.

Bailey, M. (2023, June 14). *New York Times*. From The Athletic: https://www.nytimes.com/athletic/4605403/2023/06/14/transfer-window-explained-facts/

Barnett, V., & Hilditch, S. (1993). The Effect of an Artificial Pitch Surface on Home Team Performance in Football (Soccer) . *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, 39-50.

Bell, A. R., & Brooks, C. (2012). Over the moon or sick as a parrot? The effects of football results on a club's share price. *Applied Economics*, 3435-3452.

Berrar, D. (2024, March 24). *Cross-validation.* From Berrar: berrar.com

Bhanja, S., & Das, A. (2019, January 7). *Impact of Data Normalization on Deep Neural Network for Time Series Forecasting* . From arXiv: https://arxiv.org/abs/1812.05519

Bhatia, A. (2020, April 3). *Do Football Transfers Actually Improve Performance?* . From Harvard Library: https://nrs.harvard.edu/URN-3:HUL.INSTREPOS:37364729

Bischl, B., Binder, M., Lang, M., Pielok, T., Richter, J., Coors, S., . . . Lindauer, M. (2023). Hyperparameter Optimization: Foundations, Algorithms, Best Practices and Open Challenges. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery 13(4)*.

Boido, C., & Fasano, A. (2007). Football and Mood in Italian Stock Exchange. *The ICFAI Journal of Behavioural Finance*, 32-50.

Boțoc, C., Mihancea, E., & Molcuț, A. (2019). Football and Stock Market Performance Correlation: Evidence from Italy. *Scientific Annals of Economics and Business*, 525-539.

Brownlee, J. (2019, May 23). *How Much Training Data is Required for Machine Learning?* From Machine Learning Mastery: https://machinelearningmastery.com/much-training-data-required-machine-learning/

Brownlee, J. (2020, August 26). *Train-Test Split for Evaluating Machine Learning Algorithms* . From Machine Learning Mastery: https://machinelearningmastery.com/train-test-split-for-evaluating-machine-learning-algorithms/

Bruin, M. (2023, April 30). Culture, the determinant of a profitable football transfer?

Chalabi, M. (2013, November 14). *DataBlog*. From The Guardian: https://www.theguardian.com/football/datablog/interactive/2013/nov/14/113-years-of-international-football-player-transfers

Chen, T., & Guestrin, C. (2016, August 13). XGBoost: A Scalable Tree Boosting System. *KDD '16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 785-794.

Corporate Finance Institute Team. (2024, July 20). *Abnormal return*. From Corporate Finance Institute: https://corporatefinanceinstitute.com/resources/equities/abnormal-return/

de Bakker, D. (2016, November 14). The effect of player transfers on stock prices.

De Mulder, W., Moens, M.-F., & Bethard, S. (2014, September 17). A Survey on the Application of Recurrent Neural Networks to Statistical Language Modeling. *Computer Speech & Language 30(1)*, pp. 61-94.

Doane, D. P., & Seward, L. E. (2011). Measuring Skewness: A Forgotten Statistic? *Journal of Statistics Education*.

Dubey, S. R., Singh, S. K., & Chaudhuri, B. B. (2022). Activation functions in deep learning: A comprehensive survey and benchmark. *Neurocomputing Volume 503*, 92-108.

Ebert-Uphoff, I., Lagerquist, R., Hilburn, K., Lee, Y., Haynes, K., Stock, J., . . . Stewart, J. Q. (2021, June 21). *CIRA Guide to Custom Loss Functions for Neural Networks in Environmental Sciences -- Version 1*. Retrieved from ResearchGate:

https://www.researchgate.net/publication/352558898_CIRA_Guide_to_Custom_Loss_Functions_for_Neural_Networks_in_Environmental_Sciences_--_Version_1

Frederick, S. C., Abbott, A., & Thompson, M. (1993). Returns when signal of value are frequent : the Boston Celtics. *Journal of business and economic studies*, 69-81.

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Ann. Statist. 29(5)*, 1189-1232.

Fűrész, D. I. (2021). Sztárigazolások eredményességre gyakorolt hatásának modellezése az európai labdarúgás példáján keresztül.

Fűrész, D. I., & Rappai, G. (2022). Information Leakage in the Football Transfer Market. *European Sport Management Quarterly*, 419-439.

Galloppo, G., & Boido, C. (2020). How Much Is A Goal in the Football Championship Worth? Match Results and Stock Price Reaction. *International Journal of Sport Finance*.

Günter, N., & Vischer, L. (2024, 05). Development of loan players in professional football. *Diskussionspapier des Instituts für Organisationsökonomik*.

Hashi, G., & Kroken, C. (2017, September 01). *Market Efficiency in the European Football Transfer Market.* From BI Open: https://biopen.bi.no/bi-xmlui/bitstream/handle/11250/2485695/1760412.pdf?sequence=1

Hodson, T. O. (2022, July 19). Root-mean-square error (RMSE) or mean absolute error (MAE): when to use them or not. *Volume 15, issue 14*, pp. 5481-5487.

Kenton, W. (2023, September 18). *Shapley Value Definition and Example of How It Is Applied*. From Investopedia: https://www.investopedia.com/terms/s/shapley-value.asp

Kim, N., Molnár, P., Lučivjanská, K., & Villa, R. (2019). Google searches and stock market activity: Evidence from Norway. *Finance Research Letters*, 208-220.

Kirchner, N. (2022, June 10). The Effect of Football Transfers on Investor Attention and Stock Performances. *Rotterdam School of Management - Erasmus University*.

Lundberg, S., & Lee, S.-I. (2017). A Unified Approach to Interpreting Model Predictions. *NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems*, 4768 - 4777.

Maci, G., Pacelli, V., & D'Apolito, E. (2020). The Determinants of Stock Prices of European Football Clubs: An Empirical Analysis. *International Journal of Economics, Finance and Management Sciences*, 171-172.

Massey, K. (1997). Statistical Models Applied to the Rating of Sports Teams.

McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, 115-133.

McInerney, A., & Burke, K. (2023, November 14). *Feedforward neural networks as statistical models: Improving interpretability through uncertainty quantification.* From Arxiv: arXiv:2311.08139

Meng, Y., Yang, N., Qian, Z., & Zhang, G. (2020). What Makes an Online Review More Helpful: An Interpretation Framework Using XGBoost and SHAP Values. *Journal of Theoretical and Applied Electronic Commerce Research*, 466-490.

Miles, J. (2005). R-squared, Adjusted R-squared. *Encyclopedia of Statistics in Behavioral Science*, 1655-1657.

Mizruchi, M. S. (1991). Urgency, Motivation, and Group Performance: The Effect of Prior Success on Current Success Among Professional Basketball Teams. *Social Psychology Quarterly*, 181-189.

Moody's. (2024, July 20). *Orbis*. From Moody's: https://www.moodys.com/web/en/us/capabilities/company-reference-data/orbis.html

Navest, M. (2023, June 16). Stocks of Football Clubs: Winning .

Nielsen, M. (2013). *Neural Networks and Deep Learning.*

Orimoloye, L. O., Sung, M.-C., Ma, T., & Johnson, J. E. (2020). Comparing the effectiveness of deep feedforward neural networks and shallow architectures for predicting stock price indices. *Expert Systems with Applications*.

Pengfei, Y., & Xuesong, Y. (2020). Stock price prediction based on deep neural networks. *Neural Computing and Applications*, 1610-1628.

Pennsylvania State University. (2018). *2.5 - The Coefficient of Determination, r-squared* . From PennState Eberly College of Science: https://online.stat.psu.edu/stat462/node/95/

Premier League. (2017, May 30). From Premier League News: https://www.premierleague.com/news/60258

Rácz, A., Bajusz, D., & Héberger, K. (2021). Effect of Dataset Size and Train/Test Split Ratios in QSAR/QSPR Multiclass Classification. *Molecules 26, no. 4*.

Ramsey, J. (1969). Tests for Specification Errors in Classical Linear Least-Squares Regression Analysis. *Journal of the Royal Statistical Society*, 350-371.

Romano, F. (2024, July 5). *fabriziorom*. From Instagram: https://www.instagram.com/fabriziorom/?hl=en

Schneider, P., & Xhafa, F. (2022). *Anomaly Detection and Complex Event Processing over IoT Data Streams.* Elsevier Inc.

Seger, C. (2022, June 26). *An investigation of categorical variable encoding techniques in machine learning: binary versus one-hot and feature hashing.* From Digitala Vetenskapliga Arkivet: https://www.diva-portal.org/smash/record.jsf?pid=diva2%3A1259073&dswid=9822

Smith, R. (2021, August 12). The Wisdom of the Crowd. *The New York Times*, p. 7. From https://www.nytimes.com/2021/08/12/sports/soccer/soccer-football-transfermarkt.html

Stadtmann, G. (2006). Frequent News and Pure Signals: The Case of a Publicly Traded Football Club. *Scottish Journal of Political Economy*, 485-504.

Stelmach, J. (2023, Feb 14). *TISA Group*. From https://www.tisagroup.ch/sports-business-usa-vs-europe/

Svozil, D., Kvasnicka, V., & Pospichal, J. (1997). Introduction to multi-layer feed-forward neural networks. *Chemometrics and Intelligent Laboratory Systems* , 43-62.

Tan, J., Yang, J., Wu, S., Chen, G., & Zhao, J. (2021, June 8). *A critical look at the current train/test split in machine learning.* From arXiv: https://arxiv.org/abs/2106.04525

The Economic Times. (2024, July 22). *What is 'Abnormal Rate of return'*. From The Economic Times: https://economictimes.indiatimes.com/definition/abnormal-rate-of-return

Transfermarkt. (2020, June 22). *News*. From Transfermarkt: https://www.transfermarkt.com/20-years-transfermarkt-from-non-league-football-to-the-champions-league/view/news/363452

Transfermarkt. (2024, July 13). *Transfermarkt.* From Summer Transfer Window: https://www.transfermarkt.com/transfer-statistik/transferfenster/statistik/stat/plus/0?art=s&kontinent_id=6&option=0

Tsoulos, I., Gavrilis, D., & Glavas, E. (2008). Neural network construction and training using grammatical evolution. *Neurocomputing Volume 72, Issues 1-3*, 269-277.

Tuychiev, B. (2023, December). *Understanding Skewness And Kurtosis And How to Plot Them*. From Datacamp: https://www.datacamp.com/tutorial/understanding-skewness-and-kurtosis

van Bottenburg, M. (2011). Why are the European and American sports worlds so different? Pathdependence in the European and American sports history.

Welc, J. (2022). *Evaluating Corporate Financial Performance.* Switzerland: Springer Nature Switzerland.

Wilhelmsen, H. (2020, April 21). Google Search Volume as an Attention Proxy in the Stock Market.

Wood, G. (1992). Predicting Outcomes: Sports and Stocks. *Journal of Gambling Studies*, 201-222.

Yisheng, L. (2021). *When Moneyball Meets the Beautiful Game: A Predictive Analytics Approach to Exploring Key Drivers for Soccer Player Valuation.* From Brock University Digital Repository: https://dr.library.brocku.ca/bitstream/handle/10464/15088/Brock_Li_Yisheng_2021.pdf?sequence=1&isAllowed=y

Chapter 7 Appendices

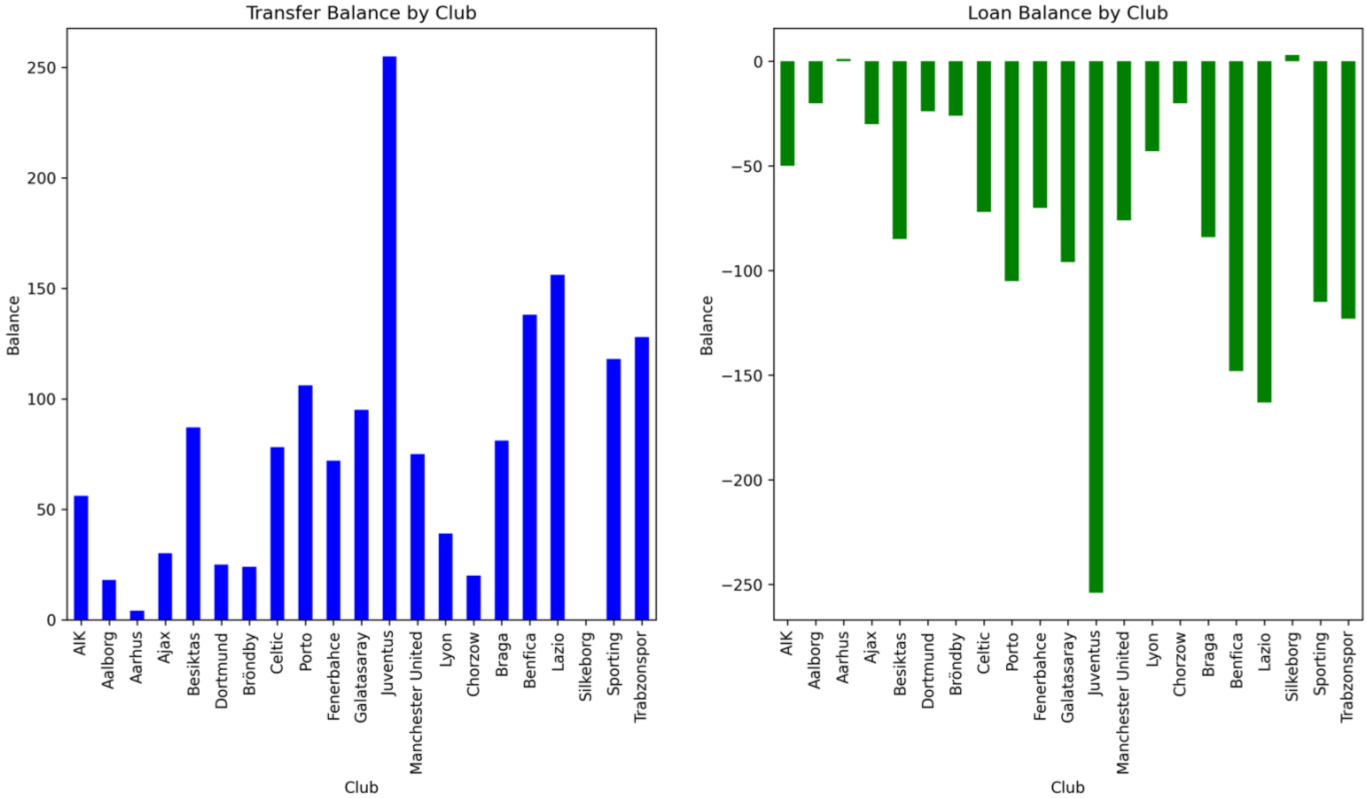

Figure 8: The transfer & loan balances of all clubs

Team	Reference Stock Index	Identification Source	Data Source
AIK	OMXS30	Infront Analytics	Nasdaq Nordic
Aarhus	OMXCGI	Yahoo Finance	Yahoo Finance
Aarhus	OMXCGI	Yahoo Finance	Yahoo Finance
Ajax	AEX-Index	Yahoo Finance	Yahoo Finance
Besiktas*	BIST 100	Yahoo Finance	Yahoo Finance
Dortmund	DAX PERFORMANCE-INDEX	Yahoo Finance	Yahoo Finance
Bröndby	OMXCGI	Yahoo Finance	Yahoo Finance
Celtic	FTSE100	Yahoo Finance	Wall Street Journal
Porto	PSI GR	Orbis	Yahoo Finance
Fenerbahce	BIST 100	Yahoo Finance	Yahoo Finance
Galatasaray	BIST 100	Yahoo Finance	Yahoo Finance
Juventus	FTSE MIB	Orbis	Yahoo Finance
Manchester United	NYSE Composite	Orbis	Yahoo Finance
Lyon	CAC 40	Yahoo Finance	Yahoo Finance
Chorzow	WIG 20	Infront Analytics	Stooq
Braga	PSI20.LS	Yahoo Finance	Yahoo Finance
Benfica	PSI GR	Orbis	Yahoo Finance
Lazio	FTSE MIB	Yahoo Finance	Yahoo Finance
Silkeborg	OMXCGI	Yahoo Finance	Yahoo Finance
Sporting	PSI GR	Orbis	Yahoo Finance
Trabzonspor	BIST 100	Yahoo Finance	Yahoo Finance

Table 4: The respective reference indexes used for each team's abnormal return calculation, including the source of identification and the data source

Output	Value
F-statistic	35.909
p-value	$3.025e^{-16}$
Degrees of Freedom (numerator)	2
Degrees of Freedom (denominator)	$7.31e^{3}$

Table 5: The output of the RESET test on the test OLS regression to validate the presence of nonlinear relationships in the data

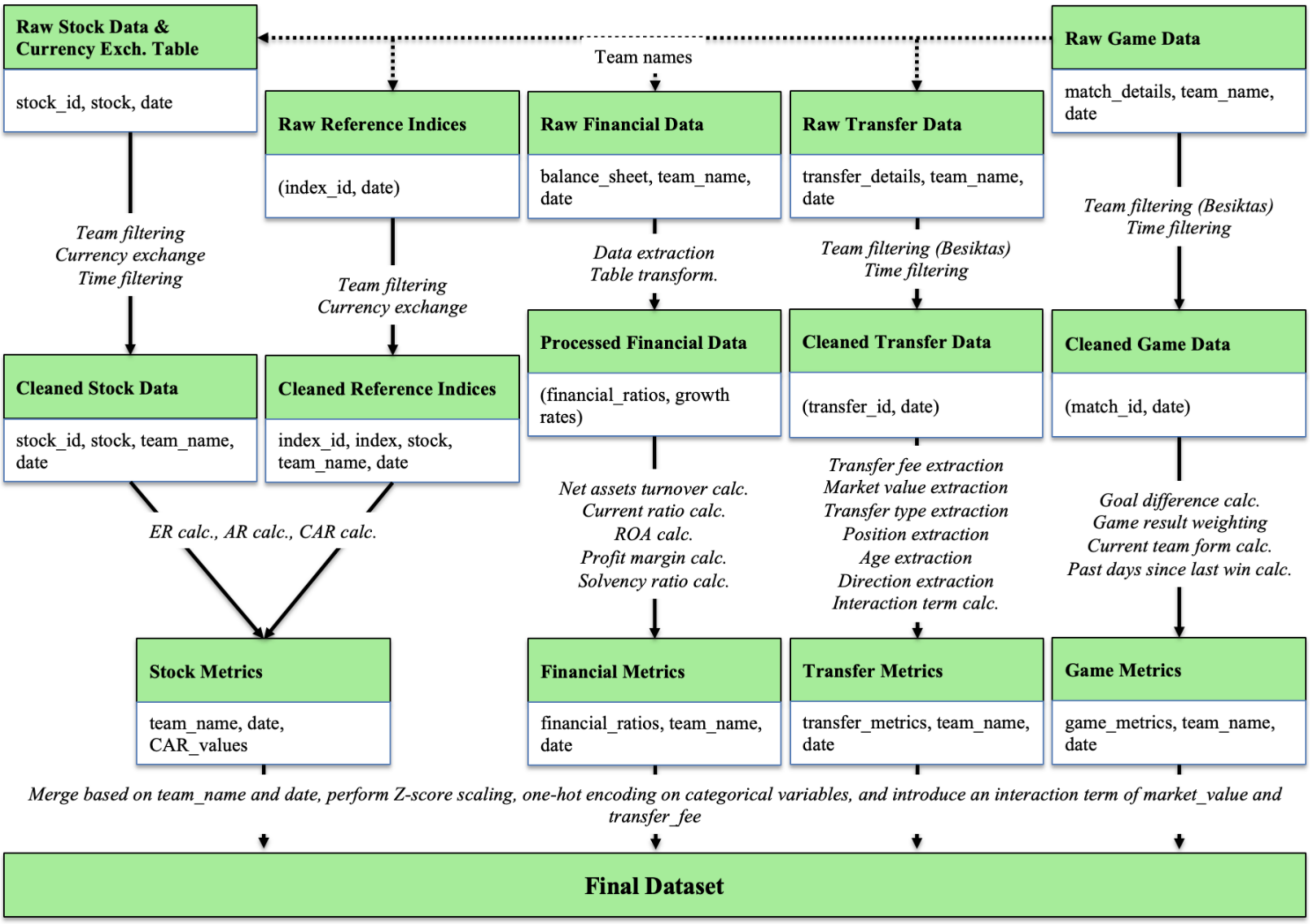

Figure 9: Data map, describing the process of assembling the final dataset, which serves as the input for XGBoost and FFNN.

Model	Parameter	Value
	colsample_bytree	0.8
	gamma	0
	learning_rate	0.1
	max_depth	5
	min_child_weight	5
	n_estimators	100
	reg_alpha	0
	reg_lambda	1
	subsample	0.8
	tree_method	hist
FFNN	units1	800
	units2	544
	l2_1	0.0001
	l2_2	0.0004
	learning_rate	0.001

Table 6: The parameters of the best-performing models

	Number of incoming transfers	Number of outgoing transfers	Number of incoming loans	Number of outgoing loans	Total transactions per team
AIK	169	113	16	66	**364**
Aalborg	124	106	16	36	**282**
Aarhus	126	122	22	21	**291**
Ajax	141	111	13	43	**308**
Besiktas	231	144	36	121	**532**
Dortmund	115	90	8	32	**245**
Bröndby	147	123	15	41	**326**
Celtic	213	135	30	102	**480**
Porto	246	140	23	128	**537**
Fenerbahce	217	145	20	90	**472**
Galatasaray	258	163	36	132	**589**
Juventus	431	176	18	272	**897**
Manchester United	161	86	8	84	**339**
Lyon	153	114	13	56	**336**
Chorzow	208	188	30	50	**476**
Braga	238	157	32	116	**543**
Benfica	294	156	15	163	**628**
Lazio	297	141	19	182	**639**
Silkeborg	98	98	11	8	**215**
Sporting	278	160	23	138	**599**
Trabzonspor	336	208	31	154	**729**
Total transactions per type	**4481**	**2876**	**435**	**2035**	**9827**

Table 7: The number of incoming & outgoing transfers and loans for the teams of interest between 04-09-2013 and 18-09-2023

Team	Number of Games Played	Distribution of Total Games (%)	Draw	Loss	Win	Team Win %
AIK	402	3.693%	94	98	210	52.239%
Aalborg BK	433	3.978%	101	151	181	41.801%
Aarhus GF	425	3.905%	109	144	172	40.471%
Ajax Amsterdam	534	4.906%	103	94	337	63.109%
Besiktas JK	524	4.814%	112	127	285	54.389%
Borussia Dortmund	539	4.952%	92	133	314	58.256%
Bröndby IF	453	4.162%	97	134	222	49.007%
Celtic FC	623	5.724%	93	99	431	69.181%
FC Porto	585	5.375%	91	95	399	68.205%
Fenerbahce	547	5.026%	119	107	321	58.684%
Galatasaray	542	4.980%	111	125	306	56.458%
Juventus FC	581	5.338%	102	98	381	65.577%
Manchester United	616	5.660%	126	151	339	55.032%
Olympique Lyon	538	4.943%	112	148	278	51.673%
Ruch Chorzow	393	3.611%	104	139	150	38.168%
SC Braga	555	5.099%	97	145	313	56.396%
SL Benfica	591	5.430%	93	94	404	68.359%
SS Lazio	540	4.961%	111	159	270	50.000%
Silkeborg IF	411	3.776%	86	155	170	41.363%
Sporting CP	550	5.053%	88	101	361	65.636%
Trabzonspor	502	4.612%	122	133	247	49.203%

Table 8: The respective match numbers and the main results statistics for all the matches in the Transfermarkt dataset

Variable Name	Explanation
Explanatory variables	
age	The players's age
market_value	The player's market value at the time of the transfer
transfer_fee	The transfer fee paid
team_form	The team of interest's form at the time of the transfer
days_since_last_win	The number of days left since the last win of the team of interest at the time of the transfer
assets_turnover	The assets turnover ratio of the team of interest based on the most recent financial report to the time of the transfer. It measures how efficiently the club uses its assets to create revenue
current_ratio	The current ratio of the team of interest based on the most recent financial report to the time of the transfer.. It displays a club's ability to pay its short-term obligations with its current assets
ROA	The return on assets ratio of the team of interest based on the most recent financial report to the time of the transfer. It shows how profitable a company is relative to its total assets
profit_margin	The profit margin ratio of the team of interest based on the most recent financial report to the time of the transfer. It shows the percentage of revenue that turns into profit
solvency_ratio	The solvency ratio of the team of interest based on the most recent financial report to the time of the transfer. It measures a company's ability to meet its long-term obligations
team_Aalborg BK	Dummy flag for team Aalborg BK
team_Aarhus GF	Dummy flag for team Aarhus GF
team_Ajax Amsterdam	Dummy flag for team Ajax Amsterdam
team_Borussia Dortmund	Dummy flag for team Borussia Dortmund
team_Bröndby IF	Dummy flag for team Bröndby IF
team_Celtic FC	Dummy flag for team Celtic FC
team_FC Porto	Dummy flag for team FC Porto
team_Fenerbahce	Dummy flag for team Fenerbahce
team_Galatasaray	Dummy flag for team Galatasaray
team_Juventus FC	Dummy flag for team Juventus FC
team_Manchester United	Dummy flag for team Manchester United
team_Olympique Lyon	Dummy flag for team Olympique Lyon
team_Ruch Chorzow	Dummy flag for team Ruch Chorzow
team_SC Braga	Dummy flag for team SC Braga
team_SL Benfica	Dummy flag for team SL Benfica
team_SS Lazio	Dummy flag for team SS Lazio
team_Silkeborg IF	Dummy flag for team Silkeborg IF
team_Sporting CP	Dummy flag for team Sporting CP
team_Trabzonspor	Dummy flag for team Trabzonspor
season_2014	Dummy flag for season 2014

season_2015	Dummy flag for season 2015
season_2016	Dummy flag for season 2016
season_2017	Dummy flag for season 2017
season_2018	Dummy flag for season 2018
season_2019	Dummy flag for season 2019
season_2020	Dummy flag for season 2020
season_2021	Dummy flag for season 2021
season_2022	Dummy flag for season 2022
season_2023	Dummy flag for season 2023
transfer_type_Transfer	Dummy flag for the transfer type (loan/transfer)
direction_out	Dummy flag for the transfer direction (in/out)
position_Goalkeeper	Dummy flag for the player position 'Goalkeeper'
position_Midfielder	Dummy flag for the player position 'Midfielder'
position_Striker	Dummy flag for the player position 'Striker'
window_w	Dummy flag for the transfer window (w/s)
fee_market_interaction	Interaction term for the market value and the transfer fee
Dependent variable	
CAR(-2,2)	The daily closing stock prices' cumulative abnormal returns with the window size of (-2, 2)

Table 9: The variable set of the dataset and their brief descriptions

Model	**RMSE**	**Adjusted R^2**
Multiple Linear Regression	0.841	0.661
Decision Tree	0.969	0.573
XGBoost	0.721	0.763
Random Forest	0.843	0.676
SVR	0.731	0.757

Table 10: Model performances for predicting soccer player valuation (Source: Yisheng, 2021)

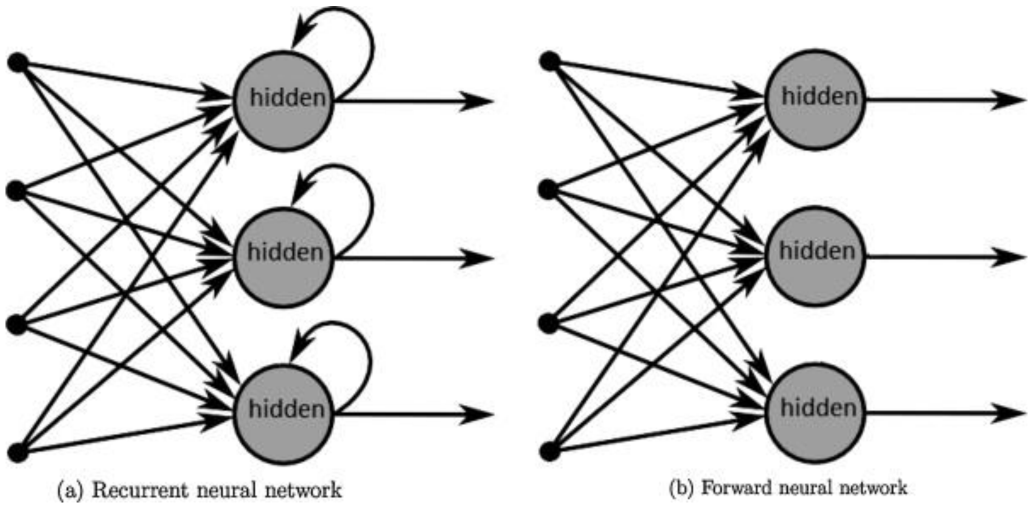

Figure 10: The layers structure difference between FFNN and RNN. Source: (De Mulder, Moens, & Bethard, 2014)

Abbreviation	Name	Weight
F	Final	1
FF	Final	1
HF	Semifinals	2
HFH	Semifinal Home	2
HFR	Semifinal Return (Away)	2
VF	Quarterfinals	3
VFH	Quarterfinal Home	3
VFR	Quarterfinal Return (Away)	3
AF	Round of 16	4
AFH	Round of 16 Home	4
AFR	Round of 16 Return (Away)	4
QRH	Qualifying Round Home	5
QRR	Qualifying Round Return (Away)	5
QR	Qualifying Round	5
1R	First Round	5
1	First Round	5
1RH	First Round Home	5
1RR	First Round Return (Away)	5
2R	Second Round	5
2	Second Round	5
2RH	Second Round Home	5
2RR	Second Round Return (Away)	5
3R	Third Round	5
3	Third Round	5
3RH	Third Round Home	5
3RR	Third Round Return (Away)	5
4R	Fourth Round	5
4	Fourth Round	5
4RH	Fourth Round Home	5
4RR	Fourth Round Return (Away)	5
5R	Fifth Round	5
5	Fifth Round	5
5RH	Fifth Round Home	5
5RR	Fifth Round Return (Away)	5
6R	Sixth Round	5
6	Sixth Round	5
6RE	Sixth Round Extra	5
VFE	Quarterfinal Extra	5
4RE	Fourth Round Extra	5
3RE	Third Round Extra	5
1RE	First Round Extra	5
ZRH	Additional Round Home	5
ZRR	Additional Round Return (Away)	5

8	Eighth Round	5
H	Home	5
A	Away	5
G	Group Stage	5
E	Group Stage	5
D	Group Stage	5
C	Group Stage	5
B	Group Stage	5
I	Intermediate Round	5
K	Knockout Round	5
L	Group Stage	5
J	Group Stage	5

Table 11: The coding and weighting of each type of match given in the source datasets. It is important that league games have been weighted as 5 regardless of home or away

Goal Difference Range	Importance 1	Importance 2	Importance 3	Importance 4	Importance 5	Interpretation
(-∞, -4)	3	2	2	0	-1	Heavy loss in low-importance match is less detrimental than in high-importance matches
(-3, -2)	3	2	1	1	-1	Moderate loss has varying impact based on importance
(-2, -1)	3	3	2	1	0	Narrow loss has a mixed impact, neutral in very high-importance matches
(-1, 0)	3	3	3	3	1	Draw or narrow loss is generally positive but less so in high-importance matches
(0, 1)	3	2	2	2	2	Narrow win generally has a positive impact
(1, 2)	3	2	3	1	2	Moderate win is highly positive, especially in more important matches
(2, 3)	3	2	2	3	2	Significant win is very positive, particularly in high-importance matches
(3, 4)	3	3	2	2	2	Heavy win is highly positive across all match importance levels
(4, ∞)	3	3	3	3	3	Very heavy win is consistently very positive

Table 12: The weighting of goal differences based on match importance, expanded with interpretation

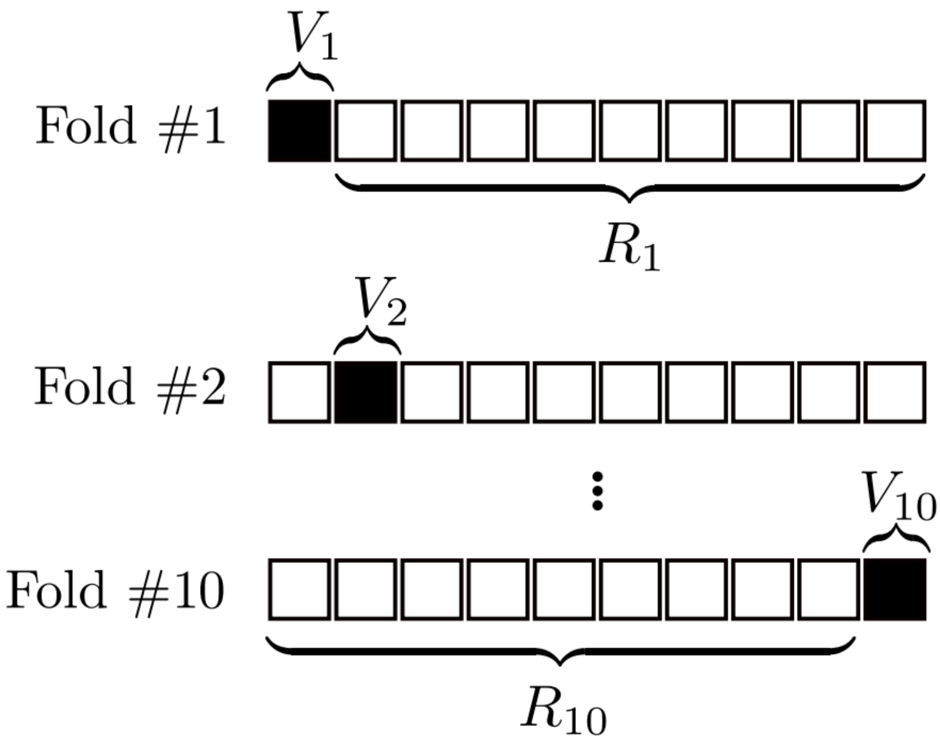

Figure 11: The visual representation of how k-fold cross-validation works. In this example, the data is randomly split into 10 parts, where 1 set (10% of the total data) serves as a validation set.

Parameter	**Description**	**Values**
n_estimators	Number of gradient boosted trees. Increasing this value makes the model more complex and likely to overfit.	50, 100, 150
learning_rate	Step size shrinkage used in update to prevent overfitting. After each boosting step, we can directly get the weights of new features, and the learning rate shrinks the feature weights to make the boosting process more conservative.	0.01, 0.05, 0.1
max_depth	Maximum depth of a tree. Increasing this value makes the model more complex and likely to overfit.	3, 5, 8, 10
subsample	Subsample ratio of the training instances. Setting it to 0.5 means that XGBoost randomly collected half of the data instances to grow trees and this will prevent overfitting.	0.2, 0.4, 0.6, 0.8, 1

colsample_bytree	Subsample ratio of columns when constructing each tree. Subsampling occurs once for every tree constructed.	0.2, 0.4, 0.6, 0.8
gamma	Minimum loss reduction required to make a further partition on a leaf node of the tree. The larger gamma is, the more conservative the algorithm will be.	0, 0.1, 0.2
reg_alpha	L1 regularization term on weights. It can be used to handle high-dimensional data.	0, 0.005, 0.5
reg_lambda	L2 regularization term on weights. It can be used to handle high-dimensional data.	0, 0.1, 1, 2

Table 13: The parameter grid used for the training process of the XGBoost model with the description of each parameter

Paramete	**Description**	**Values**
units1	Number of neurons in the first hidden layer	Range: 256 to 1024, step: 32
units2	Number of neurons in the second hidden layer	Range: 256 to 1024, step: 32
l2_1	L2 regularization factor for the first hidden layer	Range: 0.0001 to 0.001, sampled logarithmically
l2_2	L2 regularization factor for the second hidden layer	Range: 0.0001 to 0.001, sampled logarithmically
learning_rate	Learning rate for the Adam optimizer	0.1, 0.01, 0.001, 0.0001

Table 14: The parameter grid used for the training process of the FFNN model with the description of each parameter

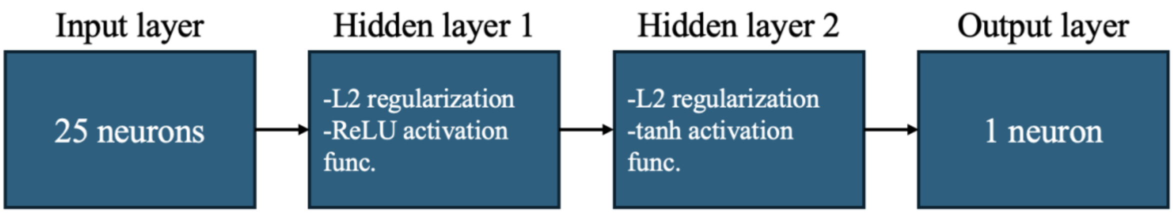

Figure 12: The final internal structure of the FFNN

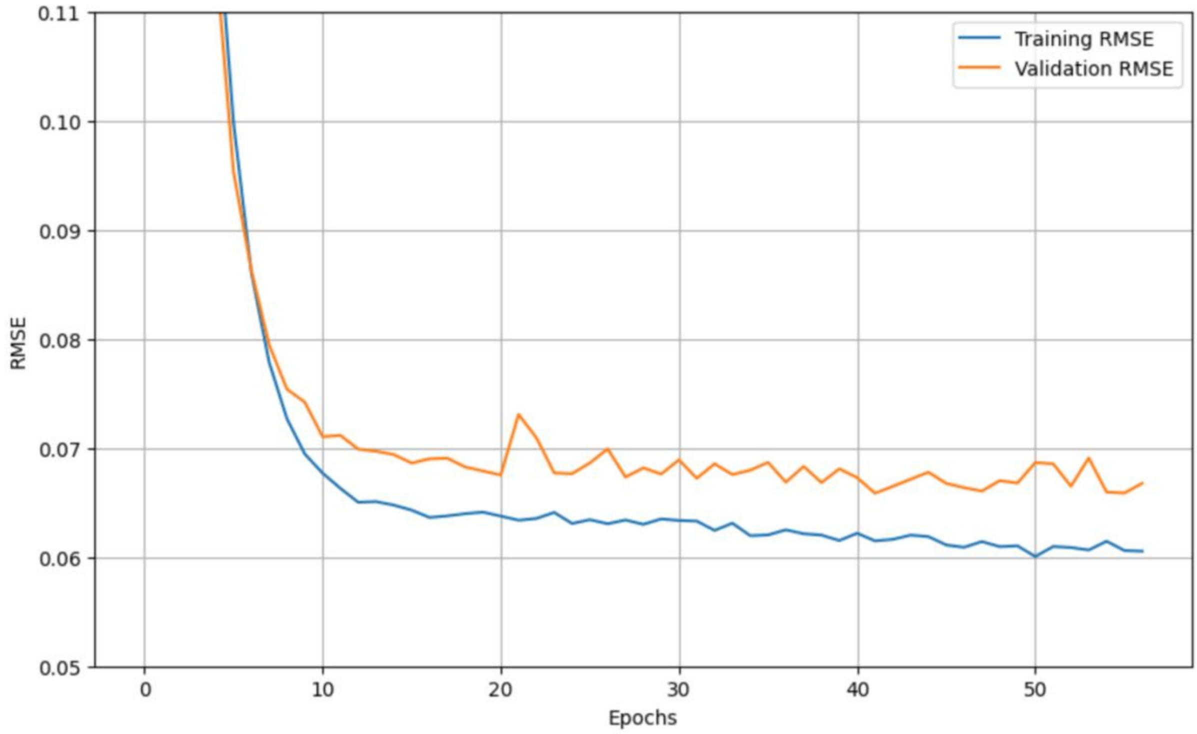

Figure 13: The loss plot of the training process of the FFNN model with RMSE as the evaluation metric

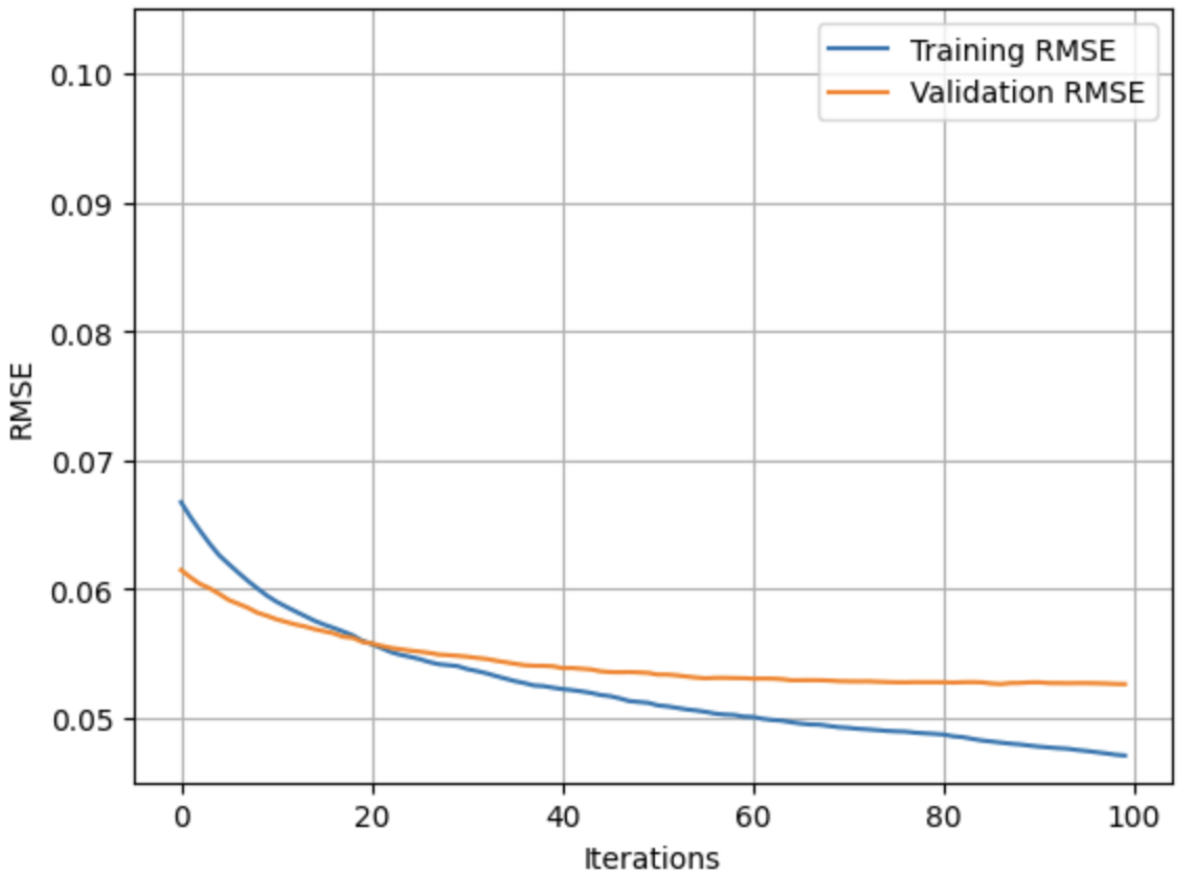

Figure 14: The loss plot of the training process of the XGBoost model with RMSE as the evaluation metric

Model	Stage	R^2	RMSE
XGBoost	Training	52.404%	0.0471
	Validation	28.250%	0.0526
FFNN	Training	28.287%	0.0471
	Validation	16.317%	0.0669
Reference OLS	Training	7.801%	0.0655
	Testing	2.382%	0.0613

Table 15: The final values of the two evaluation metrics (R^2, RMSE) for both models in training and evaluation

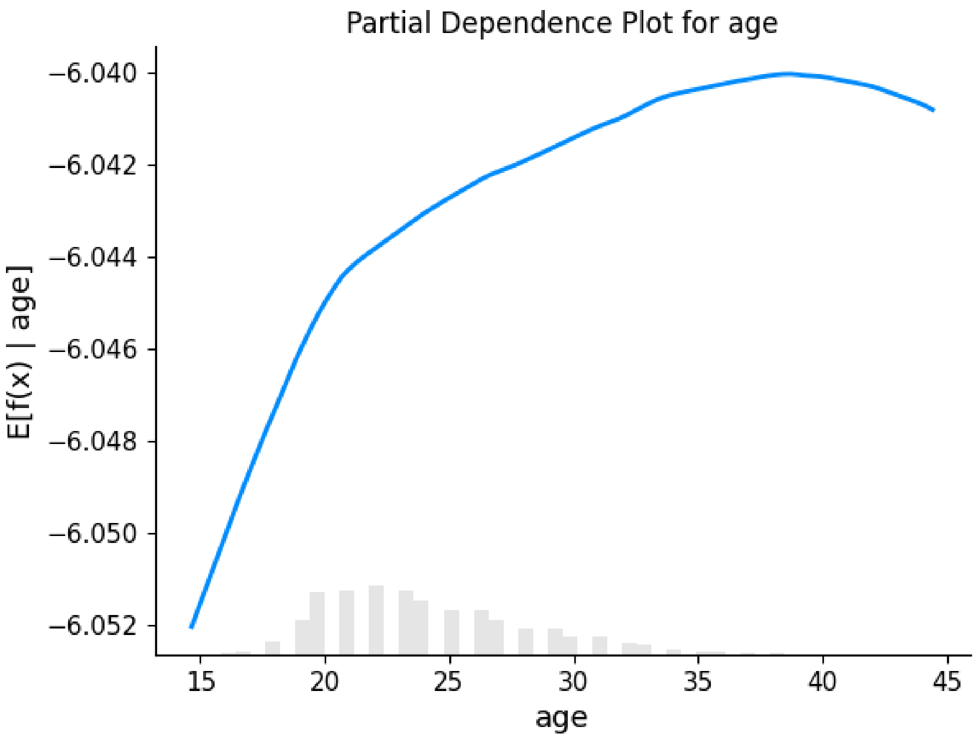

Figure 15: The partial dependence plot of feature "age" for the FFNN model

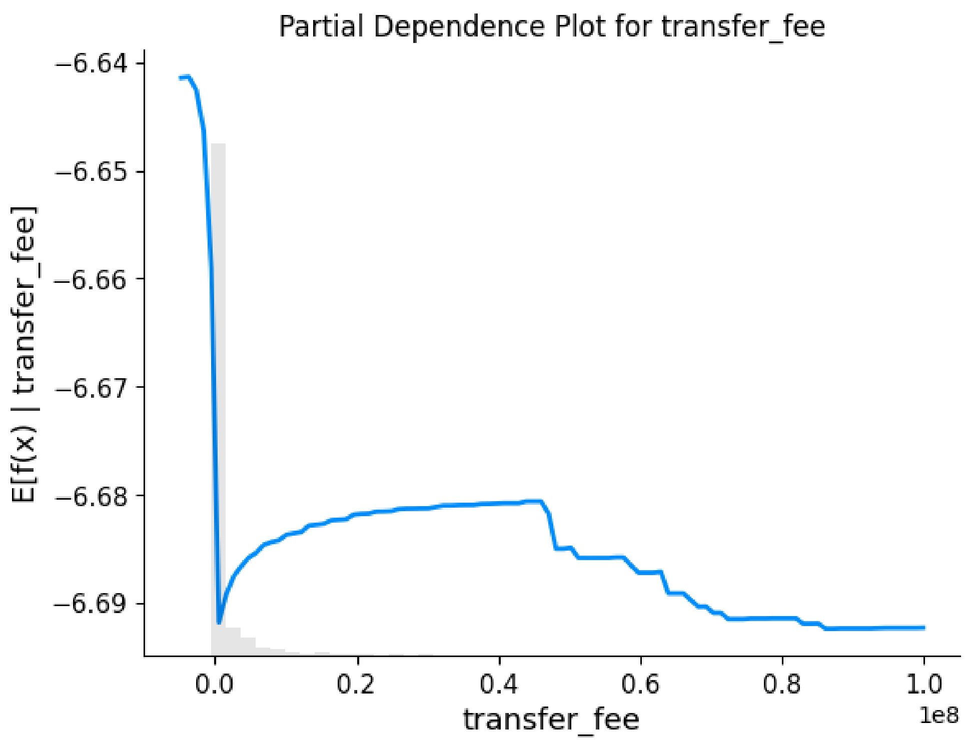

Figure 16: The partial dependence plot of feature "transfer_fee" for the FFNN model

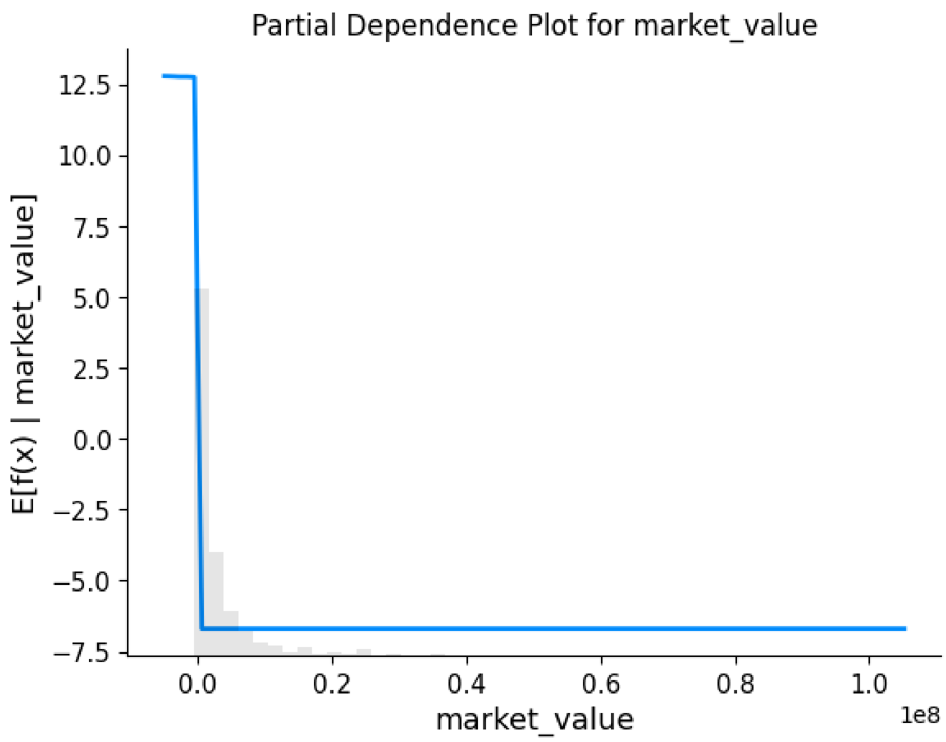

Figure 17: The partial dependence plot of feature "market_value" for the FFNN model

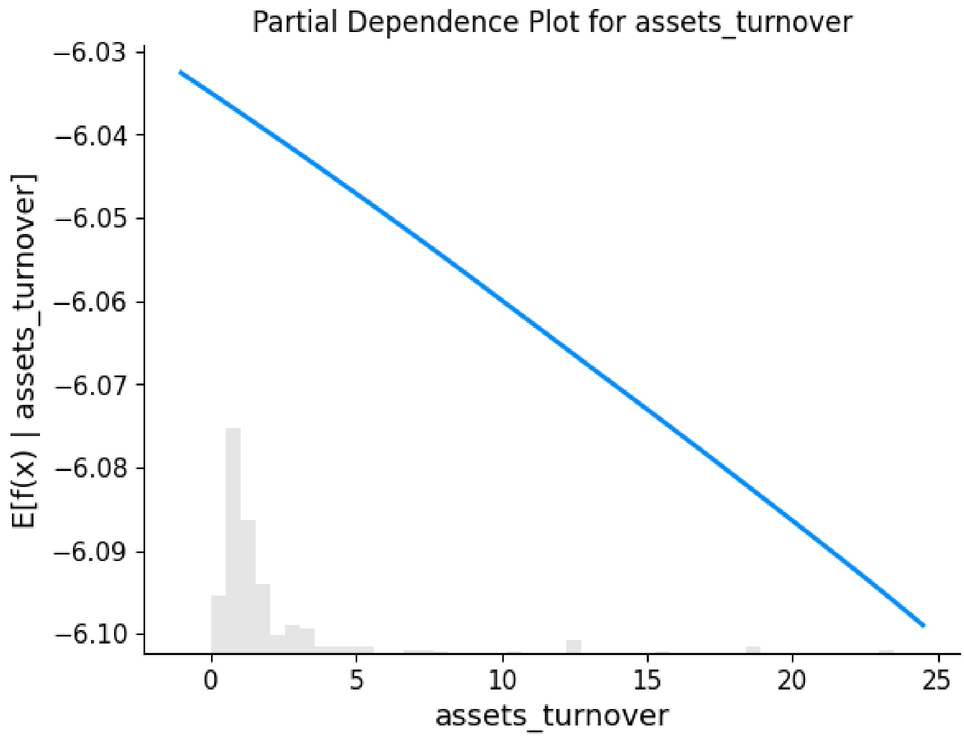

Figure 18: The partial dependence plot of feature "assets_turnover" for the FFNN model

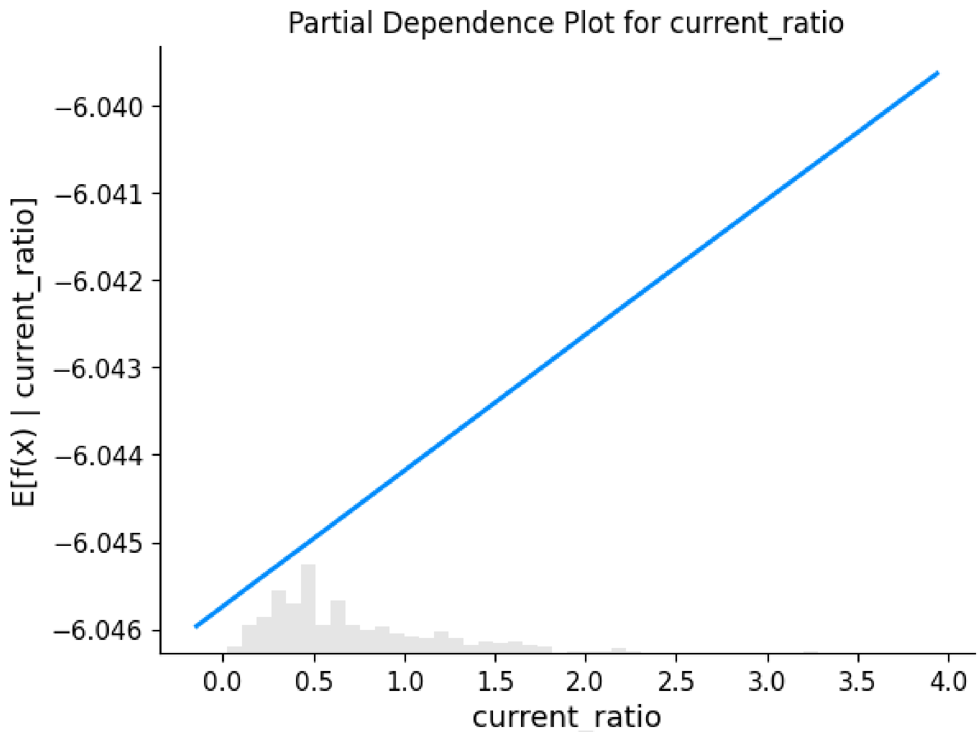

Figure 19: The partial dependence plot of feature "current_ratio" for the FFNN model

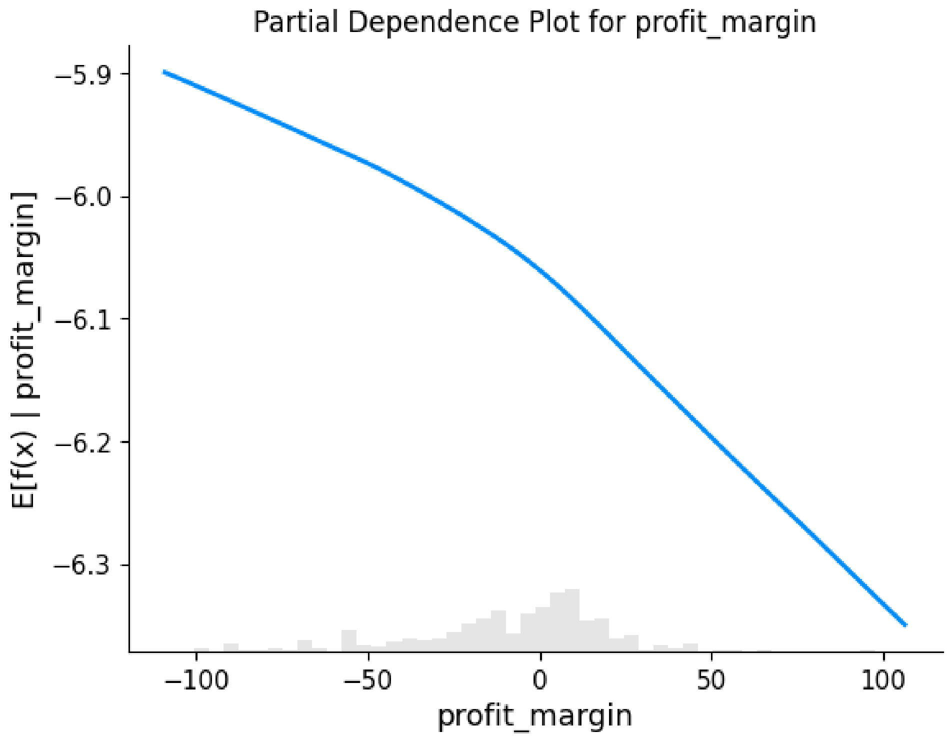

Figure 20: The partial dependence plot of feature "profit_margin" for the FFNN model

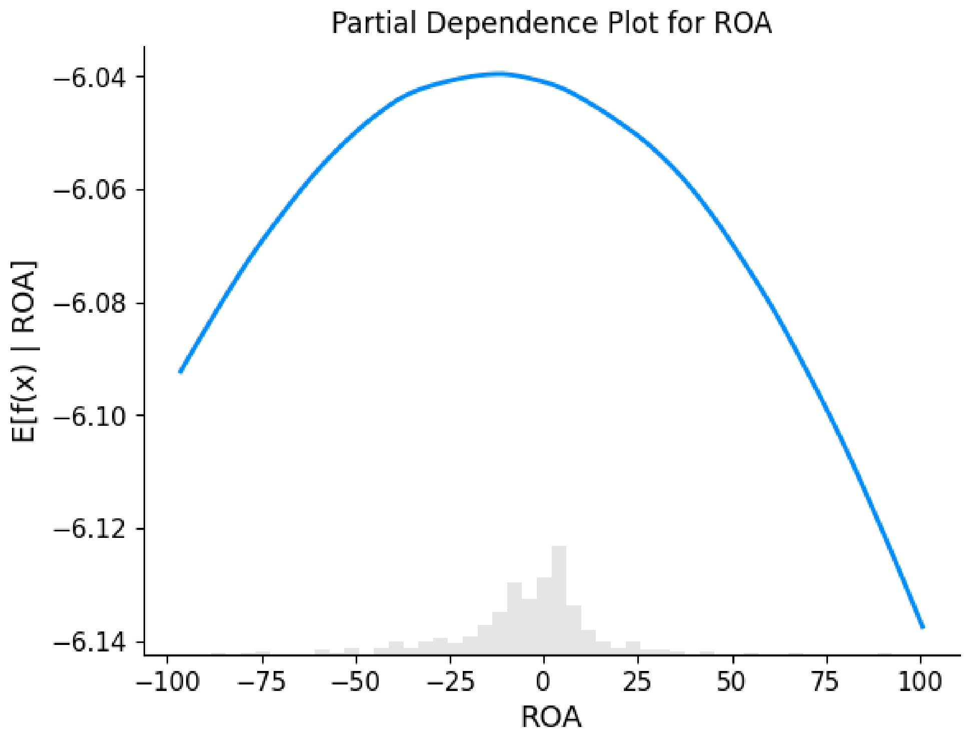

Figure 21: The partial dependence plot of feature "ROA" for the FFNN model

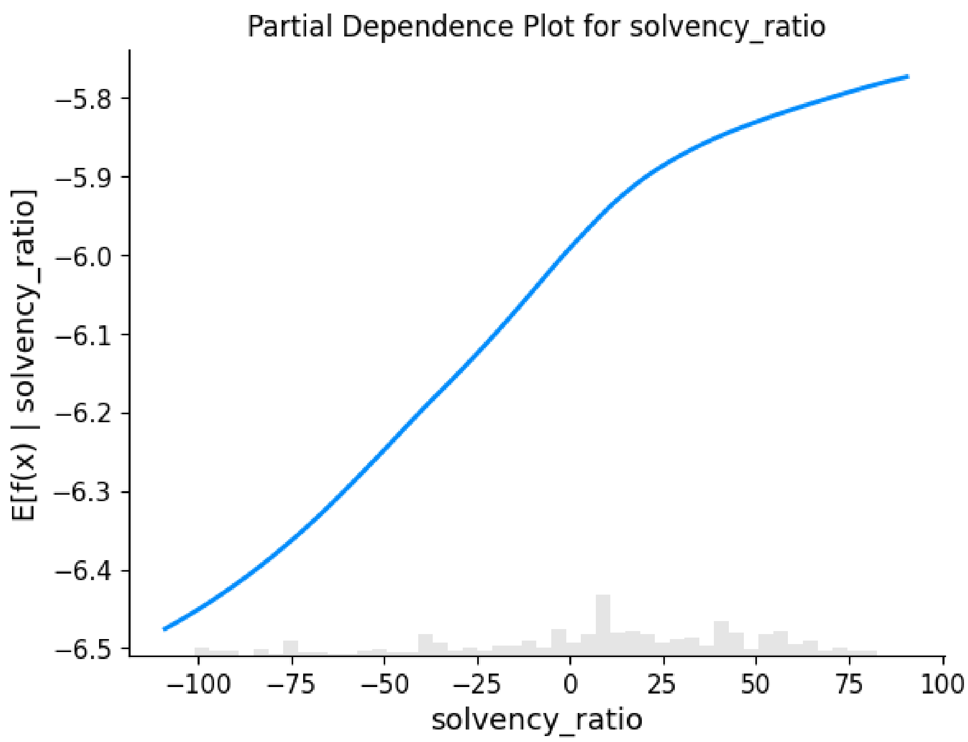

Figure 22: The partial dependence plot of feature "solvency ratio" for the FFNN model

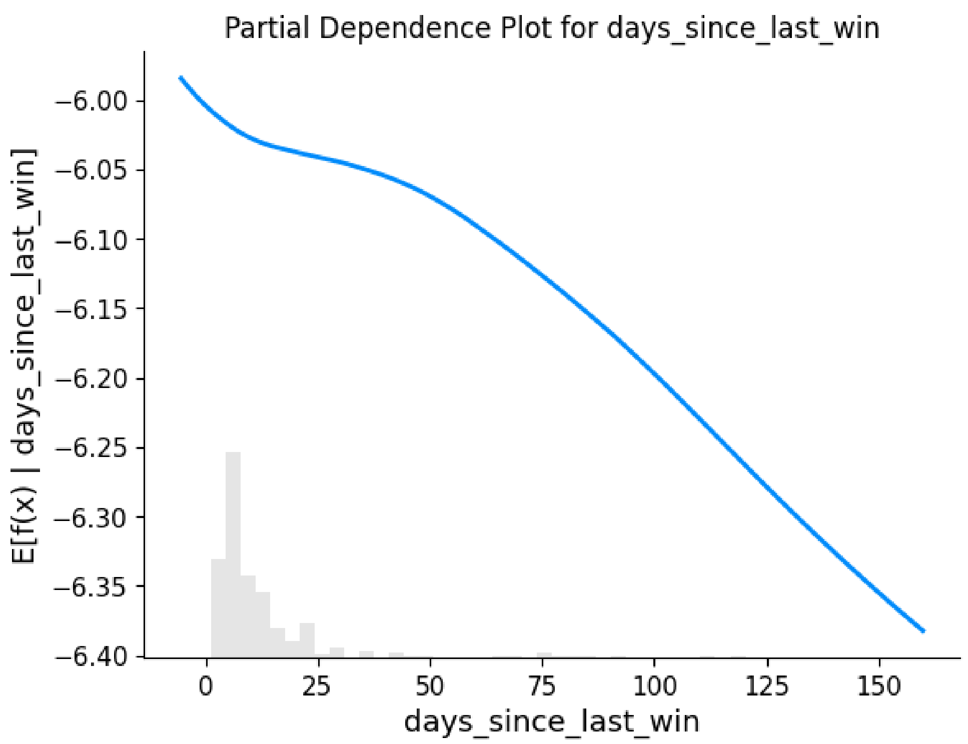

Figure 23: The partial dependence plot of feature "days_since_last_win" for the FFNN model

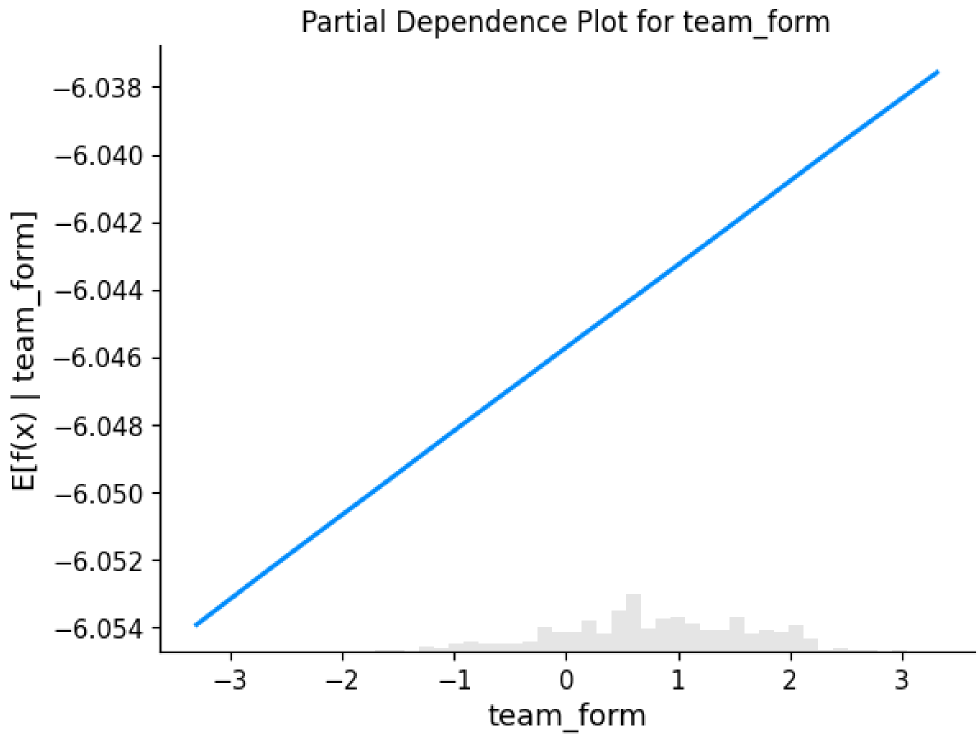

Figure 24: The partial dependence plot of feature "team_form" for the FFNN model