Forecasting the Future Development in Quality and Value of Professional Football Players

van Arem, K.W.; Goes-Smit, Floris; Söhl, J.

DOI
10.3390/app15168916

Publication date
2025

Document Version
Final published version

Published in
Applied Sciences

applied sciences

MDPI

Article

Forecasting the Future Development in Quality and Value of Professional Football Players

Koen van Arem [1,*], Floris Goes-Smit [2] and Jakob Söhl [1]

[1] Delft Institute of Applied Mathematics, Delft University of Technology, 2628 CD Delft, The Netherlands
[2] Data Science & Computer Vision, SciSports, 3703 HX Zeist, The Netherlands
* Correspondence: k.w.vanarem@tudelft.nl

Featured Application

This paper studies what models are most suitable for forecasting future values of player performance metrics in association football (soccer). The resulting forecast statistics find applications in team management and player scouting at football clubs. As transfer decisions concern whether a player should play for a club in the future, the predictions of future performance metrics offer a forward-looking improvement over the traditional backward-looking assessments.

Abstract

Transfers in professional football (soccer) are risky investments because of the large transfer fees and high risks involved. Although data-driven models can be used to improve transfer decisions, existing models focus on describing players' historical progress, leaving their future performance unknown. Moreover, recent developments have called for the use of explainable models combined with methods for uncertainty quantification of predictions to improve applicability for practitioners. This paper assesses explainable machine learning models in a practitioner-oriented way for the prediction of the future development in quality and transfer value of professional football players. To this end, the methods for uncertainty quantification are studied through the literature. The predictive accuracy is studied by training the models to predict the quality and value of players one year ahead, equivalent to one season. This is carried out by training them on two data sets containing data-driven indicators describing the player quality and player value in historical settings. In this paper, the random forest model is found to be the most suitable model because it provides accurate predictions as well as an uncertainty quantification method that naturally arises from the bagging procedure of the random forest model. Additionally, this research shows that the development of player performance contains nonlinear patterns and interactions between variables, and that time series information can provide useful information for the modeling of player performance metrics. The resulting models can help football clubs make more informed, data-driven transfer decisions by forecasting player quality and transfer value.

Keywords: football analytics; football scouting; explainable machine learning; player quality; player value; practitioner-oriented research; player development prediction; uncertainty quantification

Academic Editor: Mark King

Received: 27 June 2025
Revised: 24 July 2025
Accepted: 5 August 2025
Published: 13 August 2025

Citation: van Arem, K.; Goes-Smit, F.; Söhl, J. Forecasting the Future Development in Quality and Value of Professional Football Players. *Appl. Sci.* **2025**, *15*, 8916. https://doi.org/10.3390/app15168916

1. Introduction

Transfers in professional football (soccer) are a risky business because the average transfer fee has increased in recent years [1] and because these fees can be characterized as investments with high risks where large fees are involved [2]. Extensive knowledge about players is beneficial in making well-informed decisions about these complex transfer investments in football. By offering a practitioner-oriented study on forecasting players' future quality and monetary value, this paper offers methods to football clubs for gaining new insights and for improving their strategic transfer investments.

Models providing information about player quality and value have recently emerged with the evolution of data-driven player performance indicators. Improvements in data-capturing technologies resulted in large data sets containing in-game data about football players, which provide the opportunity to obtain more complex variables on player performance [3,4]. Numerous player performance indicators have been introduced since then. An example is the expected goals (xG) indicator, which values shot chances and shooting ability [5–8]. Next to such action-specific models, assessment methods exist for general player performance, which can be divided into bottom-up and top-down ratings [9]. Expected threat (xThreat) [10–12] and VAEP [13–16] are examples of bottom-up ratings that quantify action quality and use the quality of the actions to create general ratings. Top-down ratings such as plus–minus ratings [17–21], Elo ratings adjusted for team sports [22], and the SciSkill algorithm [23] distribute credit of player performance based on the result of a team as a whole. For the monetary value of players, many models about the estimation of transfer fees and market values have been introduced and provide indicators of the current value of football players [24]. These performance and financial models describe the quality and monetary value of a football player, and they can complement traditional scouting reports. This allows managers and technical directors of football clubs to make better-informed transfer decisions.

These models for player quality and transfer value give information about the quality and financial value of football players up to that moment, although a transfer decision regards whether a football player should be part of a team in the future. To make better-informed transfer decisions, team managers and technical directors also need insights into the future development of the indicator values that describe the financial value and the player performance. This paper examines the training of supervised learning models that forecast the development in player quality and transfer value of players.

To this end, two prediction problems are studied: forecasting the quality of a football player and their transfer value one year ahead. In the first prediction problem, models are trained to predict the development of a top-down quality indicator, the SciSkill [23]. The second prediction problem concerns the prediction of the development of the player value, described by the Estimated Transfer Value (ETV) [25]. A forecasting horizon of one year is selected in this study as it aligns with the length of one full season in football and the success of a transfer is often measured by the performances in the subsequent season. The resulting models of these prediction problems offer insight into the question of whether a player will be worth the money in the future. These models thus provide critical insights for the managerial staff of a professional football club.

To further improve the usability of the research results for staff at football organizations, the findings of models should be presented such that they can be utilized by football clubs in practice, as stressed by Herold et al. [4]. This means that models should not only be assessed on predictive accuracy but also on explainability and on methods for uncertainty quantification [26]. To this end, only explainable supervised learning models are used in this research, and the models are assessed on their methods for uncertainty quantification. Although deep learning models are common in forecasting tasks, their lack of explainability

makes them less suitable for application at football clubs, and they are thus excluded in this study. To focus even more on the applicability for practitioners, the accuracy of the models is partly determined by estimating the loss values on different groups of football players, because certain types of football players are more important from a practitioner's perspective.

The aim of this research is to find the most suitable explainable machine learning model to forecast player performance with respect to predictive accuracy and methods for uncertainty quantification. To find this, several explainable supervised learning models are studied. By reviewing the literature, favorable models are identified based on their methods for uncertainty quantification. The predictive performance is assessed by implementing them to forecast the player quality (SciSkill) and the player value (Estimated Transfer Value) of football players one year ahead. The predictive quality is studied on both the general population of players and on subgroups of players that are more interesting for practitioners, such as young or high-value players. An overview of this process is visualized in Figure 1. These results will then be combined to find an answer to the research question.

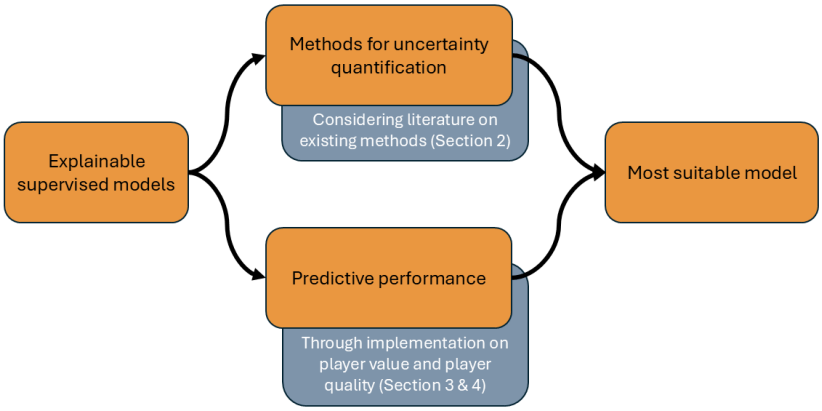


Figure 1. A visualization of the structure of this paper.

The main contribution of this paper is the illustration of how extra value can be added to football player KPIs by forecasting future values in a practitioner-oriented way. The results of this study provide knowledge on what models are most suitable for application by practitioners, lowering the threshold for real-life implementation. By taking explainability, uncertainty quantification, and performance on important subpopulations of players into account, this research is presented such that the results can be used easily by practitioners. In this way, this paper also contributes to bridging the gap between academic research and practitioners.

This paper is organized as follows: The scientific background in the existing literature is reviewed in Section 2. After a discussion of the considered supervised models and their methods for uncertainty quantification, favorable models are identified based on their methods for uncertainty quantification. Then, the literature on data-driven player value and on quality quantification methods is covered, along with research on forecasting these values. The long-term forecasting of player development is then studied for the two prediction problems, for which the methods are described in Section 3. The results of the models in the prediction problems of the player quality and player value are presented in Section 4. The conclusions of this research are then summarized in Section 5, followed by a discussion of the implications and future directions in Section 6.

2. Background

First, the considered models are discussed along with an assessment of their methods for uncertainty quantification. Then, more background is given on existing models for football player quality and value.

2.1. Supervised Models

Supervised learning models provide a possibility to forecast the player performance indicators in the future. Based on input variables, they predict output variables, and they can be trained based on an existing data set for prediction problems. Such a data set $\mathcal{D}_{\text{train}} = \{(X_1, y_1), \ldots, (X_n, y_n)\}$ consists of features X_i describing characteristics about a data point and labels y_i describing the actual values. In this research, the features X_i are values that describe the situation of a player at a specific point in time, and the labels are the player's performance indicator exactly one year later. By constructing these features and labels for multiple points through time and for different players, a data set can be created to train the models.

2.1.1. Considered Models

To align the research with practitioner needs, explainable models are studied in this paper. The considered models can be divided into linear models, tree-based models, and kNN-based models, and they will now be shortly described.

The first linear model considered is ordinary least squares (OLS), sometimes called (multiple) linear regression, as described by (Section 3.2, Hastie et al. [27]). This model assumes that the true relation is linear $y_i = X_i\beta + \varepsilon_i$, where ε_i is normally distributed noise. OLS minimizes the residual sum of squares

$$RSS(\beta) = (y - X\beta)^T(y - X\beta),$$

and has the lowest mean square error of all unbiased estimators. Lasso regression [28] is another linear model that introduces a bias in the model by minimizing the penalized residual sum of squares $RRS(\beta) + \lambda \sum_{i=1}^{p} |\beta_i|$. By introducing this bias, it reduces the variance, and it is possible to provide more accurate predictions. The third linear model is the linear mixed effects (LME) model, which assumes that

$$y_i|b_i = X_i\beta + Z_ib_i + \varepsilon_i,$$

where $\varepsilon_i \sim N(0, \sigma^2\Lambda_i)$ [29]. The random variable Z_i is called the random effect of group i and can be used to model the influence of a specific random attribute that should not influence the predictions, e.g., nationality. Because of their linearity, these models are explainable and suitable for applications.

The CART decision tree (Section 9.2, [27]) is a model that recursively splits the feature space by taking splits of the form $\{X|X_j \leq s\}, \{X|X_j > s\}$ such that it minimizes the sum of squares within each split. As described by Hastie et al. [27] (p. 312), decision trees generally suffer from a high variance. A random forest model solves this by repeatedly training decision trees on resampled data (Chapter 15, [27]). The random forest model then predicts by taking the average prediction of these decision trees. Boosting is another way of improving decision trees. A boosting algorithm fits decision trees sequentially and reweighs the data points for which the model has a bad performance (Chapter 10, [27]). The XGBoost [30] combines several techniques, such as regularization and feature quantile estimates, to provide such a boosting algorithm. Generally, the splitting behavior in tree-based algorithms mimics the if–then reasoning of humans. Additionally, they can provide

feature importances by considering how much a specific feature improves the predictions on the training set. This makes the tree-based models explainable and thus suitable models.

The last type of model considered in this paper is the k-nearest neighbor (kNN) model. Given the features, these models search for the most similar data points in the training data set and predict the associated y_i-value by taking the average of the k-nearest neighbors (Subsection 2.3.2, [27]). The closest neighbors can be weighted more heavily as illustrated by Dudani [31]. The ReliefF algorithm by [32] calculates feature importances by looking at the probability of having different values considering the neighbors of a data point. Additionally, the kNN models can provide examples of training data points to explain predictions. This gives an intuitive interpretation to practitioners, since they can see that a player's forecast performance is based on previous players in similar situations, making the rationale behind the prediction clear and relatable. To interpret a prediction, practitioners can then investigate the players in similar situations, which can help them assess the individual prediction. However, similarity becomes an abstract concept for practitioners in higher dimensions (pp. 108–109, [33]), and the kNN model suffers from the curse of dimensionality (Section 2.5, [27]), which makes application with a large feature set infeasible. To mitigate this in the current paper, kNN models are applied to a feature set of limited dimensions.

2.1.2. Uncertainty Quantification

To better align research with application in practice, the models in this study are also assessed on their methods for uncertainty quantification. Although quantile regression could be applied to obtain a form of uncertainty quantification, this requires an extra model. This makes quantile regression less useful for applications in practice. Therefore, models with uncertainty quantification methods that do not need to alter the model or results are favorable.

Linear regression has an underlying theory that gives prediction intervals as described by Neter et al. [34]. These prediction intervals are based on assumptions that are unlikely to hold for the prediction problems in this paper, such as the normality of errors. Nonetheless, they do give an indication of the uncertainty of the prediction. Similarly, the kNN models provide the neighbors, which are a group of similar data points. Uncertainty quantification can be obtained by taking the minimal and maximal values of the dependent variable within these neighbors if the number of neighbors k is of significant size. Lastly, the bagging procedure of the random forest model can be utilized to obtain uncertainty quantification for the predictions, as described by Wager et al. [35]. This method utilizes the different decision trees in the random forest to quantify the uncertainty in the prediction. From these properties, it is concluded that the linear regression, random forest, and kNN-based models are favorable with respect to uncertainty quantification.

2.2. Existing Indicators

In recent years, the increasing amount of available data in football has driven the introduction of methods to rate individual football players [36]. The performance of a professional football player has traditionally been determined via expert judgment based on video data and statistics describing the frequency of in-game events. Data-driven models have offered the possibility to reduce the bias in assessments and improve the consistency of the judgment of both player quality and player value, as discussed below. In this way, these models have provided new and consistent insights into the quality and monetary value of football players, which has aided in transfer decisions.

2.2.1. Player Performance

The creation and the comparison of models for player performance have revealed new challenges. Because teams can have different aims in football and can apply various tactics, there does not exist a ground truth for player performance [26]. A player can, for instance, be instructed to keep the ball in possession, which leads to the player performing fewer actions that might result in scoring a goal. Because of this lack of ground truth, and various models exist that describe different aspects of the game.

Bottom-Up Ratings

Some models define the players' quality by their actions, called bottom-up ratings [9]. Although action-specific models exist [5–8,37], methods to assess the quality of all types of actions have been introduced that generalize the action-specific models. These general models often define a 'good' action as one that increases the probability of scoring and decreases the probability of conceding a goal.

The VAEP model by Decroos et al. [13] calculates the probability of scoring given the last three actions, including in-game context. Because the involved machine learning techniques are considered a black-box model by practitioners, the authors in [14,15] introduced methods to make the VAEP model more accessible for practitioners. A comparable, more interpretable framework is the Expected Threat (xThreat) model by Rudd [10], which only considers the current situation, defined by the location of the ball-possessing player, to estimate the probabilities of the ball transitioning to somewhere else on the pitch. This is modeled by a Markov chain, which can be used to describe the probabilities of scoring before and after each action to find the quality of an action. Van Roy et al. [11] have compared the xThreat and VAEP models and have found that, although the xThreat model is more interpretable to practitioners, it can only take into account the position of an action and excludes contextual information such as the position of defenders from its model. This means that there is a trade-off between explainability and the inclusion of in-game context when choosing either VAEP or xThreat models. Van Arem and Bruinsma [12] have extended the xThreat model by including variables describing the defensive situation and height of the ball. This Extended xThreat model can take into account in-game context like the VAEP model, while maintaining explainability as an xThreat model.

These bottom-up ratings describe the quality of a professional football player by assessing in-game on-the-ball actions. These on-the-ball actions are often offensive actions, and they are better at capturing the quality of attackers and attacking midfielders, which leads to a bias when these models are studied. This bias makes it harder to assess the quality of defensive players. Moreover, these models need more detailed event data describing in-game events, limiting the scale on which they can be applied. Consequently, the quality of football players in this study is not assessed with bottom-up ratings.

Top-Down Ratings

These problems of bottom-up ratings [9] can be reduced by using top-down models that describe player quality using the lineups and outcomes. Plus–minus ratings are an example of such ratings and were first used in ice hockey and basketball [18], and were later applied to football by Sæbø and Hvattum [17]. For plus–minus ratings, the game is partitioned into game segments that contain the same lineups, which correspond to the data points in the data set. In this data set, the result of the game segment, the goal difference, for example, is the dependent variable. Indicators describing whether a player was active in the segment are the independent variables. Linear regression is then applied to estimate the influence of players on the results. The coefficients of the regression describe the average

impact of a player on the game result and give an indicator for player performance over the period of time covered by the data.

Because substitutions are infrequent in football, a game does not have many game segments with different lineups. Moreover, football is a low-scoring sport. This creates a situation where a low number of segments with limited distinction in outcomes must be used to infer player quality. To deal with this, other quantities have been used as the dependent variable to describe the result of a segment, like the expected number of goals (xG), the expected number of points (xP), and the created VAEP values by Kharrat et al. [18] and Hvattum and Gelade [9]. Pantuso and Hvattum [19] additionally showed the potential of taking age, cards, and home advantage into account, and Hvattum [20] illustrated how separate defensive and offensive ratings can be obtained. Because putting a player in the lineup is an action that a coach performs, De Bacco et al. [21] adapted the plus–minus ratings using a causal model to better describe the influence of the selection of a player. These studies show how plus–minus ratings have been adapted to the application of rating players in football.

Whereas the plus–minus ratings describe the quality of a player using multiple historical games, there also exist models that determine the quality of a player after each game using the lineups and final score. Elo ratings are such ratings and were originally developed to evaluate performance in one-on-one sports. The concept was subsequently adapted to the game of football by Wolf et al. [22]. This adapted algorithm provides ratings for each individual football player and calculates the team rating via the average of players in a game weighted by the number of minutes played. The ratings are then used to predict the match outcome using a fixed logistic function, and after each game, the individual ratings are adjusted. If the outcome is better than predicted, the player rating is increased, and if the outcome is worse than expected, it is decreased. The authors also introduced an indicator for player impact to deal with the fact that this rating undervalues good players at below-average teams.

The SciSkill [23] is an industry-validated, more elaborate version of the Elo rating. Instead of only considering one player's quality, it describes football players combining a defensive rating and an offensive rating. These offensive and defensive scores are then combined to obtain one value, the SciSkill. For each game, the outcome is predicted using a model via an expectation-maximization algorithm. After the prediction, the SciSkill is updated by adjusting the SciSkill values based on the difference with the actual game result as with other Elo-ratings. Compared to the Elo algorithm, which was implemented and validated by Wolf et al. [22], the SciSkill model extracts more detailed information from the matches and describes the player quality more elaborately.

2.2.2. Player Value

In contrast with player quality, there does exist a ground truth for the concept of player values. The value of a transfer fee is based on the value of the player for each of the involved clubs and historical transfers of similar players [38]. Thus, the historical values of these fees can be used to predict the transfer value of a player, albeit at the cost of some selection bias.

A well-known medium that describes the value of a player is Transfermarkt. This company uses crowd estimation to assign values to players [39]. These market values describe the general value of a player and do not take into account the temporary situation of a football player, like the current club and contract length. This means that they describe a different quantity than the expected value of a transfer fee. Nonetheless, these values are strongly correlated with the real transfer fees as shown by Herm et al. [39]. Therefore, both

market values by Transfermarkt and transfer values are often used interchangeably when describing the monetary value of a football player.

The literature review by Franceschi et al. [24] showed that many studies have been performed to describe the monetary value of a football player based on data. The authors found that most of these studies used linear regression to find which variables have a significant linear dependence on the value of a football player. The review by Franceschi et al. [24] considered 111 trained models that were used in the scientific literature to investigate the transfer value of a football player. The vast majority (85%) of the models are based on ordinary least squares. The authors have also shown the importance of different variables in the considered models. For instance, they have found that age, the square of the age, and the number of matches played by a player are frequently studied variables. These variables are also most often found to be significant. Players frequently increase in value as they get better with experience that is gained over the years, but they also decrease in value as a player loses the potential to improve when getting older. Consequently, the influence of age is often measured with a quadratic term. Similarly, the number of games played can be expected to be an important variable because players gain experience by playing games, which makes them more valuable. Moreover, players who play a lot of games are often the better players on a team. This explains why these variables are important, as found by Franceschi et al. [24]. Their study also shows that almost no variables describing defensive behavior were considered when other researchers trained models to describe transfer values. As a consequence, the resulting models can be expected to describe the value of offensive players better than that of defensive players. In addition, the linear models in these studies are mostly trained to determine the influence of variables on the transfer value of a football player, and they are generally not trained and tested for out-of-sample prediction, which limits the application of predicting based on new, unseen data.

In contrast, Al-Asadi and Tasdemır [40] have trained multiple models to predict market values of football players with features from the video game FIFA. They have found that a random forest model gives improved predictions over linear methods. Steve Arrul et al. [41] have studied the application of artificial neural networks for the same problem, also considering features from the video game FIFA. They obtain similar loss values as the random forest model of Al-Asadi and Tasdemır [40]. The research by Behravan and Razavi [42] introduces a methodology to train a support vector regression model via particle swarm optimization for the prediction of market values. Although the data set is somewhat similar to the studies above, this model attains worse loss values. In the study of Yang et al. [1], random forests, GAMs, and QAMs are applied to predict the transfer fees of players based on variables describing the player. The authors have inferred from their random forest model that the expenditure of the buying club and the income of the selling club are important features in predicting the transfer fee, as well as the age and the remaining contract duration. This research additionally illustrates how GAMs and QAMs can be used to investigate the dependency between the player transfer value and the given features. The QAM models show that this relation varies for different quantiles, indicating the need to study the influence of models on different groups of players. On the other hand, the GAMs show that the relationship between the transfer values and the features is often nonlinear. A study with extensive types of player performance metrics as features has been performed by McHale and Holmes [2], in which linear regression, linear mixed effects, and XGBoost models have been trained. These models not only include statistics such as the number of minutes played, height, and position, but also plus–minus ratings based on xG, expert ratings from the video game FIFA, and GIM ratings, which are similar to VAEP ratings. The results show that the best predictive performance is attained by the XGBoost model, although the linear mixed effects with the buying and selling clubs as random effects also

provide good results. Their results indicate that their model outperforms Transfermarkt market values when predicting the transfer fees on average, although the market values are a better predictor for transfers of more than EUR 20 million. These studies show how supervised learning can be used to obtain models that predict the value of a football player. They show that nonlinear methods generally predict more accurately and that the patterns can differ for different groups of players.

2.3. Predicting Future Values

Although the studies discussed above are concerned with the quantification of the quality and monetary value of football players at the present moment, only limited work has been conducted on the future development of player performance. Apostolou and Tjortjis [43] have conducted a small-scale study with the aim of predicting the number of future goals of the two football players Lionel Messi and Luis Suárez using a random forest, logistic regression, a multi-layer perceptron classifier, and a linear support vector classifier. Pantzalis and Tjortjis [44] have predicted the expert ratings in the next season of 59 center-backs in the English Premier League based on one season of player attributes from a popular football manager simulation game. Their method uses a linear regression model to describe the in-sample patterns. Giannakoulas et al. [45] have trained linear regression, random forest, and multi-layer perceptron models to predict the number of goals of a football player in a season before the start of the corresponding season. Their data set entails around 800 football players. Similarly, the models by Markopoulou et al. [46] have been trained to predict the number of goals by looking at the creation of different models per competition. This study on 424 football players shows that the best results are often obtained using XGBoost models for this prediction problem and that making different models for different competitions might be beneficial.

Barron et al. [47] have tried to predict the tier within the English first three leagues in which a football player would play next season as an indicator of player quality. Three artificial neural networks are trained to predict in which league a player would play with the data of 966 football players. Their models are only able to recognize the differences between players in the lowest and highest tiers (League One and the English Premier League).

Little literature exists about the prediction of the future transfer values of football players. Baouan et al. [48] have applied lasso regression and a random forest model to identify important features for the development of around 22,000 football players. They have trained these supervised models for players of different positions to predict a player's transfer value two years in the future based on performance statistics. The feature importances of the models show, for instance, that the average market value of a league is an important feature for the future values of players in that league. Although cross-validation is performed for the hyperparameter tuning, this study focuses on finding in-sample patterns.

Our research treats forecasting of both player quality and transfer value. The current paper builds on the existing literature about forecast player performance by studying the long-term forecasting of a model-based player quality indicator on a larger data set. Additionally, this research avoids the bias in the current literature that better describes offensive players by using a more general top-down rating. Our research contributes to the existing knowledge of forecasting the monetary value because the models are trained to perform out-of-sample prediction, which makes it possible to apply them to unseen situations. Moreover, the combination of forecasting the development of both player quality and transfer value gives a comprehensive summary of the most important aspects in transfer decisions. In this way, this paper builds upon the existing literature by forecasting

the development of model-based indicators for player quality and value on a significant data set in a predictive setting.

3. Methods

The goal of this study is to find the most suitable machine learning model to forecast the development in player performance with respect to predictive accuracy and uncertainty quantification methods. The assessment of the uncertainty quantification methods was performed using the literature on the models in Section 2. To compare the predictive performance, the models are trained to predict player performance indicators one year ahead in two prediction problems, and the corresponding loss values are determined. The two prediction problems concern the prediction of the development in player quality based on the top-down SciSkill rating and the development in monetary value described by the Estimated Transfer Value (ETV) model.

3.1. Data

Two data sets were obtained to study the predictive performance of the models in this paper. The features of the data points in both prediction problems represent the historical situation of professional football players. The corresponding label is the player performance indicator that was recorded one year later. The data availability for different years is visualized in Figure 2. For the player quality prediction problem, data is available from 2014 up to 2022, while the data for the player value covers the years 2016 up to 2021.

Since the used SciSkill and ETV models are proprietary, exact reproduction of the results using the same models is not possible. However, similar models exist for both the SciSkill [22] and ETV models [1,2,40–42], as discussed in Section 2, ensuring that the overall approach remains, in principle, reproducible.

Figure 2. A visualization of the partition of the years in the test and training set in both prediction problems.

3.1.1. Player Quality: SciSkill

The first prediction problem concerns predicting the development in the subsequent year of the player quality described by an EM algorithm called the SciSkill [23], which is a generalization of the Elo rating, as discussed in Section 2. The data set is restructured to contain monthly data points describing the situation of each player at that time. Each monthly data point consists of the features and the dependent variable. The dependent variable is the difference between the player quality one year ahead and the current player quality value, which is the development in player quality in one year. The features set consists of 86 features constructed with domain knowledge describing, for instance, the month of year, the current player performance, league strength, time since the most recent game, player characteristics, time series information of the SciSkill, the club's transfer situation, and the difference in quality between the player and his teammates. The dependent variable and features are described in Tables S1–S3 in the Supplementary Materials.

Only male players with more than 20 games and more than 2 years of data are considered. This filtering might introduce a bias in the data against player groups on which less data is available, such as younger or injury-prone players. However, it is necessary to obtain enough data points per player. The results on players with limited data availability should thus be interpreted with care. The final data set consists of 80,568 male professional football players playing in the years 2012 to 2023. As the data set consists of 3,834,539 data points, there are on average 47.6 monthly data points per player, which corresponds to roughly 4 seasons of data.

3.1.2. Player Value: Estimated Transfer Value

In the second prediction problem, the monetary player values are considered, obtained by the Estimated Transfer Value model (ETV). This model is a supervised tree-boosting model, trained on historical transfers to predict the transfer fees based on features that describe the situation of the player at the time of transfer. The model is then used to describe the transfer value for professional football players to obtain the transfer values for the general population of players over time. In this way, the supervised ETV model provides the monetary values of players over time.

The data set for this prediction problem of the development in player value describes the current situation of professional football players with 58 features. These features were constructed based on domain knowledge and data availability. Because there were fewer data points than for the player quality case study, the number of features was reduced by eliminating highly correlated features. These features consist of indicators of league strength, age, experience, contract situation, the current quality (SciSkill), or the monetary value (ETV) of a football player. The feature set also includes player characteristics, like playing position and age, as well as the differences in quality with teammates, the transfer history of his current club, and league strength. A description of all features and the dependent variable in this prediction problem can be found in Tables S4–S6 in the Supplementary Materials.

For this prediction problem, the data set consists of biannual data points describing the players' transfer values in January and July within the period from 2014 to 2021. Similarly to the SciSkill, players with less than 20 games, less than 2 years of data, or missing values are excluded. The remaining data set includes 60,175 male professional football players described with 413,177 biannual data points. This corresponds to an average of 6.87 data points per player, equivalent to approximately 3.5 seasons of data.

3.2. Model Implementation

3.2.1. Model Assessment

Supervised models are trained on the prediction problems of player quality and player value. The root mean square error (RMSE) and mean absolute error (MAE) are determined for the test set. The RMSE is more strongly influenced by large errors, which can be expected to happen for 'superstar' players. Since these players are especially of interest to practitioners, the performance in the RMSE is considered the most important.

The losses are determined on different parts of the test set. First, the loss values of the RMSE and MAE are calculated using the complete test sets. Second, the loss function of the RMSE is also considered for different age groups on the test sets because estimating the potential of a player is mostly interesting for young players. Third, the RMSE is studied on important subgroups of players, like players with large positive or negative development, players with good performance, or players with a high transfer value. By determining the test losses separately on the general population of players, young players, and important subgroups of players, the models are studied on their predictive performance.

Beforehand, 5% of all football players are left out of both data sets based on stratified sampling for internal studies by the data provider. Because the data of player performance indicators is often dependent on time [16,26], time-dependent train–test splits are applied to study the predictive performance on unseen data, as visualized in Figure 2. In both case studies, all data up to 2020 is considered as the training set, and from 2021 and later as the test set.

3.2.2. Model Training

In this study, linear, tree-based, and kNN-based models are implemented. The linear models are ordinary least squares (OLS), also known as multiple linear regression (MLR), lasso regression, and a linear mixed effect model (LME). For the OLS model [49], feature selection is performed by applying backwards selection with a threshold of the p-values of 0.0001 for the player quality and 0.001 for the player value. These values are smaller than commonly used for significance testing due to the predictive nature of this study and the large data sets combined with the fact that common assumptions of the linear regression model, like normality, do not hold. These p-values were selected via trial and error, considering values of the form 10^k, such that roughly half of the features is excluded. For lasso regression [50], feature selection is applied by selecting only the variables with nonzero coefficients. The linear mixed effect (LME) model [49] is trained to take into account the influence of a player's nationality as a random influence. The feature selection for the LME model is performed by taking the 20 variables corresponding to the largest feature importances within the lasso model. This is conducted for computational feasibility. For these three linear models, no interaction effects are included.

The tree-based models are the decision tree [50], random forest [50], and XGBoost [30] models. For the tree-based models, feature selection is performed by first adding noise variables to the data set and training the models. All features with a larger feature importance than the noise variables are then selected. As tree-based methods can have a bias favoring non-discrete variables, both discrete and continuous variables are added. The discrete features are compared with the discrete noise variables, and the continuous features with the continuous noise variables.

To investigate the predictive power of time series, three k-nearest neighbors (kNN) models have been implemented using the Hierarchical Navigable Small Worlds indexer provided by Douze et al. [51]. This was carried out on an altered feature set consisting of time series information with lagged versions of the most important player indicators. Because the kNN model suffers from the curse of dimensionality (Section 2.5, [27]), the feature set is kept relatively small. The features and labels of this altered data set are described in Tables S3 and S6 in the Supplementary Materials. First, a normal kNN model is applied to the predictive problem. This model searches for the most similar data points with respect to the Euclidean norm. The prediction is then obtained by applying a weighted average, where closer data points are weighted more heavily as described by Dudani [31]. Possible weighting methods are the reciprocal of the absolute value of the distance, the distance with min–max scaling, or uniform weights. The method of calculating these weights is considered a hyperparameter. The second kNN model is constructed in the same way as the first, but it calculates the distances based on the Mahalanobis distance instead of the Euclidean distance. The Mahalanobis distance projects the features onto a decorrelated feature space and calculates the Euclidean distance in this changed feature space. This makes it possible to better distinguish differences because the lagged time series features are heavily dependent. Lastly, an adapted RReliefF method [52] is implemented to calculate feature importances of the normal kNN model in a regression context. The features are then multiplied by the feature importances before calculating Euclidean distances to introduce

new feature weights. The kNN model with Euclidean distance is then trained on these reweighted features.

3.2.3. Hyperparameter Tuning

To select the best hyperparameters of the models in the case studies, Bayesian optimization [53] is applied to find hyperparameters minimizing the RMSE loss. Because of the time dependencies of player performance indicators, cross-validation with an expanding windows split was implemented. This time series split strategy splits the data per year and incrementally grows the training set as visualized for the player quality prediction problem in Figure 3. After the hyperparameter tuning, the models are trained on the complete training set with the optimized hyperparameters.

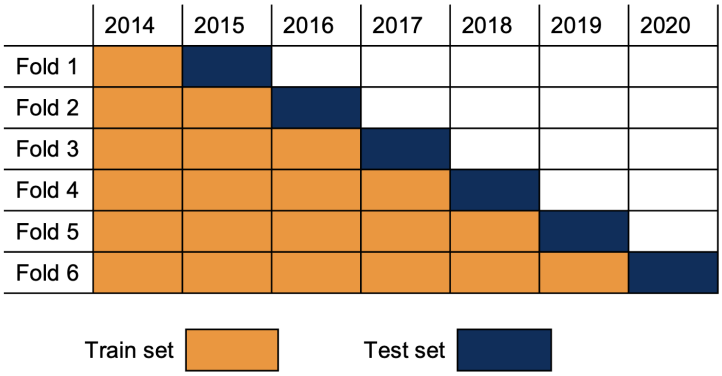


Figure 3. A visualization of the distribution of the data points in the test and training set for the player quality prediction problem of the adjusted version of cross-validation used for hyperparameter optimization.

3.3. Feature Importances

To illustrate the advantages of using explainable models, the feature importances are studied in both predictive problems. Among the models in this paper, the linear and tree-based models have methods that can be used to calculate the feature importances of the models. To investigate what features are important for the development of professional players, the feature importances are calculated. This is similar to the methods by Baouan et al. [48]. However, because the development of the indicator is considered instead of the indicator value itself, it is possible to describe the influence of the indicator value on the development. Min–max scaling is applied to the feature importances to be able to compare them across the different models.

4. Results

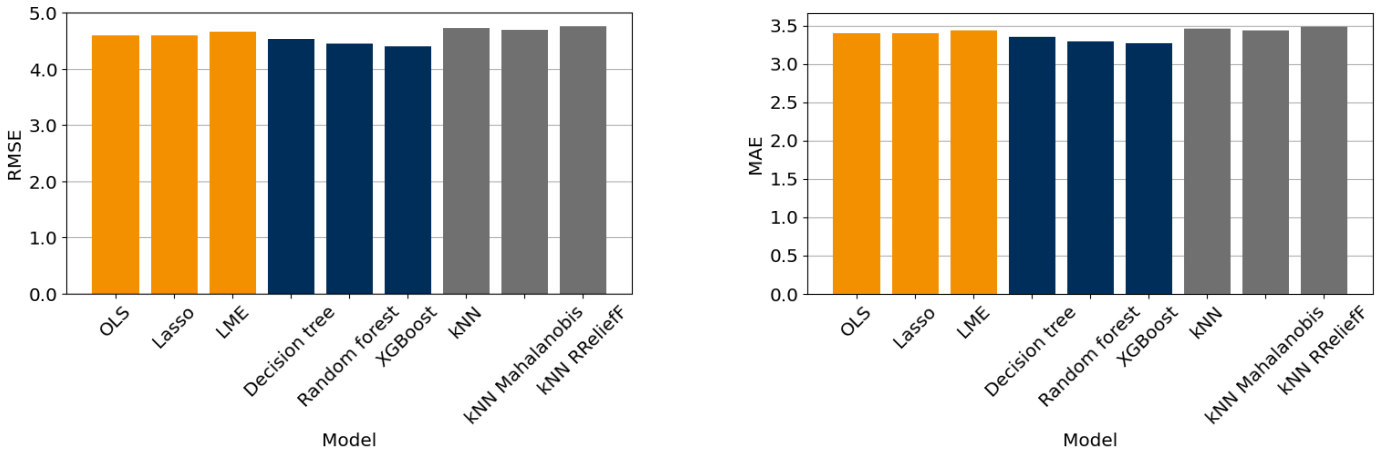
4.1. Prediction Problem on Player Quality

4.1.1. General Population of Players

The loss values on the test set are shown in Figure 4. The RMSE and MAE agree on the predictive performance of the models to forecast player development in general player quality, as they show similar patterns. The results indicate that the XGBoost model attains the lowest loss values of all models. The decision tree and random forest models also attained low loss values, with the latter having slightly better loss values.

The tree-based models generally obtain the lowest loss values. The difference in performance with the linear models implies that there is some nonlinear or interaction effect in the true underlying relation. On the other hand, the kNN models based on the time series attain the worst loss values. This implies that these kNN-based models missed out on important information by relying on time series information with a local

method. It shows that the development of player quality is nonlinear and dependent on contextual information.



(**a**) (**b**)

Figure 4. The test loss values for the different models on the general population of players in the player quality prediction task for two different loss functions: (**a**) RMSE and (**b**) MAE.

4.1.2. Predictions per Age

The RMSE loss is also determined for the players of each age in the test set, as visualized in Figure 5. In general, Figure 5 shows that all models predict best on players with ages 24 up to 28. The quality of younger or older players is more volatile because young players tend to increase in quality, and older players tend to decrease in quality, albeit over a longer time period. Additionally, more data is available on players around their peak age. The data set consists of roughly 10,000 data points for players aged 18 and 20,000 for players aged 34, while around 70,000 data points describe situations of players aged 25. The worse predictive performance on the younger and older players can be explained by the lower number of data points combined with a higher volatility in player quality. Moreover, the low number of data points for young players indicates that the RMSE estimates for young players may be less stable than those of players aged 24 up to 28.

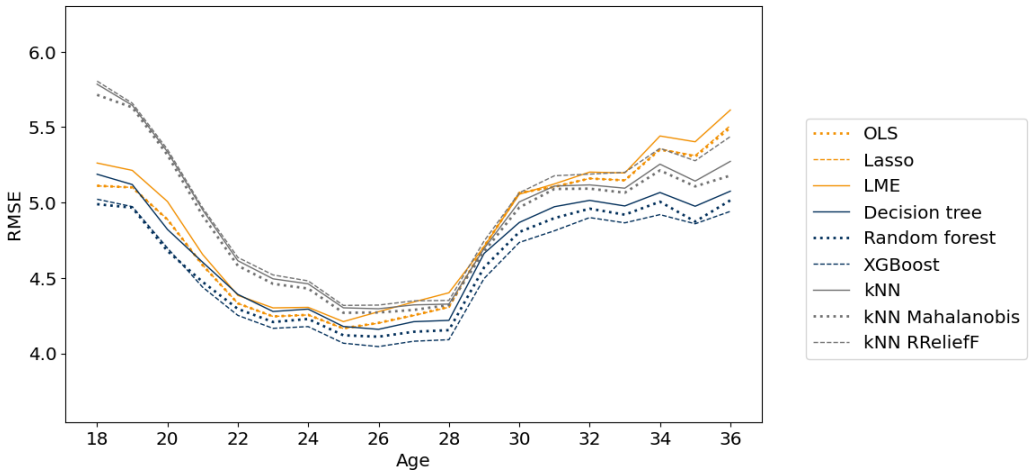


Figure 5. The RMSE values for each age for the different models in the player quality prediction problem.

The results show that the XGBoost model attains the lowest loss values for almost all ages, especially for ages above 22. For the ages below 22, the random forest model

attains similar loss values. Because the younger ages are most important for estimating the potential of players in practice, these results indicate that both the XGBoost and random forest models are favorable models.

Additionally, the results show that the kNN models perform worse on young players. Generally, younger players are harder to model because of their volatility. Combined with the fewer available data points, this explains why the local kNN models can predict the development of younger players less well. Conversely, the linear methods tend to perform worse for older players. This indicates that the patterns for older players involve nonlinearities or interaction effects. Because tree-based models are non-local methods that can take into account nonlinearities and interaction effects, they do not suffer from these problems. This explains the low loss values of the tree-based models on the general test set as shown in Figure 4a,b.

4.1.3. Prediction on Important Player Groups

The test losses are determined on three different groups of players that are important for the application of the models in this prediction problem: high-quality players, players with a large decrease in performance, and players with a large improvement in performance. Based on domain knowledge, high-quality players were defined as players with a SciSkill of more than 100, while an increase or decrease of more than 10 was defined as large. The RMSE values on the test set for these subgroups are shown in Figure 6. In general, the vertical scales show that the RMSE on these subgroups is larger than that of the general population. This is even more evident for players with a large increase, and it indicates that the development of these players of interest is harder to predict.

The results also show that the tree-based models are best at predicting the development of all three of these subgroups of players. The XGBoost model attains the best RMSE values for all three of the important groups of players. Figure 6c shows that the XGBoost model outperforms all other models on the group of players with a large increase in quality, whereas the differences in loss values with the decision tree and the random forest are less apparent in Figure 6a,b.

The results indicate that the linear models predict significantly less accurately for the group of players with a large decrease in quality compared to the other models. This shows that interaction effects or nonlinearities are particularly important for predicting a decrease in player quality. On the other hand, the kNN models based on time series predict significantly worse for players with large increases in quality, as the large increases are probably for the young players. This is in line with Figure 5, where kNN methods do not perform well on the young players. These differences in losses show that the large increases in player quality are better predicted by global models with contextual information and that decreases are better predicted by methods that include nonlinearities or interaction effects.

4.1.4. Feature Importances

The feature importances of the linear and tree-based models are shown in Figure 7. The feature importances indicate that the age (‘age_years’), age squared (‘age_years_squared’), and the difference between the age and the peak age (‘years_diff_peak_age’) are important features amongst the models. These features are strongly correlated, as they all capture age-related effects. As a consequence, the models often identify at most two of these features as important. The importance of age-related features is in line with domain knowledges since it is known that young players tend to improve in quality, whereas older players often decrease in quality. Therefore, it is reasonable that these age-related factors are indeed important in the development of the quality of professional football players.

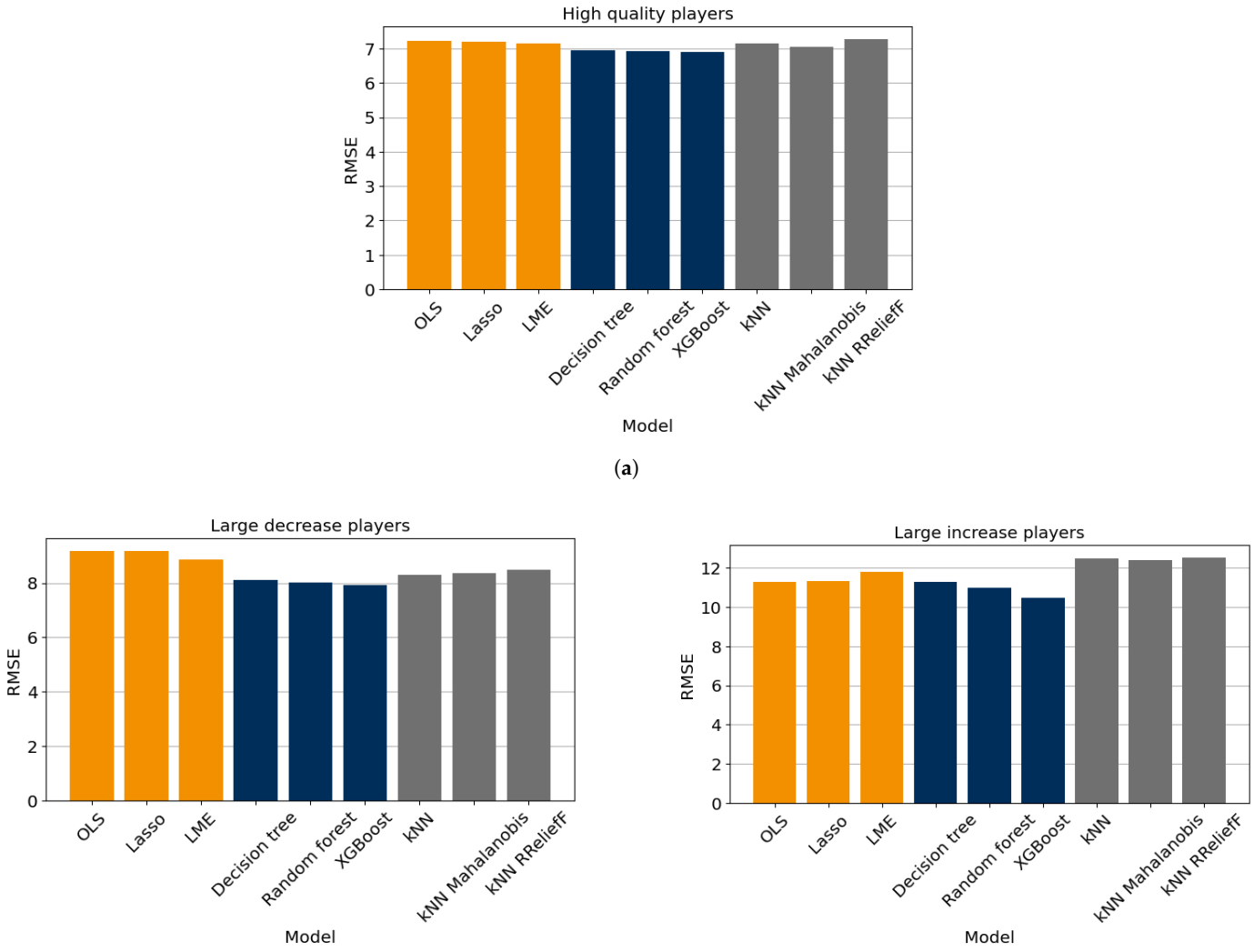


Figure 6. The RMSE loss per model in the player quality prediction task for various player subsets: (**a**) players with a SciSkill of at least 100, (**b**) players with a SciSkill decrease of at least 10, and (**c**) players with a SciSkill increase of at least 10.

Next to that, other features have also been assigned a large importance by the models. For instance, the player quality at prediction time ('sciskill') has a large feature importance. This indicates that the development patterns depend on the player level, which is reflected by the larger loss values in Figure 6a compared to Figure 4a. The importance of the difference between player quality and average team quality ('sciskill_diff_mean_team') can be explained by the fact that players tend to have good performances when their team plays well. Consequently, the player's quality grows towards the average team quality, which explains the importance of the difference in quality between the player and his team. Lastly, the feature importance assigns a large importance to the number of months since the last registered game ('previous_zero_months'). The SciSkill model penalizes players when they have not played games for a long time. As this penalty is applied after the next game of a player, the quality of a player can be expected to decrease when the number of months since the last game is large. The importance of the number of months since the last registered game shows that the models can find this pattern.

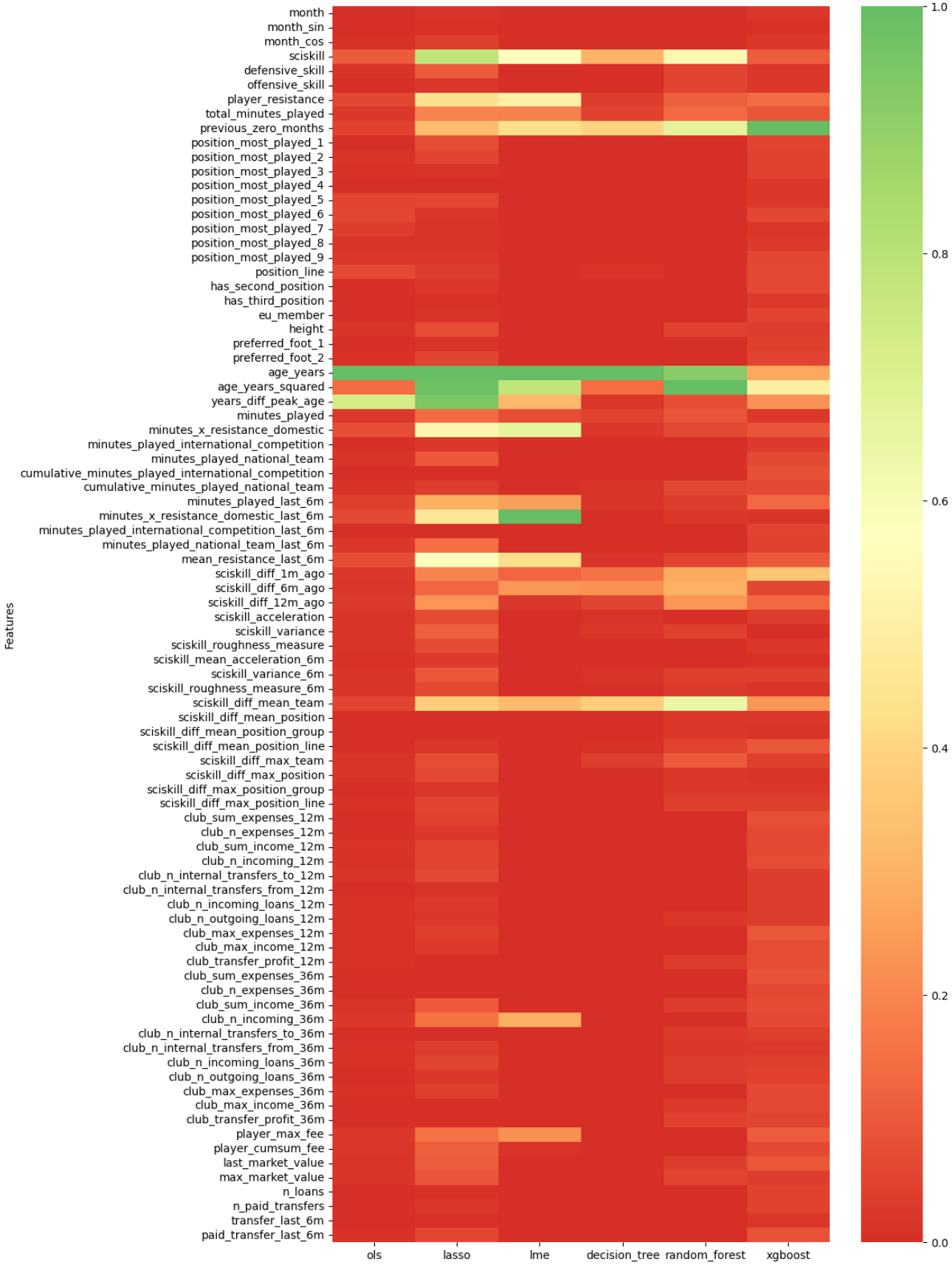


Figure 7. The feature importances of the linear and tree-based models in the prediction problem of player quality. Min–max scaling has been applied to the feature importances.

To summarize, the models indicate that the player age, player quality, the difference in quality between player and team, and the number of months since the last game are the most important features to predict future player quality. In this way, the feature importance can be used to identify what factors are important for the development of professional football players and to test whether models behave as expected.

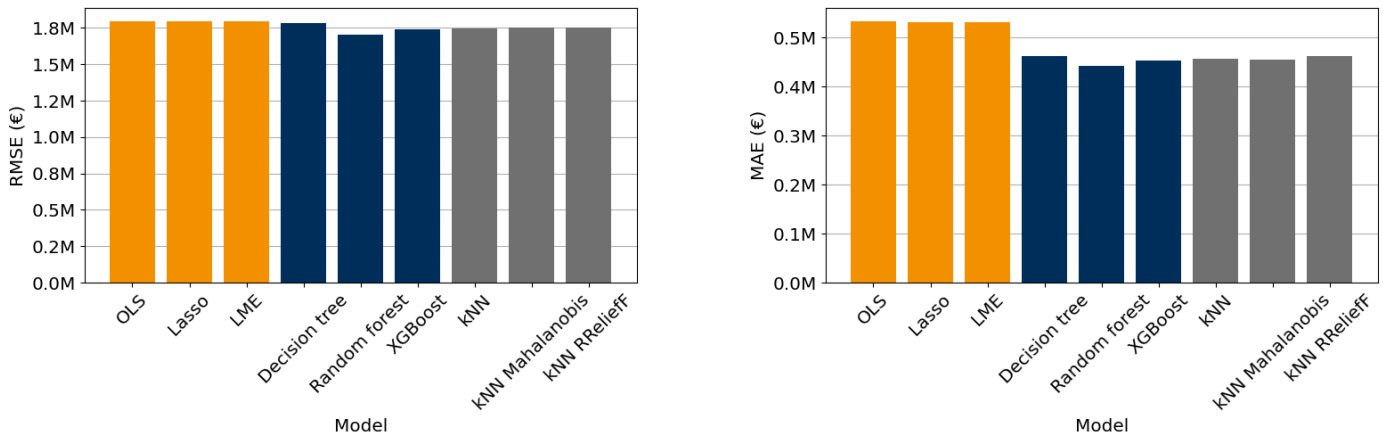
4.1.5. Predictive Accuracy

The results show that the tree-based models tend to attain the lowest loss values when predicting the development of a player's quality in this specific prediction problem. The XGBoost model seems to provide the most accurate predictions for the general population of players in the data set. Moreover, the XGBoost model attains low loss values on young players, high-quality players, and players with either a large decrease or increase in football quality in this data set. The random forest model attains similar performances on the young players and most of the interesting subgroups of players. Overall, the XGBoost model seems to have the lowest loss values on the data set for predicting player quality, followed by the random forest model.

4.2. Prediction Problem on Player Value

4.2.1. General Population of Players

The loss values for predicting player value development in the prediction problem of the monetary player value are given in Figure 8. The results show that the tree-based and kNN-based models perform relatively well in this prediction problem. Although the differences in RMSE are less obvious than the differences in the MAE, the results show that the random forest model attains the lowest loss values for both the RMSE and MAE, while the XGBoost and kNN-based models attain only slightly higher loss values.



(**a**) (**b**)

Figure 8. The test loss values for the different models on the general population of players in the player value prediction task for two different loss functions: (**a**) RMSE and (**b**) MAE.

The two loss functions show differences in performance as the linear models predict relatively worse with respect to the MAE than in terms of the RMSE compared to the other models. It can be reasoned that the linear models have relatively few large errors, which corresponds to the players that are harder to predict. On the contrary, the linear models predicted worse for players who are easier to predict. This means that the linear models predicted well for the players whose development is less predictable, but relatively badly for the players who are relatively easy to predict.

Additionally, the kNN models, which have time series information as features, predict better than the linear models. These results show that the time series of the player performance indicators contains important information for the development of the transfer value and that nonlinearities and interactions are involved.

4.2.2. Predictions per Age

The RMSE values on the test set for each age are given in Figure 9. The results show that the loss values of the models are similar at all ages and no clear distinction can be made between the quality of the models based on this.

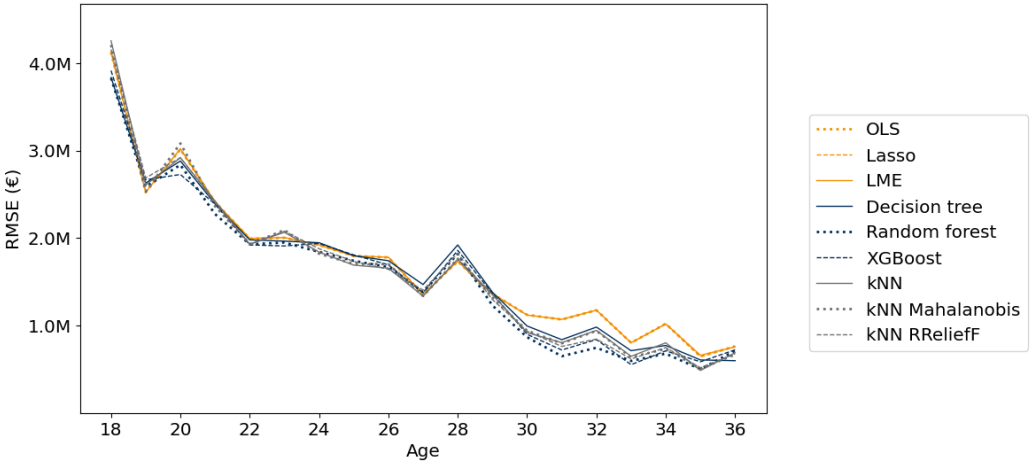


Figure 9. The RMSE values for each age for the different models in the player value prediction problem.

Moreover, the linear models perform worse on older players. This suggests that the patterns in the development of old players involve interaction effects or nonlinearities that were not captured by the linear models.

Generally, the models tend to predict the future development of player value worse for young players and better for older players. The development of a young player is less predictable, similarly as found for the prediction of player quality in Section 4.1. This is due to the fact that the KPIs for young players are more volatile. On the other hand, the player values of older players are smaller and more predictable because older players tend to decrease in value. In this way, it can be explained why the models have large loss values for young players and small loss values for older players.

4.2.3. Prediction on Important Player Groups

For the development of the transfer value, the four different groups of players that have been identified as important for the application of the models are high-quality players, high-value players, players with a large decrease in transfer value, and players with a large improvement in transfer value. Based on domain knowledge, high-quality players were defined as players with a SciSkill of more than 100, while high-value players are defined as players with an ETV of at least 10 M EUR. An increase or decrease of more than EUR 2.5 M was defined as large. The RMSE for these groups of players is visualized in Figure 10. Similar to the prediction problem for player quality, the errors in the predictions are larger for these interesting groups of players.

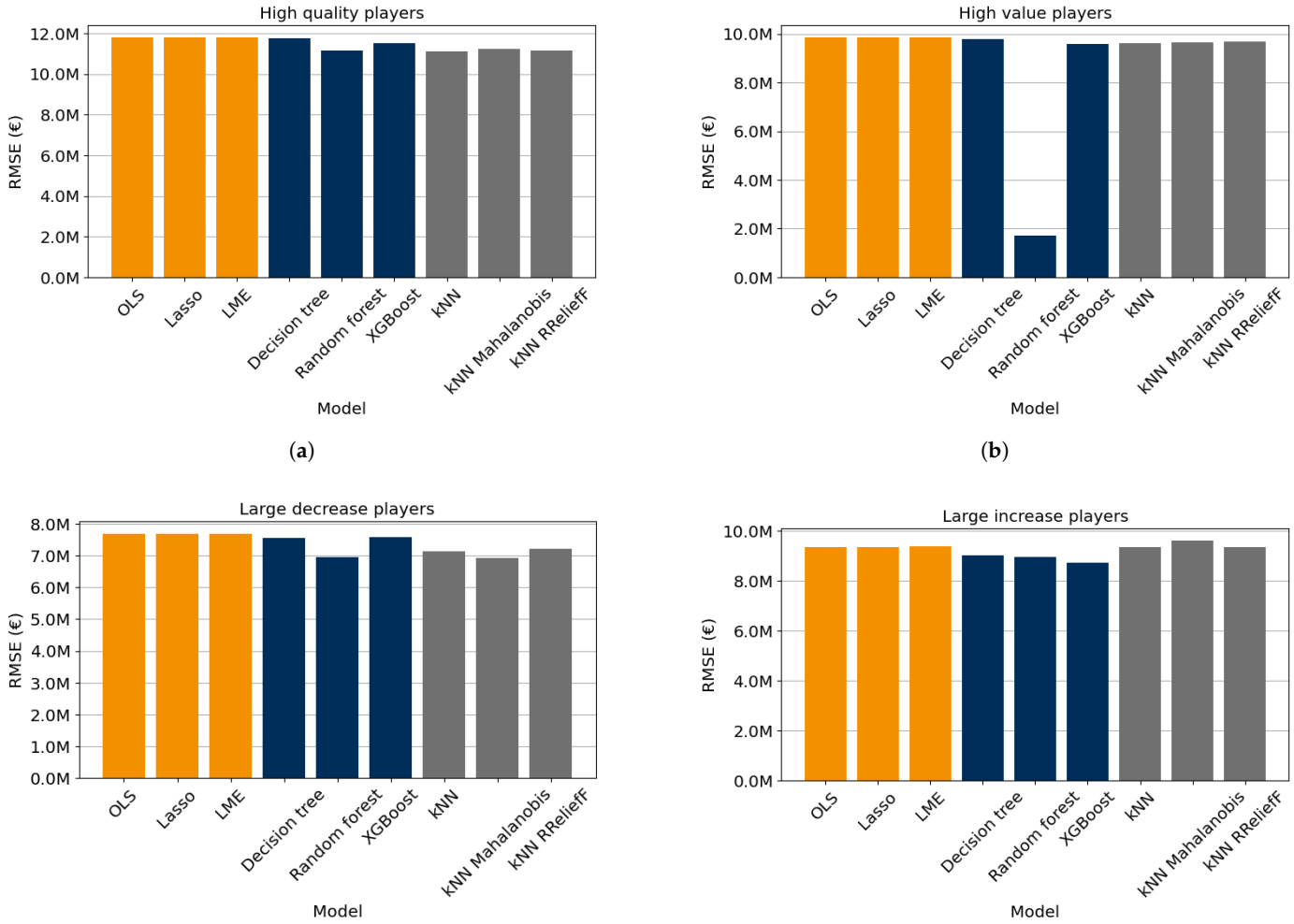


Figure 10. The RMSE loss per model in the player value prediction task for various player subsets: (**a**) players with a SciSkill of at least 100, (**b**) players with an ETV of at least EUR 10 M, (**c**) players with a value decrease of at least EUR 2.5 M, and (**d**) players with a value increase of at least EUR 2.5 M.

The results in Figure 10a,c show that the random forest model and kNN-based models attain the lowest loss values for the players of high quality and the players with a large decrease in value. The low loss values of the kNN-based models indicate that the time series features, described in Table S6 in the Supplementary Materials, contain most of the important information to predict the development of these subsets of players. The good performance of the kNN also indicates that a local method is suitable for the prediction of these types of players. On the other hand, Figure 10d shows that, for players with a large increase, the kNN models attain worse loss RMSE values compared to the tree-based models. Similarly, the tree-based models also outperform the linear model on the subset of players with a large increase. These differences suggest that a large increase in player value is influenced by many different features and contains nonlinear patterns or interaction effects.

Figure 10b shows that the random forest model has a distinctly lower RMSE on the players with a high player value than the other models. Players often attain high transfer values for a short period in their careers when they show peak performance and are young. Consequently, players with high transfer values can sometimes be expected to decrease in value shortly after obtaining high transfer values. The random forest model seems to be the only model that captures this pattern, as will be indicated by the high feature importance

of the current transfer value in Section 4.2.4. This explains the lower loss values of the random forest model on the high-value players.

4.2.4. Feature Importances

The feature importances for the development of the player transfer value after min–max scaling are shown in Figure 11. The results indicate that the most important features for predicting player value development are the features describing the most recent transfer value and the developments within the last 6 and 12 months of the player's transfer value. Especially the random forest model assigns a large value to the player value at the time of prediction ('etv'), which explains its good performance on high-value players in Figure 10b. Figure 11 also shows that additional time series information about the player quality ('sciskill_diff_6m_ago') provides information about the development of the player transfer value. This implies that the time series information is most important in predicting the future development, which is in line with the relatively good predictive performance of the kNN models that are based on time series information.

Moreover, it is found that the month of the year, which indicates whether it is January or July, is an important factor. Transfers in the winter transfer period are often driven by a necessity of a specific type of player at that moment, while transfers to improve the squad in the long term are more frequent in the summer transfer period. This changes the transfer fees, which is reflected in the ETV values. Additionally, players with only 6 months of contract left are able to negotiate a free transfer with a club, which decreases the transfer values of these players. As this commonly happens in the winter and the Estimated Transfer Value model contains the indicator whether a player has less than 6 months on his contract, the development of players' transfer value is different depending on the time of the year. Therefore, the fact that the month of the year is an important factor is in line with domain knowledge.

4.3. Predictive Accuracy

The results showed that, with this feature set, the random forest, XGBoost, and kNN-based models attained low loss values on the general population of football players. No clear distinction was found between the RMSE values for players of different ages. The random forest and kNN models attained low RMSE scores on the subsets of high-quality players and players with a large decrease in monetary value. The tree-based models had the lowest loss values on players with a large increase in value, with the XGBoost model attaining the lowest loss. The random forest model attained a lower loss value on the subset of high-value players. Taking this into account, the random forest model has the best predictive accuracy on the data set of this prediction problem, while the XGBoost model and the kNN-based models also attain low loss values.

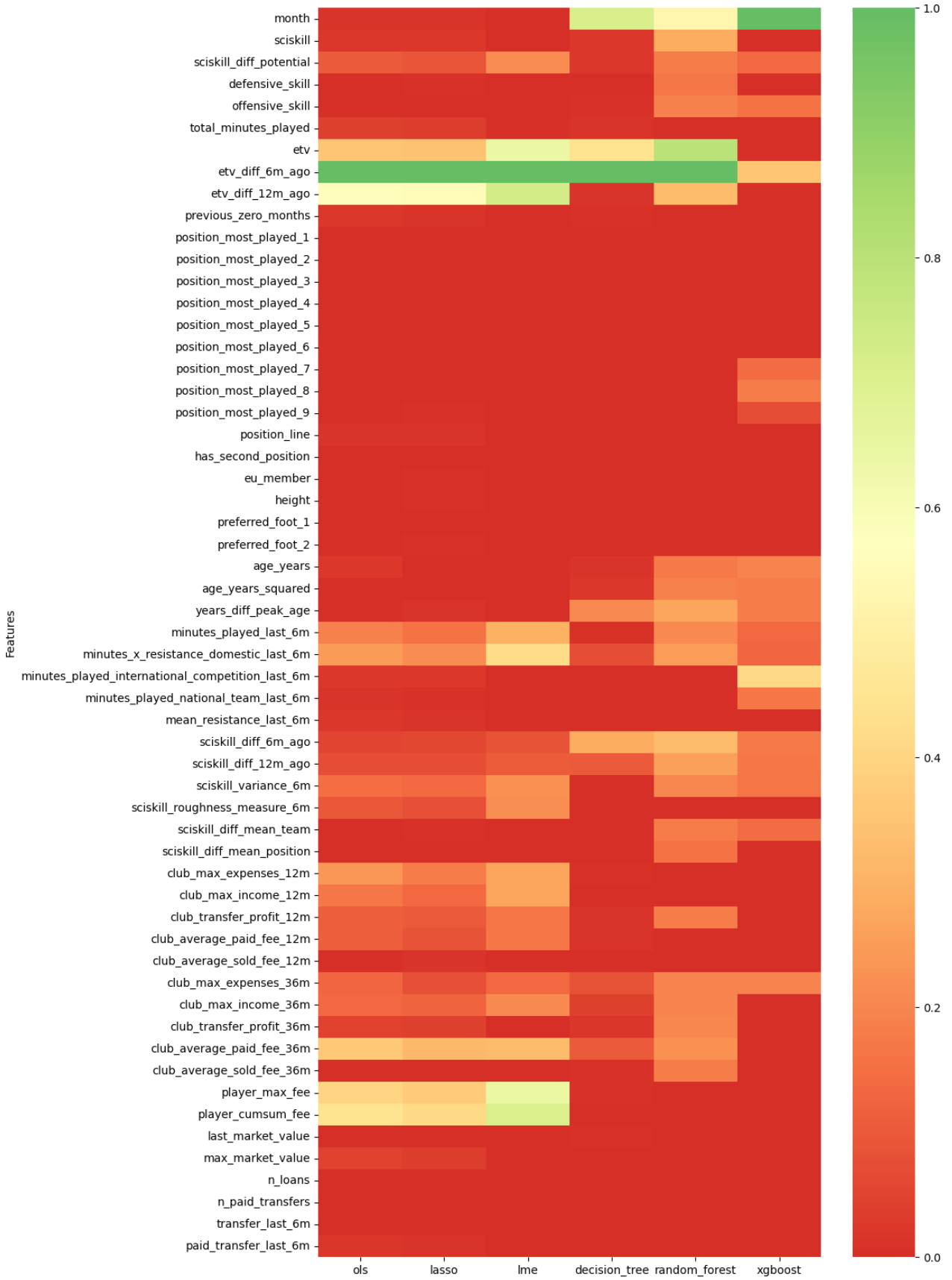


Figure 11. The feature importances of the linear and tree-based models in the prediction problem of player value. Min–max scaling has been applied to the feature importances.

5. Conclusions

This paper aims to find the most suitable supervised learning model for forecasting the development of player performance indicators one year ahead in a practitioner-oriented way. Through the literature, it was found that linear regression, random forest, and kNN-based models are favorable based on their uncertainty quantification methods.

Two prediction problems were considered to study the predictive performance of linear, tree-based, and kNN-based models. The results show that the XGBoost model attained the lowest loss values for predicting the development of the player quality with this data set, while the random forest also had relatively low loss values. For the prediction of the player value, it was found that the random forest had the lowest loss values. The XGBoost and kNN-based models also had relatively accurate predictions according to some of the studied losses.

Because the random forest had a good predictive accuracy on these data sets and it provides methods for uncertainty quantification, it seems to be a suitable model for predicting the development of player quality and value in football. Nonetheless, it is important to note that random forests can be sensitive to data imbalance, which regularly occurs in elite football. Additionally, given the large number of features needed to predict the development of football players, random forests may risk overfitting if not properly tuned and validated. These limitations should be considered when applying the model in practice, and appropriate techniques, like resampling or feature selection, may be necessary to mitigate them.

Taking these considerations into account, it is concluded that the random forest is the most suitable explainable machine learning model to predict the development of player performance indicators one year ahead for the above prediction problems.

6. Implications and Future Directions

By addressing the two prediction problems, two random forest models have been obtained that can predict the development of both the quality and the monetary value of football players. These random forest models can be used to aid in transfer decisions combined with a critical view of a domain expert. Suppose the manager of a football club has a long-term interest in a football player. If the model predicting the development of player quality indicates that a player will grow in quality, it means that a player is more interesting to a football club in the long term. If the prediction of the transfer value indicates an increase in the transfer value, it might be better to buy the player sooner rather than later. The two models can also be combined to give extra insights. Suppose a manager currently has a veteran player who is predicted to start decreasing in quality, and he can buy a young player who is predicted to develop into a first-team player within the next year. Assume that the transfer value of the young player is predicted to only slightly increase. In this case, it might be better to buy the young player one year later and then sell the veteran player after a critical evaluation of the possible influences of noise and expected transfer market development by the manager. This would be beneficial because the young player will only have a slightly higher transfer fee, and the veteran player will have a better quality in the meantime. These examples illustrate the added value of our models to predict player development for the improvement of data-informed decision-making.

The models from this paper can also be used to complement existing methods in the literature. Pantuso and Hvattum [19] introduced a method to optimize transfer decisions based on indicators that describe a player's quality and transfer value and their future values. Their methodology needs to know the 'future' values of the player quality and their transfer values. To solve this, they consider transfer situations from over a year prior so that the values of one year later are already known. Consequently, their model can only

be applied to historical situations. By using our models to predict the future values in quality and transfer value, it is possible to obtain these predictions and optimize transfer decisions with a current-time application. This makes it possible to advise data-driven transfer decisions to optimize the squad in a real-life transfer period.

When applying the models from this research at football clubs, possible shortcomings should be taken into consideration. The models will most likely capture patterns of development for players that are typical for the population on which they are trained. They will perform less on nontypical players, which makes it harder to predict, e.g., late bloomers. Young or injury-prone players can be considered vulnerable players who are nontypical for the general population, and the predictions of the models on such players should thus be interpreted with extra care. This highlights the value of explainable modeling, which can be complemented with interpretation techniques like the use of Shapley values by Lundberg and Lee [54].

Another advantage of the explainable models in this research is that they provide methods to gain insight into important factors for prediction. The models for the prediction problems show that this can be achieved via feature importance, similar as carried out by Baouan et al. [48]. Because our models predict the difference between the current performance indicators and those of one year later, our research can also show the influences of the indicator itself. Our results show that the value of the indicator and the historical values of the indicators are the most important features, which adds to the knowledge obtained by Baouan et al. [48] that describes which features are important. Additionally, time-dependent variables like the period in the year and the months without games have been found to be important features in forecasting player development. This indicates that the time series of the indicators themselves contains important information on the development of football players, although our findings also suggest that contextual information gives improved predictive performance of the models.

In short, this paper studied explainable supervised models to predict player development via performance indicators. Two prediction problems were studied in which explainable models were trained to predict the development of both player quality and player value. It was found that the random forest model is the most suitable model for forecasting player development, because of the accurate predictions for both performance indicators, combined with the method for uncertainty quantification arising from the bagging procedure.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/app15168916/s1, Table S1: Dependent variable of the player quality prediction problem. Table S2: Feature descriptions of the player quality prediction problem. Table S3: Feature descriptions of the kNN models in the player quality prediction problem. Table S4: Dependent variable of the player value prediction problem. Table S5: Feature descriptions of the player value prediction problem. Table S6: Feature descriptions of the kNN models in the player value prediction problem.

Author Contributions: Conceptualization, K.v.A., F.G.-S., and J.S.; methodology, K.v.A.; software, K.v.A.; validation, K.v.A.; visualization, K.v.A.; investigation, K.v.A.; data curation, K.v.A.; writing—original draft preparation, K.v.A.; writing—review and editing, K.v.A., F.G.-S., and J.S.; supervision, F.G.-S. and J.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Restrictions apply to the availability of this data. Data was obtained from SciSports.

Acknowledgments: The authors thank Geurt Jongbloed for his comments on a draft version of this paper and the company SciSports for the data, computational resources, and insight into the needs of practitioners. The authors would also like to thank the two anonymous referees who helped to improve the paper with their suggestions.

Conflicts of Interest: Author Floris Goes-Smit was employed by the company SciSports. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Yang, Y.; Koenigstorfer, J.; Pawlowski, T. Predicting transfer fees in professional European football before and during COVID-19 using machine learning. *Eur. Sport Manag. Q.* **2024**, *24*, 603–623. [CrossRef]
2. McHale, I.G.; Holmes, B. Estimating transfer fees of professional footballers using advanced performance metrics and machine learning. *Eur. J. Oper. Res.* **2022**, *306*, 389–399. [CrossRef]
3. Rein, R.; Memmert, D. Big data and tactical analysis in elite soccer: Future challenges and opportunities for sport science. *SpringerPlus* **2016**, *5*, 1410. [CrossRef]
4. Herold, M.; Goes, F.; Nopp, S.; Bauer, P.; Thompson, C.; Meyer, T. Machine learning in men's professional football: Current applications and future directions for improving attacking play. *Int. J. Sports Sci. Coach.* **2019**, *14*, 798–817. [CrossRef]
5. Green, S. Assessing the Performance of Premier League Goalscorers. 2012. Available online: https://www.statsperform.com/resource/assessing-the-performance-of-premier-league-goalscorers/ (accessed on 27 November 2023).
6. Eggels, H.; van Elk, R.; Pechenizkiy, M. Explaining soccer match outcomes with goal scoring opportunities predictive analytics, 2016. In Proceedings of the Workshop on Machine Learning and Data Mining for Sports Analytics 2016, Riva del Garda, Italy, 19 September 2016.
7. Anzer, G.; Bauer, P. A Goal Scoring Probability Model for Shots Based on Synchronized Positional and Event Data in Football (Soccer). *Front. Sports Act. Living* **2021**, *3*, 624475. [CrossRef]
8. Mead, J.; O'Hare, A.; McMenemy, P. Expected goals in football: Improving model performance and demonstrating value. *PLoS ONE* **2023**, *18*, e0282295. [CrossRef] [PubMed]
9. Hvattum, L.M.; Gelade, G.A. Comparing bottom-up and top-down ratings for individual soccer players. *Int. J. Comput. Sci. Sport* **2021**, *20*, 23–42. [CrossRef]
10. Rudd, S. A Framework for Tactical Analysis and Individual Offensive Production Assessment in Soccer Using Markov Chains, 2011. In Proceedings of the New England Symposium on Statistics in Sports, Cambridge, MA, USA, 24 September 2011.
11. Van Roy, M.; Robberechts, P.; Decroos, T.; Davis, J. Valuing on-the-ball actions in soccer: A critical comparison of xT and VAEP, 2020. In Proceedings of the 2020 AAAI Workshop on AI in Team Sports, New York, NY, USA, 8 February 2020.
12. Van Arem, K.; Bruinsma, M. Extended xThreat: An explainable quality assessment method for actions in football using game context, 2024. In Proceedings of the 15th International Conference on the Engineering of Sport (ISEA 2024), Loughborough, UK, 8–11 July 2024. [CrossRef]
13. Decroos, T.; Bransen, L.; Davis, J. Actions Speak Louder Than Goals: Valuing Player Actions in Soccer, 2019. In Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, Anchorage, AK, USA, 4–8 August 2019. [CrossRef]
14. Decroos, T.; Davis, J. Interpretable prediction of goals in soccer, 2020. In Proceedings of the 2020 AAAI Workshop on AI in Team Sports, Hilton Midtown, New York, NY, USA, 8 February 2020.
15. Van Haaren, J. Why Would I Trust Your Numbers? On the Explainability of Expected Values in Soccer, 2021. In Proceedings of the Workshop on Artificial Intelligence for Sports Analytics (AISA 2021), Virtual, 17 August 2021. [CrossRef]
16. Mendes-Neves, T.; Meireles, L.; Mendes-Moreira, J. Valuing Players Over Time. *arXiv* **2022**, arXiv:2209.03882. [CrossRef]
17. Sæbø, O.D.; Hvattum, L.M. Evaluating the efficiency of the association football transfer market using regression based player ratings, 2015. In Proceedings of the 28th Norsk Informatikkonferanse (NIK 2015), Høgskolen i Ålesund, Ålesund, Norway, 23–25 November 2015.
18. Kharrat, T.; McHale, I.G.; López Peña, J. Plus-minus player ratings for soccer. *Eur. J. Oper. Res.* **2020**, *283*, 726–736. [CrossRef]
19. Pantuso, G.; Hvattum, L.M. Maximizing performance with an eye on the finances: A chance-constrained model for football transfer market decisions. *TOP* **2021**, *29*, 583–611. [CrossRef]
20. Hvattum, L.M. Offensive and Defensive Plus-Minus Player Ratings in Soccer. *Appl. Sci.* **2020**, *20*, 7345. [CrossRef]

21. De Bacco, C.; Wang, Y.; Blei, D.M. A causality-inspired adjusted plus-minus model for player evaluation in team sports, 2024. In Proceedings of the Third Conference on Causal Learning and Reasoning (CLeaR 2024), Los Angeles, CA, USA, 1–3 April 2024.
22. Wolf, S.; Schmitt, M.; Schuller, B. A football player rating system. *J. Sports Anal.* **2020**, *6*, 243–257. [CrossRef]
23. SciSports. SciSkill Index—Why and Hows. 2020. Available online: https://www.scisports.com/sciskill-index-why-and-how/# (accessed on 27 November 2023).
24. Franceschi, M.; Brocard, J.F.; Follert, F.; Gouguet, J.J. Determinants of football players' valuations: A systematic review. *J. Econ. Surv.* **2024**, *38*, 577–600. [CrossRef]
25. SciSports. Player Valuation Model. 2024. Available online: https://www.scisports.com/player-valuation-model/ (accessed on 5 December 2023).
26. Davis, J.; Bransen, L.; Devos, L.; Jaspers, A.; Meert, W.; Robberechts, P.; Van Haaren, J.; Van Roy, M. Methodology and evaluation in sports analytics: Challenges, approaches, and lessons learned. *Mach. Learn.* **2024**, *113*, 6977–7010. [CrossRef]
27. Hastie, T.; Tibshirani, R.; Friedman, J.H.; Friedman, J.H. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*; Springer: Boston, MA, USA, 2009; Volume 2.
28. Tibshirani, R. Regression Shrinkage and Selection Via the Lasso. *J. R. Stat. Soc. Ser. B Stat. Method.* **1996**, *58*, 267–288. [CrossRef]
29. Lindstrom, M.J.; Bates, D.M. Newton-Raphson and EM Algorithms for Linear Mixed-Effects Models for Repeated-Measures Data. *J. Am. Stat. Assoc.* **1988**, *83*, 1014–1022. [CrossRef]
30. Chen, T.; Guestrin, C. XGBoost: A Scalable Tree Boosting System, 2016. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KKD '16), San Francisco, CA, USA, 13–17 August 2016. [CrossRef]
31. Dudani, S.A. The Distance-Weighted k-Nearest-Neighbor Rule. *IEEE Trans. Syst. Man Cybern.* **1976**, *SMC-6*, 325–327. [CrossRef]
32. Robnik-Sikonja, M.; Kononenko, I. An adaptation of Relief for attribute estimation in regression, 1997. In Proceedings of the Fourteenth International Conference on Machine Learning, San Francisco, CA, USA, 8–12 July 1997.
33. Molnar, C. *Interpretable Machine Learning: A Guide for Making Black Box Models Explainable*; Version 2019-02-21; Springer: Boston, MA, USA, 2019.
34. Neter, J.; Kutner, M.H.; Nachtsheim, C.J.; Li, W. *Applied Linear Statistical Models*, 5th ed.; McGraw-Hill/Irwin: Boston, MA, USA, 2004.
35. Wager, S.; Hastie, T.; Efron, B. Confidence Intervals for Random Forests: The Jackknife and the Infinitesimal Jackknife. *J. Mach. Learn. Res.* **2014**, *15*, 1625–1651. [CrossRef] [PubMed]
36. Arntzen, H.; Hvattum, L.M. Predicting match outcomes in association football using team ratings and player ratings. *Stat. Model.* **2021**, *21*, 449–470. [CrossRef]
37. Goes, F.; Kempe, M.; Meerhoff, L.A.; Lemmink, K.A.P.M. Not evey pass can be an assist: A data-driven model to measure pass effectiveness in professional soccer matches. *Big Data* **2018**, *7*, 57–70. [CrossRef] [PubMed]
38. Poli, R.; Besson, R.; Ravenel, L. Econometric Approach to Assessing the Transfer Fees and Values of Professional Football Players. *Economies* **2022**, *10*, 4. [CrossRef]
39. Herm, S.; Callsen-Bracker, H.M.; Kreis, H. When the crowd evaluates soccer players' market values: Accuracy and evaluation attributes of an online community. *Sport Manag. Rev.* **2014**, *17*, 484–492. [CrossRef]
40. Al-Asadi, M.A.; Tasdemır, S. Predict the Value of Football Players Using FIFA Video Game Data and Machine Learning Techniques. *IEEE Access* **2022**, *10*, 22631–22645. [CrossRef]
41. Steve Arrul, V.; Subramanian, P.; Mafas, R. Predicting the Football Players' Market Value Using Neural Network Model: A Data-Driven Approach, 2022. In Proceedings of the 2022 IEEE International Conference on Distributed Computing and Electrical Circuits and Electronics (ICDCECE), Ballari, India, 23–24 April 2022. [CrossRef]
42. Behravan, I.; Razavi, S.M. A novel machine learning method for estimating football players' value in the transfer market. *Methodol. Appl.* **2021**, *25*, 2499–2511. [CrossRef]
43. Apostolou, K.; Tjortjis, C. Sports Analytics algorithms for performance predictions, 2019. In Proceedings of the 10th International Conference on Information, Intelligence, Systems and Applications (IISA 2019), Patras, Greece, 15–17 July 2019. [CrossRef]
44. Pantzalis, V.C.; Tjortjis, C. Sports Analytics for Football League Table and Player Performance Prediction, 2020. In Proceedings of the 11th International Conference on Information, Intelligence, Systems and Applications (IISA 2020), Piraeus, Greece, 15–17 July 2020. [CrossRef]
45. Giannakoulas, N.; Papageorgiou, G.; Tjortjis, C. Forecasting Goal Performance for Top League Football Players: A Comparative Study. In *Proceedings of the Artificial Intelligence Applications and Innovations*; Maglogiannis, I., Iliadis, L., MacIntyre, J., Dominguez, M., Eds.; Springer: Cham, Switzerland, 2023; Volume 676, pp. 304–315. [CrossRef]
46. Markopoulou, C.; Papageorgiou, G.; Tjortjis, C. Diverse Machine Learning for Forecasting Goal-Scoring Likelihood in Elite Football Leagues. *Mach. Learn. Knowl. Extr.* **2024**, *6*, 1762–1781. [CrossRef]
47. Barron, D.; Ball, G.; Robins, M.; Sunderland, C. Artificial neural networks and player recruitment in professional soccer. *PLoS ONE* **2018**, *13*, e0205818. [CrossRef]

48. Baouan, A.; Bismuth, E.; Bohbot, A.; Coustou, S.; Lacome, M.; Rosenbaum, M. What should clubs monitor to predict future value of football players. *arXiv* **2022**, arXiv:2212.11041. [CrossRef]
49. Seabold, S.; Perktold, J. Statsmodels: Econometric and statistical modeling with python, 2010. In Proceedings of the 9th Python in Science Conference, Austin, TX, USA, 28 June–3 July 2010.
50. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
51. Douze, M.; Guzhva, A.; Deng, C.; Johnson, J.; Szilvasy, G.; Mazaré, P.E.; Lomeli, M.; Hosseini, L.; Jégou, H. The Faiss library. *arXiv* **2024**, arXiv:2401.08281. [CrossRef]
52. Slifka, M.K.; Whitton, J.L. Theoretical and Empirical Analysis of ReliefF and RReliefF. *Mach. Learn.* **2003**, *53*, 23–69. [CrossRef]
53. Head, T.; Kumar, M.; Nahrstaedt, H.; Louppe, G.; Shcherbatyi, I. scikit-optimize/scikit-optimize, 2021. Available online: https://zenodo.org/records/5565057 (accessed on 11 December 2023).
54. Lundberg, S.; Lee, S.I. A Unified Approach to Interpreting Model Predictions. *arXiv* **2017**, arXiv:1705.07874. [CrossRef]