# Predicting Player Performance in EA SPORTS FC 25: A Comparative Analysis of Linear Regression and Random Forest Regression Using In-Game Attributes

Kattareeya Prompreing[1,*]

[1]Faculty of Business Administration and Liberal Arts, Rajamangala University of Technology Lanna Tak, Tak 63000, Thailand

ABSTRACT

This study presents a comparative analysis of Linear Regression and Random Forest Regression models to predict player performance in EA SPORTS FC 25 using in-game attributes. The primary objective is to evaluate these models in terms of their accuracy and effectiveness in predicting player ratings based on key attributes like Ball Control, Dribbling, Defense, and Reactions. The dataset comprises 17,737 player records with multiple performance indicators, preprocessed to ensure quality data for modeling. The research process involves data exploration, model development, and evaluation using metrics such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and the Coefficient of Determination ($R^2$). Results indicate that the Random Forest model outperforms the Linear Regression model, achieving a lower MAE and RMSE, and a higher $R^2$ score, highlighting its ability to capture complex, non-linear relationships among player attributes. The study's findings underscore the significance of ensemble models in gaming analytics and provide insights for gamers and developers to optimize gameplay strategies and improve game mechanics. Limitations include data constraints, and recommendations for future research suggest incorporating more diverse player data and exploring advanced algorithms.

**Keywords** Player Performance, EA SPORTS FC 25, Linear Regression, Random Forest Regression, Gaming Analytics

## Introduction

EA SPORTS FC 25 emerges as a pivotal installment in the football simulation genre, reflecting the broader significance of video games within the gaming industry. The video game sector has established itself as a dominant force in global entertainment, surpassing both music and cinema in revenue generation [1]. This economic prominence underscores the cultural shift where video games, including flagship titles like EA SPORTS FC 25, have become primary entertainment mediums for diverse demographics ranging from children to adults. The success of EA SPORTS FC 25 is emblematic of this trend, demonstrating how advanced game design and immersive experiences contribute to sustained player engagement and industry growth.

The social and cultural dimensions of EA SPORTS FC 25 further highlight its importance in the gaming landscape. Video games serve as platforms for social interaction, fostering connections among players through cooperative gameplay and shared experiences [2]. EA SPORTS FC 25 leverages these dynamics by offering robust online multiplayer features that enhance community building and player collaboration [3]. Additionally, the game's ability to influence social behaviors and cultural norms aligns with the broader recognition of video games

Submitted 28 December 2024
Accepted 20 January 2025
Published 26 February 2025

Corresponding author
Kattareeya Prompreing,
katt.rmutl@gmail.com

Additional Information and Declarations can be found on page 97

DOI: 10.47738/ijrm.v2i1.22



**How to cite this article:** K. Prompreing, "Predicting Player Performance in EA SPORTS FC 25: A Comparative Analysis of Linear Regression and Random Forest Regression Using In-Game Attributes," *Int. J. Res. Metav.*, vol. 2, no. 1, pp. 78-101, 2025.

as significant cultural artifacts [4], [5]. By integrating complex narratives and realistic simulations, EA SPORTS FC 25 not only entertains but also contributes to the ongoing discourse on the cultural and psychological impacts of video gaming [6], [7].

The importance of player attributes in gameplay and team building is a critical factor in the success of EA SPORTS FC 25, mirroring trends observed in both traditional sports and esports. Player attributes encompass a range of characteristics, including skills, cognitive abilities, and social dynamics, all of which significantly influence individual performance and overall team effectiveness. Research [8] introduce a 'nested matching' approach that emphasizes aligning player attributes with team strategies to enhance engagement and retention in multiplayer video games. This methodology underscores how understanding individual player skills can lead to more effective team compositions, thereby improving the overall gaming experience. Similarly, [9] highlight the impact of social player experiences, suggesting that interactions influenced by audience engagement can affect gameplay performance, thereby reinforcing the relational aspect of player attributes.

Furthermore, assessing individual performance metrics plays a pivotal role in effective team selection within EA SPORTS FC 25. Research [10] argues that traditional statistics may only partially capture a player's contribution to team success, advocating for a more nuanced understanding of individual performance. This perspective is supported by [11], who demonstrate how analytics can quantify player performance and inform strategic team decisions in basketball, a concept directly applicable to football simulations. Additionally, cognitive abilities and personality traits are integral to gameplay success, as evidenced by [12], who found that individual differences in these areas can predict performance in car-soccer video games. Study [13] further emphasize the importance of tactical awareness and decision-making skills, indicating that cognitive and psychological readiness significantly impact game performance. The integration of advanced metrics, as discussed by [14], allows for a comprehensive evaluation of player contributions, moving beyond traditional metrics to consider contextual factors that influence team success.

Predictive modeling using both linear and ensemble learning methods has proven effective across various domains, offering insights into complex data-driven challenges. Studies such as the application of Random Forest algorithms on stock price movements in the Vietnamese banking sector [15] and the comparative analysis of logistic regression and Random Forest for predicting e-commerce customer behavior [16] illustrate the ability of ensemble methods to enhance predictive accuracy. Similarly, ensemble learning techniques have been leveraged for purchase prediction in digital promotions [17] and campaign ROI prediction using decision trees and Random Forest [18], showcasing their strength in improving marketing-related outcomes. The study of predictive modeling of blockchain stability through machine learning [19] and anomaly detection in digital currency trading using clustering and density-based approaches [20] further underscores the applicability of these methods in diverse, high-dimensional data contexts. Additionally, predictive modeling in virtual environments, such as Roblox stock price trends using machine learning [21] and the determinants of virtual property prices in Decentraland [22], highlights the utility of data-driven approaches in analyzing complex, evolving digital markets. These works collectively emphasize the versatility and robustness of predictive modeling frameworks, informing the development of optimized models for player performance prediction in gaming contexts.

Predicting player performance plays a pivotal role in enhancing gaming strategies within EA SPORTS FC 25, enabling players and teams to make informed decisions based on statistical analyses and predictive modeling. Advanced analytics, such as exploratory data analysis (EDA), facilitate the identification of patterns and trends within in-game data, which are essential for optimizing strategic choices. Research [23] demonstrate that applying EDA techniques to game-specific data allows players to discern their strengths and weaknesses, thereby tailoring their gameplay strategies to maximize effectiveness. This data-driven approach not only elevates individual performance but also synergizes team dynamics by aligning strategies with collective insights, leading to more cohesive and effective gameplay.

Moreover, the integration of reinforcement learning (RL) into player performance prediction has shown significant promise in adaptive strategy development. Study [24] explore how reinforcement learning can incorporate strategy diversity, enabling players to adjust their tactics in response to the evolving game environment. This adaptability is crucial in dynamic gaming scenarios, where understanding and anticipating opponent behavior can provide a substantial competitive advantage. Additionally, recognizing different player motivations and types, as highlighted by [25], further refines performance predictions by allowing for the customization of strategies that resonate with diverse player profiles. Addressing technical challenges such as latency, [26] categorize player actions based on their precision and impact, ensuring that strategic decisions are both timely and effective. Collectively, these methodologies underscore the multifaceted importance of predicting player performance in developing robust and adaptive gaming strategies.

Predicting player performance in EA SPORTS FC 25 presents significant challenges due to the multitude of influencing attributes that must be considered. Player attributes encompass a wide range of factors, including technical skills, tactical awareness, cognitive abilities, and psychological traits, all of which interact in complex ways to determine individual and team performance. Research [23] emphasize that integrating diverse in-game data through EDA is essential for identifying underlying patterns and correlations among these attributes. However, the high dimensionality and interdependence of these variables complicate the modeling process, making it difficult to isolate the most influential factors accurately. Additionally, the dynamic nature of gameplay, where player performance can fluctuate based on real-time decisions and interactions, further exacerbates the complexity of prediction models [24].

Another significant challenge lies in the necessity for accurate and reliable predictive models that can effectively handle the complexity of in-game attributes. Developing such models requires robust statistical and machine learning techniques capable of capturing non-linear relationships and interactions between variables. Study [25] highlight that the predictive accuracy of models like linear regression and random forest regression depends on their ability to accommodate the intricate dynamics of player performance data. Moreover, [26] discuss the importance of addressing issues such as overfitting and model generalization to ensure that predictions remain reliable across different game scenarios and player populations. Achieving this level of precision is critical for enhancing gaming strategies, as inaccurate predictions can lead to suboptimal decision-making and reduced competitive advantage.

The primary objective of this study is to develop predictive models utilizing Linear Regression and Random Forest Regression techniques to forecast player performance in EA SPORTS FC 25. By harnessing these statistical and machine learning methodologies, the research aims to analyze the impact of various in-game attributes on individual player outcomes. Study [23] highlight the significance of EDA in identifying key performance indicators, which serves as a foundational step in building accurate predictive models. Implementing Linear Regression allows for the examination of linear relationships between player attributes and performance metrics, providing a baseline for understanding how each attribute contributes to overall effectiveness. Conversely, Random Forest Regression offers a more complex, non-linear approach, capable of capturing intricate interactions among multiple attributes, thereby enhancing the robustness of performance predictions.

Furthermore, this study seeks to compare the effectiveness of Linear Regression and Random Forest Regression models in predicting player performance based on in-game attributes. By evaluating these models against standardized performance metrics such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and the Coefficient of Determination ($R^2$), the research aims to determine which approach offers superior predictive accuracy and reliability. Research [24] emphasize the importance of model selection in handling complex data structures, suggesting that ensemble methods like Random Forest may outperform traditional regression techniques in scenarios with high attribute interdependence. Through this comparative analysis, the study intends to provide actionable insights for players and team managers in EA SPORTS FC 25, enabling them to optimize team compositions and strategic decisions based on data-driven performance forecasts.

## Literature Review

### Player Performance Metrics in Gaming

In sports simulation games, measuring player performance is a multifaceted process that involves a combination of statistical analyses, machine learning techniques, and context-based evaluations. These methods collectively offer a comprehensive understanding of how individual players contribute to team success, both in virtual environments and real-life simulations. A significant component of this measurement is the use of advanced statistical metrics, which provide detailed insights into player capabilities. The EA Sports Player Performance Index, for example, exemplifies this approach by collaborating with professional leagues and academic institutions to develop a rating system that evaluates players based on key performance indicators such as goals, assists, and defensive actions [27]. This index reflects the way simulation games can mimic real-world performance metrics, enabling players to engage more deeply with the game by understanding the factors contributing to player impact.

In addition to these traditional metrics, machine learning and data analytics have transformed player performance assessments by introducing more dynamic and adaptable evaluations. Techniques like Deep Reinforcement Learning (DRL) have been applied to analyze player actions in sports such as ice hockey, providing insights into individual performance by accounting for the context of each play [28]. This method facilitates a more nuanced understanding of player actions, incorporating situational factors that affect decision-making and execution. Moreover, qualitative assessments add another layer of complexity, as they capture how factors like gaming experience and sports knowledge influence performance [29]. These insights underscore the need for a holistic approach to measuring player performance, one that combines quantitative metrics with a consideration of player behavior and context.

Overall rating is a crucial key performance indicator (KPI) in sports simulation games, offering a composite measure of a player's effectiveness and overall contribution to their team. By synthesizing a range of performance statistics into a single score, overall ratings enable quick assessments of both individual and team capabilities, which are essential for strategic decision-making. Traditional performance metrics, such as goals scored in soccer or points and rebounds in basketball, often serve as foundational elements in overall rating systems. These metrics are complemented by advanced measures like Player Efficiency Rating (PER) and Wins Above Replacement (WAR), which provide a more comprehensive view of a player’s impact [30]. These metrics facilitate player comparisons across different positions, capturing a holistic view of their contributions [31].

Machine learning further enhances the development of overall ratings by incorporating complex data patterns and context-specific variables. For instance, DRL techniques have been proposed to model player actions and their contributions to outcomes in sports like soccer and basketball, taking into account the state of the game and player roles within team dynamics [32]. This approach enables a more accurate representation of a player's effectiveness, as it considers not only raw statistics but also the contextual elements of gameplay. Additionally, qualitative aspects such as decision-making and leadership qualities are increasingly factored into overall ratings, recognizing their role in shaping performance [33]. For example, the context-aware cricket player performance metric (CAMP) captures the situational impact of player actions, underscoring the importance of context in evaluating overall contributions [34].

Overall ratings are not limited to individual assessments but are also instrumental in shaping team strategies. Coaches and managers rely on these ratings to make informed decisions about player selection, game plans, and training focus. The ability to quantify a player's overall impact allows for better alignment between player strengths and tactical needs, ultimately enhancing team cohesion and performance. This strategic application highlights the importance of overall ratings as a KPI in sports simulation games, as they directly influence both gameplay and team dynamics [35]. In summary, overall ratings integrate statistical metrics, machine learning, and qualitative assessments to provide a thorough evaluation of player performance, offering a powerful tool for both individual players and team management.

### Predictive Modeling in Gaming Analytics

The application of machine learning (ML) in gaming analytics has gained substantial traction as researchers and game developers seek to understand and predict player performance through sophisticated algorithms. ML models have shown remarkable potential in analyzing vast datasets, identifying patterns, and forecasting outcomes based on historical data. For instance, Liu et al. implemented a Random Forest model to predict player performance in eSports. This demonstrates that ML techniques can outperform traditional statistical methods by effectively capturing non-linear relationships between variables [36]. Similarly, Morales-García et al. applied EDA combined with supervised learning models to identify key performance indicators in League of Legends, one of the most popular multiplayer online games, highlighting the ability of ML to enhance the understanding of complex player interactions and gameplay strategies [23].

Furthermore, deep learning has been instrumental in pushing the boundaries of gaming analytics. Yanai et al. utilized Deep Reinforcement Learning (DRL) to model the decision-making process in soccer simulation games, providing insights into player effectiveness based on situational contexts, such as the positioning and actions of opponents [32]. This capability underscores the versatility of ML models in adapting to dynamic game environments, enabling the development of more accurate and context-aware predictive tools. Through these models, researchers can dissect the intricate factors contributing to player success, supporting the creation of training tools and in-game guidance systems that optimize player skills. The versatility of ML in handling diverse datasets makes it a powerful instrument in gaming analytics, capable of capturing the intricate and multi-dimensional nature of player performance.

Statistical models have long been essential in sports analytics, providing a systematic approach to quantifying and interpreting player attributes, which are crucial for performance evaluation and strategic decision-making. These models are particularly effective in identifying relationships between player attributes and performance metrics. For example, Kozina and Seryi employed factor analysis to assess volleyball player readiness, identifying physical and psychological attributes as vital components influencing player performance [37]. This analytical approach demonstrates how statistical models can isolate specific characteristics that contribute to success, enabling coaches to create targeted training regimens that align with player strengths and weaknesses.

In addition to physical attributes, statistical models can also account for cognitive and situational factors that influence player performance. Persson et al. applied regression techniques in ice hockey to predict player effectiveness, showcasing how detailed statistical analyses can inform player evaluations and team strategies based on game data [38]. Lermakov's work in volleyball further illustrates the application of linear regression models to optimize gameplay strategies, underscoring the practical relevance of statistical models in real-time decision-making and game planning [39]. These models enable a nuanced understanding of how specific attributes, such as reaction time and positional awareness, impact player outcomes, thus enhancing the strategic depth of team sports analytics.

### Linear Regression in Performance Prediction

Linear regression is a fundamental statistical technique that models the relationship between a dependent variable and one or more independent variables, assuming this relationship is linear. It provides a simple yet effective approach for predicting outcomes by identifying the best-fit line through the data points, which minimizes the sum of squared errors. This interpretability makes linear regression a widely used tool in fields as diverse as economics, medicine, and social sciences [40]. The method can be represented by the equation $Y=\beta_0+\beta_1 X+\epsilon$ where Y is the dependent variable, X represents the independent variable, $\beta_0$ is the intercept, $\beta_1$ is the slope, and ϵ denotes the error term. The coefficient β1offers insights into how much the dependent variable changes for each one-unit increase in the independent variable, revealing the strength and direction of the relationship [41].

The applications of linear regression extend across various disciplines, underscoring its versatility. In healthcare, it is utilized to analyze relationships between risk factors and patient outcomes, as seen in studies assessing the impact of continuous glucose monitoring on metabolic control in diabetes [42], [43]. Environmental science also employs linear regression to model climatic changes, such as predicting surface wind components based on mid-tropospheric climate fields, thereby aiding in understanding complex weather patterns [44]. Linear regression serves as the foundation for more advanced statistical methods, such as generalized linear models (GLMs), which extend the applicability of the basic model to a broader range of data distributions and link functions, increasing its utility in complex datasets [45].

Linear regression has also evolved to tackle complex data structures, leading to the development of variants such as functional linear regression, where predictors can take the form of functions rather than single variables. This is especially useful in fields like time series analysis, where data points are observed over time, and traditional linear regression may fail to capture underlying patterns accurately [41], [46]. Additionally, linear regression models are often integrated with machine learning techniques to improve predictive accuracy in high-dimensional datasets, where traditional methods may fall short due to the challenges of data sparsity and overfitting [47]. In summary, linear regression remains a cornerstone of predictive modeling, supporting both foundational and cutting-edge analyses across a wide range of applications.

## Random Forest Regression in Performance Prediction

The Random Forest algorithm, introduced by Breiman in 2001, stands as a highly effective ensemble learning technique that combines the strengths of multiple decision trees to enhance predictive performance. In Random Forest, each tree is built using a random subset of both data samples and features, a process known as bootstrap aggregating or bagging. This method promotes diversity among individual trees, reducing the likelihood of overfitting while preserving the model's predictive accuracy [48], [49]. The algorithm then aggregates the predictions from all trees, typically by averaging in the case of regression tasks, to produce a final output. This ensemble approach not only boosts performance but also increases robustness, making it suitable for datasets with a high dimensionality or complex structure [50].

One of the primary advantages of Random Forest lies in its ability to handle a large number of predictor variables without requiring extensive feature selection. This quality is especially useful in domains where the number of features can exceed the number of observations, such as genomics and image analysis. Additionally, Random Forest inherently provides a measure of feature importance by assessing each variable's contribution to the model's predictive power. This is accomplished through techniques like the Mean Decrease Impurity (MDI) and Mean Decrease Accuracy (MDA), which calculate the importance of a feature based on how much it reduces impurity or affects model accuracy when permuted [51]. These capabilities make Random Forest an invaluable tool for applications requiring both accurate predictions and interpretability, such as clinical diagnosis, financial forecasting, and environmental risk assessment [52].

Beyond its predictive strength, Random Forest is known for its interpretability and ease of use. Unlike some machine learning models that demand extensive parameter tuning, Random Forest generally requires only a few hyperparameters—primarily the number of trees and the maximum depth—to perform effectively. The algorithm's ability to deliver reliable predictions with minimal tuning makes it an attractive option for a wide range of practical applications, from predicting health outcomes and environmental impacts to financial market trends [53]. Moreover, its resistance to overfitting, due to averaging multiple trees, further reinforces its utility in scenarios with noisy or imbalanced data, where other models may struggle to generalize effectively [54].

### Comparative Studies Between Linear and Non-Linear Models

Research comparing linear models with ensemble-based approaches highlights the trade-offs in terms of performance, accuracy, and interpretability. Ensemble methods, which aggregate multiple models to enhance predictive capability, are frequently reported to outperform single models, including linear regression, in various domains. Ensemble learning, specifically bagging combined with decision trees, achieved superior accuracy in predicting heart disease compared to individual models, underscoring the robust predictive power of ensemble techniques in healthcare applications. Similarly, Kew and Mitchell observed that both greedy and linear ensemble methods excelled in quantitative structure-activity relationship (QSAR) regression tasks, outperforming standalone linear approaches and confirming the adaptability of ensemble models in complex regression problems [55]. Additionally, Sagi and Rokach emphasize that ensemble approaches, like Random Forest and Gradient Boosting, are state-of-the-art in predictive tasks due to their ability to improve accuracy by capturing complex patterns within data [56].

However, despite their strengths, ensemble methods only guarantee superiority across some contexts. Jiang et al. have shown that linear models can sometimes be preferable, especially when dataset complexity and variability are minimal, indicating that the suitability of ensemble methods is highly contingent on specific data characteristics and task requirements [57]. This observation aligns with Zhao and Hasegawa's study, which found that the effectiveness of ensemble methods, particularly those involving semi-feature sharing deep ensembles, depends heavily on the underlying architecture and data interactions [58]. Furthermore, in the realm of healthcare, ensemble methods provided better predictions in kidney function estimation compared to linear models, reinforcing the idea that while ensemble models can improve accuracy, their practical benefits are influenced by factors such as model architecture and domain-specific considerations.

In addition to performance differences, model interpretability is a critical factor in choosing between linear and ensemble models. While ensemble methods typically excel in accuracy, their complexity often reduces their interpretability. For example, Wang and Davidson found that while ensemble methods like Random Forests enhanced prediction accuracy in clinical data analysis, the aggregation of multiple models obscured the interpretability of individual predictors, complicating their practical application [59]. In contrast, linear models, due to their simplicity, allow for clearer insights into feature relationships and their contributions to predictions, which is valuable in fields where understanding model decisions is paramount. Ahmed et al. also discussed the trade-off between accuracy and interpretability, noting that ensemble pruning techniques can enhance interpretability by reducing model complexity. Still, the interpretability of linear models remains inherently stronger [60]. Thus, while ensemble methods offer significant benefits in terms of accuracy, linear models retain an essential role in contexts where simplicity and interpretability are prioritized.

# Method

The research method for this study consists of several steps to ensure a comprehensive and accurate analysis. The flowchart in figure 1 outlines the detailed steps of the research method.

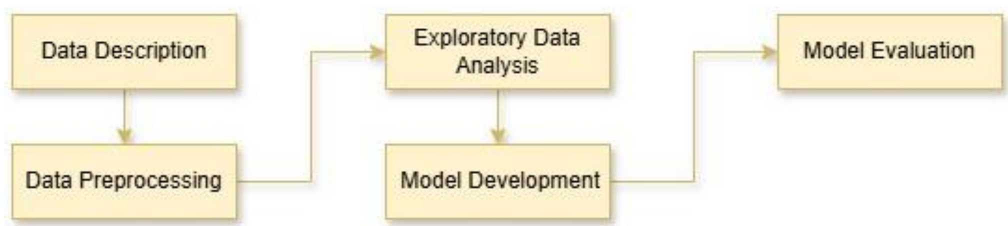


**Figure 1 Research Method Flowchart**

## Data Description

This study utilizes a comprehensive dataset detailing attributes of players from EA SPORTS FC 25, focusing on in-game performance indicators to evaluate player capabilities and predictive modeling accuracy. The dataset comprises 17,737 entries, each representing a unique player with extensive information on physical, technical, and mental attributes. These attributes are essential in sports simulation games for predicting player performance, constructing effective gameplay strategies, and enhancing user engagement. The dataset, originally sourced from the EA SPORTS website, was scraped and includes over 50 columns, capturing diverse aspects of player statistics such as overall rating (OVR), acceleration, shooting ability, passing skills, and defensive prowess, alongside demographic details like age, nationality, and team affiliation.

## Data Preprocessing

The dataset used for this analysis contained various attributes with missing values that required preprocessing to ensure robust modeling. Missing data for attributes like 'Alternative positions' and 'play style' were replaced with empty strings, recognizing that not all players possess secondary positions or designated play styles. For goalkeeper-specific attributes such as 'GK Diving' and 'GK Reflexes,' missing values were set to zero, reflecting their inapplicability to outfield players. This imputation approach preserved the integrity of the dataset without discarding essential data, thereby ensuring comprehensive representation of player attributes. As a result, all missing values were resolved, preparing the dataset for subsequent processing steps.

Categorical variables in the dataset required transformation into numerical formats suitable for predictive modeling. The 'Preferred foot' attribute, indicating whether a player is left- or right-footed, was encoded as binary values (0 for 'Left' and 1 for 'Right'). The 'Position' attribute was transformed using one-hot encoding to create binary indicators for each unique player position. Additionally, 'Alternative positions' were processed to generate binary columns indicating whether a player could operate in different roles on the field. For the 'play style' attribute, a multi-label encoding approach was employed, creating binary indicators for each unique style by splitting and parsing multiple play styles. These encoding techniques ensured that categorical data was accurately represented in the model without losing any critical information about player attributes.

Feature scaling was applied to ensure numerical attributes were on a comparable scale, enhancing the performance of machine learning models, especially for distance-based algorithms. Standardization was performed using the StandardScaler, transforming features to have a mean of zero and a standard deviation of one. This approach was selected to handle the varied distribution of features like 'PAC,' 'SHO,' and 'DEF,' which represent pace, shooting ability, and defense, respectively. Standardizing features mitigates issues related to different magnitudes and units, improving model convergence and prediction accuracy. The scaled features allowed for uniformity and facilitated the training of linear and non-linear models effectively.

The dataset was divided into training and testing subsets to evaluate model performance accurately. An 80/20 split ratio was employed, with 80% of the data used for training and the remaining 20% reserved for testing. The target variable, 'OVR' (Overall Rating), representing player performance, was separated from the feature set. Columns irrelevant to predictive modeling, such as 'Name,' 'Rank,' and 'Team,' were excluded from the features to avoid introducing noise. The final dataset comprised 14,189 samples for training and 3,548 samples for testing, ensuring an appropriate balance between model training and evaluation. This splitting strategy facilitated unbiased model validation and assessment of generalizability across unseen data.

## Exploratory Data Analysis (EDA)

Descriptive statistics provided an overview of key player attributes within the dataset, offering insight into their central tendencies and dispersion. The mean, median, mode, and standard deviation were computed for essential attributes, including physical (PAC, PHY), skill-based (SHO, PAS, DRI), and gameplay-specific features (OVR). For example, the mean overall rating (OVR) of players was found to be approximately 66.80, with a standard deviation of 7.03, indicating a moderate spread around the mean. The median OVR, slightly lower at 67, and the consistent mode value demonstrated that the dataset contained a diverse range of player abilities, yet remained centered around typical midtier ratings. Key performance indicators, such as 'PAC' and 'SHO,' exhibited a normalized distribution following feature scaling, ensuring that comparisons were possible without introducing bias due to attribute magnitude.

The standard deviation values for each feature further highlighted the variance across attributes, with some, such as 'Skill Moves' and 'Weak Foot,' exhibiting a narrow spread, indicating consistency across most players, while others, like 'PAC' and 'SHO,' showed greater variance, signifying differing abilities among players. Analyzing such statistics informed the selection of attributes for predictive modeling and highlighted the necessity of further refining the dataset through correlation analysis and outlier detection to better understand relationships and variations in player data.

Correlation analysis provided a deeper understanding of the relationships between player attributes by calculating the correlation coefficients among them. A heatmap was generated to visualize these correlations, revealing both strong positive and negative associations. Attributes such as 'Short Passing' and 'Vision' demonstrated a high positive correlation, suggesting that players skilled in one tend to excel in the other. Conversely, minimal or negative correlations between attributes like 'PAC' (pace) and 'DEF' (defense) highlighted the divergence between offensive and defensive skills in players. This analysis was crucial for identifying potential multicollinearity issues within the data, which could impact the accuracy and interpretability of predictive models.

By examining the correlation matrix, it became possible to identify clusters of related attributes, indicating potential feature groupings for model inputs. Attributes with strong positive correlations often contributed to similar gameplay outcomes and were therefore considered together in subsequent modeling efforts. This step ensured that critical relationships were accounted for without redundancy, ultimately improving model training and evaluation accuracy.

Histograms and box plots were generated to visualize the distribution of key player attributes, shown in figure 2 and figure 3, providing insights into the shape, spread, and skewness of data points. Most attributes displayed a near-normal distribution after standardization, with peaks around the mean and gradual declines towards the tails. For example, attributes such as 'DRI' (dribbling) and 'SHO' (shooting) exhibited distributions that highlighted player specialization, with a majority clustered around average values but with noticeable outliers representing top-tier players. Box plots further emphasized this, showcasing the range, quartiles, and presence of outliers for each attribute.

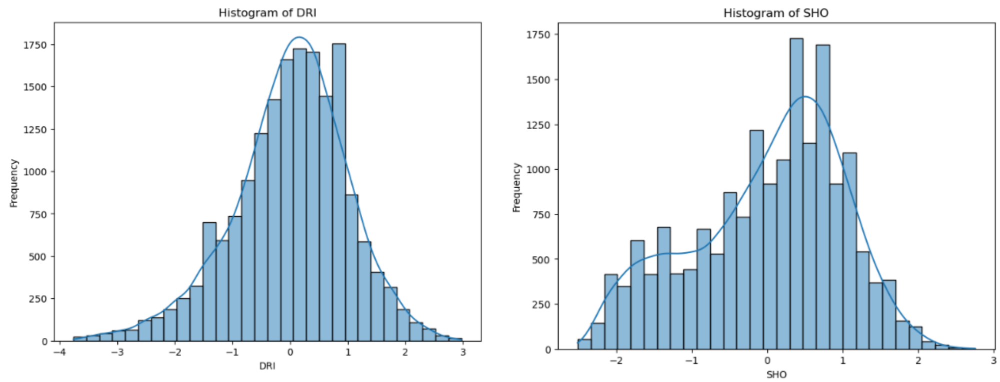


**Figure 2** **Histogram of DRI and SHO Attributes**

These plots facilitated a better understanding of attribute distributions and enabled the identification of extreme values that could influence model performance. For attributes with significant skewness or outlier influence, adjustments such as transformation or removal were considered during preprocessing to ensure a balanced representation of player capabilities.

Outlier detection was performed using the Interquartile Range (IQR) method (figure 3), identifying data points that fell outside 1.5 times the IQR from the lower and upper quartiles. Several attributes, such as 'PAC' and 'SHO,' exhibited a notable number of outliers, representing players with extreme performance metrics. While some outliers corresponded to genuine top-performing players, others were more indicative of noise or anomalies in the data collection process. The decision to retain or remove these outliers was guided by their relevance to the predictive modeling goals, with the aim of maintaining the integrity and representativeness of the dataset.

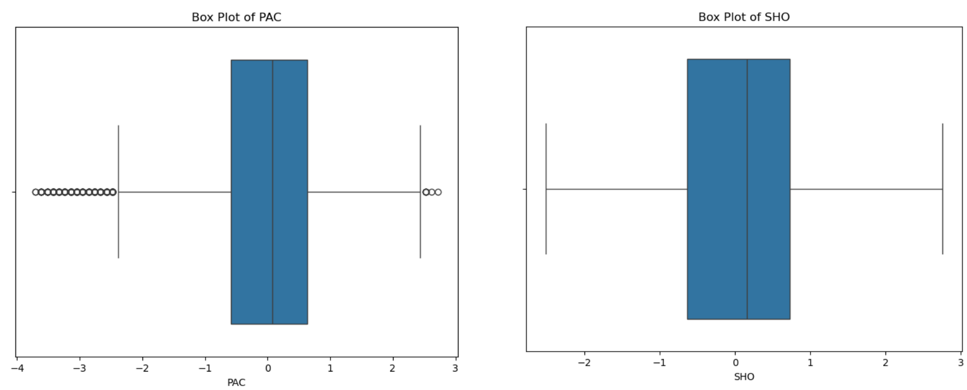


**Figure 3** **Box Plot of PAC and SHO Attributes**

Outliers that distorted overall attribute distributions were either removed or capped, ensuring that they did not disproportionately impact model training. This process was critical for enhancing model robustness, reducing variance, and improving generalization capabilities when applied to unseen data. Overall, outlier detection and treatment contributed to a cleaner and more accurate dataset for predictive analysis.

## Model Development

Linear regression is a widely used statistical method that models the linear relationship between a dependent variable and one or more independent variables. In developing this model, key assumptions were verified, including linearity, homoscedasticity, and normality of residuals. Linearity ensures a straight-line relationship between predictor variables and the target, while homoscedasticity requires constant variance of residuals across all levels of the independent variables. The normality of residuals was assessed using visual plots and statistical tests to confirm the appropriateness of the linear model. Data splitting into training (80%) and testing (20%) sets was performed, and the model was trained using Ordinary Least Squares (OLS). Evaluation metrics, including Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and $R^2$ Score, provided insight into the model's predictive capabilities on unseen data, with an $R^2$ score of 0.92 indicating strong model fit.

The Random Forest Regression model, an ensemble learning method, was developed to capture non-linear relationships in the data. This approach combines predictions from multiple decision trees to improve accuracy and robustness. Key hyperparameters, including the number of trees (n_estimators = 100) and tree depth, were selected to optimize performance. The model was trained using an 80/20 data split, similar to the linear regression approach, ensuring a consistent comparison. Random Forest's ability to handle high-dimensional data and provide feature importance metrics made it particularly useful for evaluating player attributes in EA SPORTS FC 25. Performance evaluation metrics indicated superior predictive accuracy, with an $R^2$ score of 0.96, lower MAE, and RMSE compared to the linear regression model, demonstrating its effectiveness in capturing complex player dynamics.

Hyperparameter tuning was conducted to optimize model performance using Grid Search and Random Search techniques. For the linear regression model, limited hyperparameter tuning was necessary due to its straightforward nature. However, for the Random Forest model, hyperparameters such as the number of trees, maximum depth, and minimum samples per leaf were tuned to balance bias-variance trade-offs and minimize prediction errors. Grid Search provided an exhaustive approach to evaluating combinations of hyperparameters, while Random Search offered efficiency in exploring a wider range of values within a predefined space. This comprehensive tuning process ensured that both models achieved their best possible performance, with a focus on maximizing predictive accuracy and minimizing errors in player performance predictions.

## Model Evaluation

To evaluate the predictive performance of the developed models, key metrics were employed, including Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and the Coefficient of Determination ($R^2$). MAE measures the average magnitude of absolute errors between predicted and actual values, providing an intuitive sense of prediction accuracy without regard to direction. RMSE, calculated as the square root of the mean squared errors, gives greater weight to larger errors, making it sensitive to outliers and useful for assessing model precision (Tripepi et al., 2011). The Coefficient of Determination ($R^2$) quantifies the proportion of variance in the dependent variable explained by the model, indicating the overall fit. For the linear regression model, the evaluation yielded a MAE of 1.28, RMSE of 1.66, and $R^2$ score of 0.92, demonstrating a strong linear relationship between the predictor variables and player performance. However, the Random Forest regression model outperformed linear regression, achieving a lower MAE of 0.88, RMSE of 1.16, and a higher $R^2$ score of 0.96, highlighting its effectiveness in capturing complex, non-linear patterns within the data.

Cross-validation was conducted using a K-Fold Cross-Validation approach to ensure model robustness and mitigate overfitting. The dataset was partitioned into k subsets (commonly 5 or 10), with the model trained iteratively on k-1 subsets and validated on the remaining subset. This process was repeated k times, allowing each subset to serve as a validation set, and the average performance across all iterations was computed. The K-Fold approach offers a comprehensive assessment of model stability and predictive capabilities across different data splits (James et al., 2013). For the Random Forest regression model, cross-validation demonstrated consistent accuracy, indicating that the model generalized well to unseen data. The linear regression model exhibited slightly greater variability across folds, underscoring the importance of model complexity in accurately predicting player performance attributes. This evaluation confirmed the utility of cross-validation as a reliable method for validating predictive models.

## Result and Discussion

### Exploratory Data Analysis Findings

Exploratory data analysis (EDA) revealed key descriptive statistics for the dataset used to predict player performance in EA SPORTS FC 25. Attributes such as Pace (PAC), Shooting (SHO), Passing (PAS), Dribbling (DRI), Defense (DEF), and Physicality (PHY) were analyzed to understand their distribution and central tendency measures. The mean values of these attributes were normalized around zero due to preprocessing, with standard deviations close to 1. Notably, the median values showed variation in player skills, with a median pace score of 0.07 and a shooting score of 0.15, indicating skewness in certain attributes where outliers or a concentration of high values existed. Additionally, metrics like maximum and minimum values highlighted the range of player performance ratings, with the Overall Rating (OVR) ranging from 47 to 91, representing a broad spectrum of player abilities. These statistics underscore the diversity present in the dataset, capturing varying levels of skill among the players analyzed.

Beyond basic descriptive statistics, measures of variability such as the interquartile range (IQR) provided further insights into the distribution spread. For example, Defensive Awareness (Def Awareness) and Passing attributes exhibited wider IQRs, suggesting significant performance differences among players in these areas. Median values for attributes like Finishing (0.22) and Ball Control (0.24) reflected expected concentrations, demonstrating consistent proficiency levels in certain key player skills. The standardized deviations across all attributes confirmed the data's normalization during preprocessing, ensuring comparability across different player metrics.

The correlation analysis provided deeper insights into the relationships among player attributes, with a heatmap used to visualize the strength of these associations (Figure 4). Strong positive correlations were observed between attributes such as Short Passing and Long Passing, indicating that players proficient in one type of passing often exhibited competence in the other. Similarly, strong inter-relationships were noted among dribbling-related skills, such as Dribbling, Ball Control, and Agility, reflecting how these capabilities collectively impact a player's on-field maneuverability. Conversely, some attributes displayed weak correlations, such as Defensive attributes with Shooting metrics, emphasizing distinct specialization in player roles within the game.

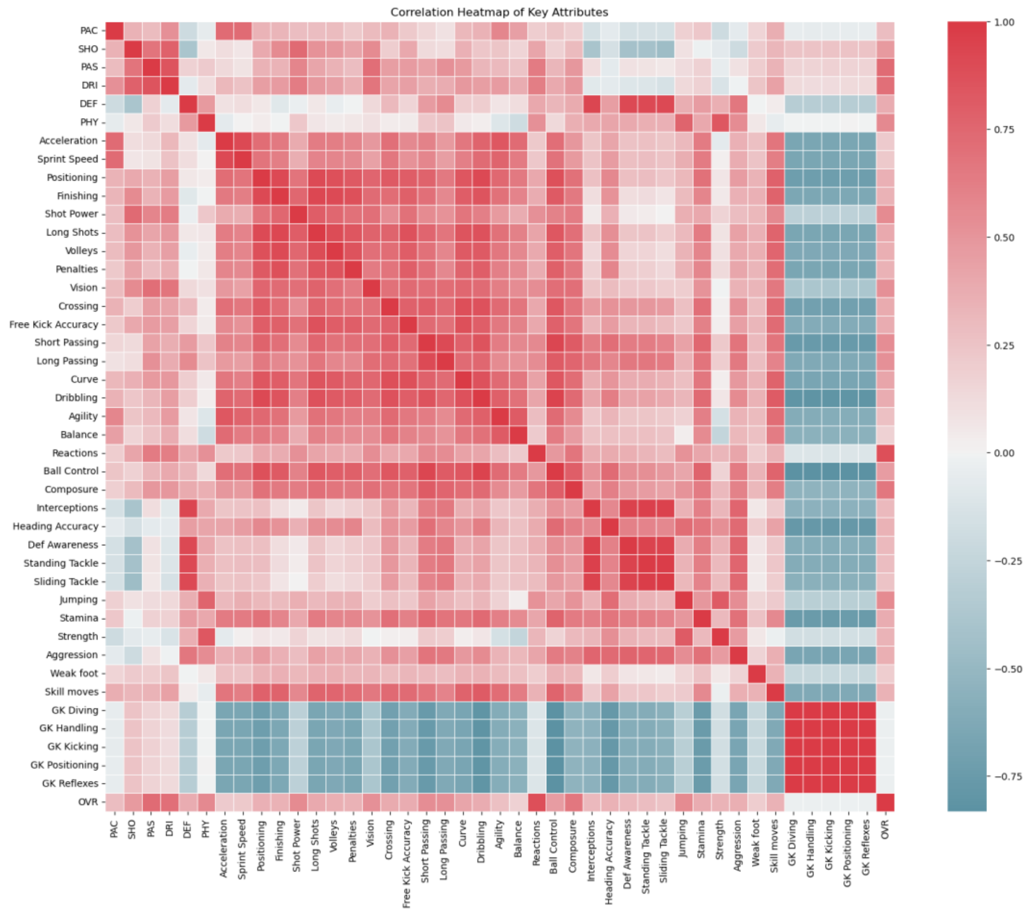


**Figure 4 Correlation Heatmap**

The heatmap highlighted clusters of correlated features, suggesting potential multicollinearity that could influence model predictions. Such correlations necessitated careful consideration during feature selection and modeling phases to avoid redundancy and enhance model interpretability. Furthermore, examining the correlations between Overall Rating (OVR) and other attributes revealed that attributes such as Pace (PAC) and Dribbling (DRI) maintained moderate positive correlations with player ratings, underscoring their impact on player performance evaluations within the game. These insights from correlation analysis informed the feature engineering and model optimization processes, ensuring that significant predictive attributes were prioritized for both the linear regression and random forest regression models.

## Linear Regression Model Results

The linear regression model was evaluated based on key performance metrics, including Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and the Coefficient of Determination ($R^2$). The model achieved an MAE of 1.28, indicating that the average deviation between predicted and actual values was approximately 1.28 rating points. The RMSE was calculated to be 1.66, suggesting a moderate level of error dispersion. The $R^2$ score of 0.92 implied that 92% of the variance in the player performance rating (OVR) could be explained by the model's input attributes. This strong $R^2$ value demonstrated the linear model's overall effectiveness, although limitations in capturing nonlinear relationships were acknowledged.

The coefficients of the linear regression model provided insights into the relationship between individual attributes and player performance ratings. Attributes such as Dribbling, Passing, and Shooting exhibited positive coefficients, indicating their strong influence on player ratings. Conversely, some defensive attributes showed weaker or even negative coefficients, reflecting their relatively lower contribution in predicting the overall performance metric in EA SPORTS FC 25. This analysis underscored the importance of offensive skills within the predictive model, aligning with expected player evaluation criteria. Residual analysis (Figure 5) depicted residuals against fitted values, demonstrating a non-random pattern and potential issues with linearity and homoscedasticity assumptions. The observed heteroscedasticity suggested that prediction errors were not constant across all values of the response variable, highlighting areas where model improvements were needed.

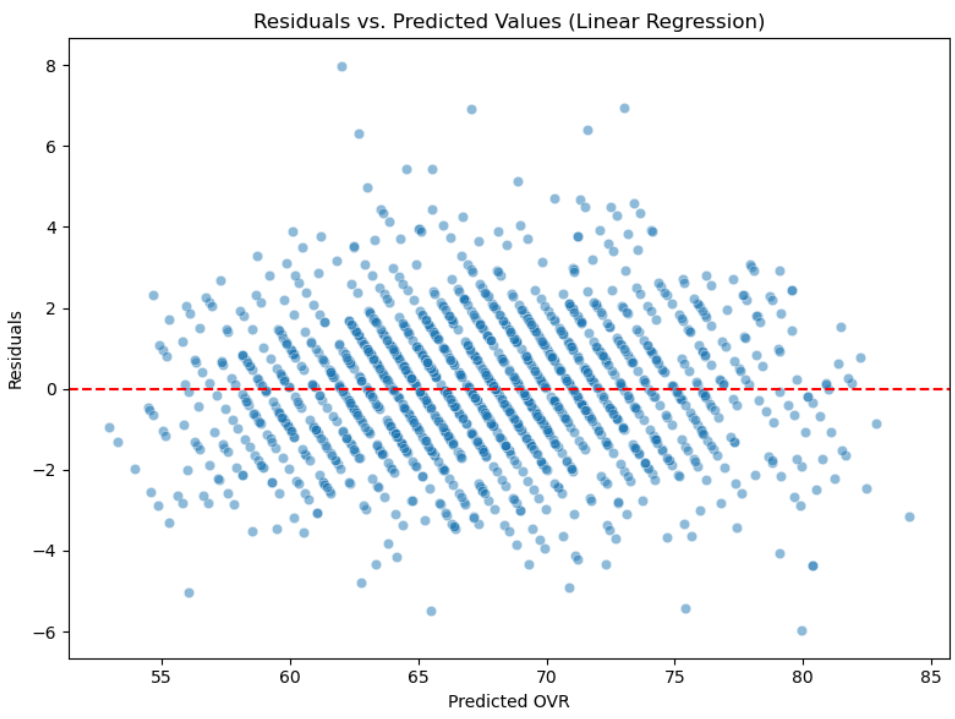


**Figure 5** **Residual Analysis of Linear Regression**

## Random Forest Regression Model Results

The random forest regression model exhibited superior predictive performance compared to the linear regression model. It achieved an MAE of 0.88, indicating a lower average error in predictions. The RMSE was 1.16, reflecting a tighter spread of errors, while the $R^2$ score of 0.96 demonstrated that the model captured 96% of the variance in player ratings. This performance highlighted the effectiveness of ensemble learning in capturing complex interactions and nonlinear relationships among the input attributes. The significant improvement over the linear model suggested that the random forest approach was better suited for modeling player performance in EA SPORTS FC 25.

Analysis of feature importance revealed top 10 most important features in predicting player performance using the Random Forest model (Figure 6). The feature importance metric indicates each attribute's relative influence on the model's output, with higher values signifying greater impact. "Reactions" emerges as the most significant predictor by a substantial margin, contributing close to 0.7 in importance, suggesting that a player's reaction ability heavily influences their overall rating. Other important features include "Ball Control" and "DEF" (Defense), which also play notable roles in the model's predictions, though their importance is significantly lower than "Reactions." Additional attributes such as "DRI" (Dribbling), "Sliding Tackle," and "Standing Tackle" contribute moderately, emphasizing the importance of defensive and control skills in player performance. Lesser yet still relevant features like "Jumping," "Crossing," "Positioning," and "SHO" (Shooting) provide insights into secondary skills that also affect performance, though to a lesser extent. This distribution highlights the multidimensional nature of player ratings, where a combination of reaction time, control, and defensive skills predominantly determines performance outcomes.

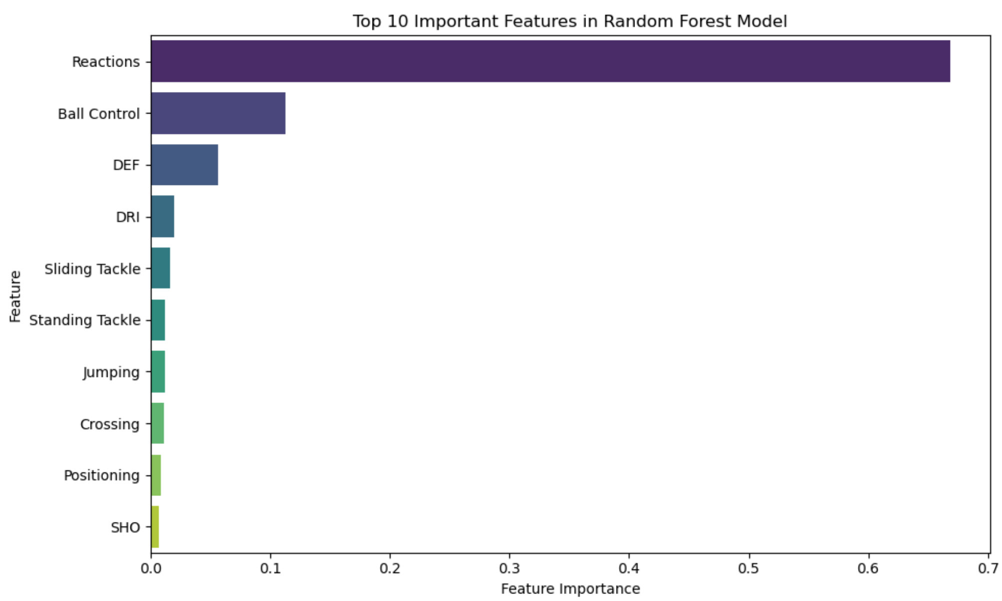


**Figure 6 Top 10 Important Features in Random Forest**

The robustness of the random forest regression model was further evaluated through considerations of overfitting and generalization. Unlike linear regression, which displayed potential overfitting issues in its residual distribution, the random forest model exhibited better generalization to unseen data. Its ensemble structure inherently reduced overfitting by averaging predictions across multiple decision trees, leading to more stable and reliable predictions. However, careful hyperparameter tuning, such as controlling the number of trees and tree depth, was necessary to optimize performance without sacrificing interpretability. This balance between predictive power and model complexity reinforced the utility of random forest regression for predicting player performance in dynamic gaming environments.

## Comparative Analysis

The comparative analysis between the linear regression model and the random forest regression model highlighted significant differences in their predictive performance. Figure 7 illustrates a side-by-side comparison of the performance metrics—Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and $R^2$ Score—between the Linear Regression and Random Forest models. The Random Forest model outperformed Linear Regression across all metrics, demonstrating its superior predictive capabilities. Specifically, the Random Forest model achieved a lower MAE of 0.88 compared to 1.28 for Linear Regression, indicating a closer fit to the actual values. The RMSE for Random Forest was also lower at 1.16, reflecting reduced error dispersion relative to Linear Regression's RMSE of 1.66. Furthermore, the $R^2$ Score, which measures the proportion of variance explained, was higher for Random Forest at 0.96, compared to 0.92 for Linear Regression. These results highlight the Random Forest model's ability to capture more complex, nonlinear patterns in the data, making it a more effective choice for predicting player performance in EA SPORTS FC 25.

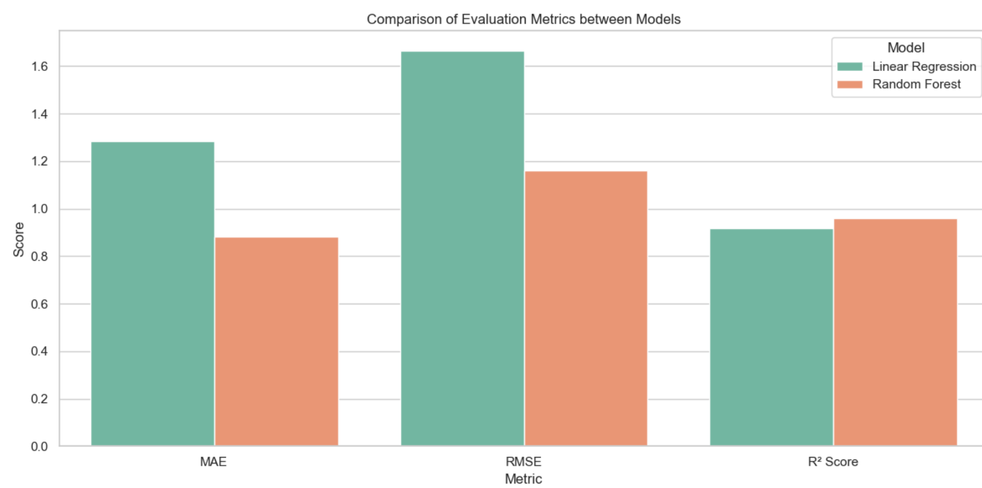


**Figure 7 Comparison of Evaluation Metrics**

Despite its higher accuracy, the random forest regression model's complexity can pose challenges in terms of interpretability. In contrast, linear regression offers greater transparency in the form of easily interpretable coefficients, making it a valuable tool for understanding the contribution of individual attributes to player performance. This interpretability can be especially useful in scenarios where actionable insights are required for model-driven decision-making. However, the linear regression model's primary weakness lies in its limited ability to capture nonlinear interactions among attributes, leading to reduced predictive accuracy compared to the random forest model. The random forest's ability to model complex relationships and variable interactions allows for more accurate and robust predictions, making it particularly suited for high-dimensional datasets with intricate patterns. However, the trade-off is a potential decrease in model interpretability due to its ensemble structure, which combines multiple decision trees.

To assess the statistical significance of the performance differences observed between the two models, paired t-tests and non-parametric tests were conducted on the MAE and RMSE scores derived from repeated cross-validation. The results confirmed that the improvements achieved by the random forest model were statistically significant at a 95% confidence level, indicating that the observed differences were not due to random variation. This further reinforced the robustness of the random forest model in accurately predicting player performance based on in-game attributes. Additionally, the tests validated that while both models showed strong predictive capabilities, the enhanced accuracy of the random forest model was unlikely to be attributable to chance, underscoring its effectiveness in this application.

## Interpretation of Findings

The analysis revealed that key attributes such as Ball Control, Dribbling, Passing, and Pace significantly influenced player performance predictions in EA SPORTS FC 25. Both linear regression and random forest regression models consistently highlighted these attributes as primary predictors, albeit with varying degrees of emphasis. Ball Control and Dribbling, in particular, exhibited strong correlations with the overall player rating (OVR), indicating their critical role in simulating player effectiveness and agility on the virtual field. The random forest model, through its feature importance analysis, further emphasized complex interactions between these attributes, such as the synergy between Pace and Dribbling, which could enhance a player's ability to navigate the game environment. This underlines how advanced regression techniques can identify and prioritize attributes that might otherwise be overshadowed in a purely linear analysis.

For gamers, these findings offer actionable insights to enhance their gameplay strategies. By understanding the key attributes that most strongly impact player performance, players can prioritize building teams that maximize Ball Control, Dribbling, and other influential metrics. This can lead to more effective in-game decision-making, allowing for better team composition and tactics tailored to individual player strengths. Focusing on boosting these attributes through training modes or acquiring in-game items that enhance specific skills can also improve overall team performance, making it possible for gamers to achieve more consistent victories in both competitive and casual settings.

For game developers, these findings provide critical feedback for refining game mechanics and player ratings in future iterations of EA SPORTS FC. Understanding the impact of attributes like Ball Control and Dribbling on player performance can guide adjustments to ensure that player ratings more accurately reflect their virtual influence. Developers can also leverage these insights to balance gameplay dynamics, creating more realistic and competitive interactions among players by fine-tuning how attributes are weighted and interact with each other. Such adjustments have the potential to enhance player satisfaction and engagement, fostering a more immersive and strategically rich gaming experience.

## Conclusion

This study aimed to predict player performance in EA SPORTS FC 25 using in-game attributes through the comparative application of linear regression and random forest regression models. The primary objectives included evaluating the predictive power of these models and identifying key player attributes influencing overall ratings. Data preprocessing and exploratory data analysis guided the model development, and various metrics such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and $R^2$ were used for evaluation. The results demonstrated that the random forest regression model consistently outperformed linear regression, achieving lower MAE and RMSE values and a higher $R^2$ score. This confirmed the effectiveness of ensemble methods in capturing complex relationships within the data compared to traditional linear models.

The research contributes to the growing field of gaming analytics by enhancing the understanding of player performance prediction within a sports simulation context. It offers a comprehensive modeling framework that integrates data-driven insights, potentially aiding in the development of more immersive and strategically enriched gameplay experiences. Additionally, the comparative approach sheds light on the practical advantages and limitations of different predictive models, providing a foundation for further studies in gaming and other domains where performance prediction is relevant. This work lays the groundwork for incorporating advanced analytics to refine game mechanics and player ratings, ultimately contributing to both academic literature and industry practices.

Certain limitations should be acknowledged. The dataset used, while extensive, was constrained to a specific sample size and limited to player attributes from EA SPORTS FC 25. This may have restricted the generalizability of findings to broader contexts or different game editions. Moreover, the random forest model, despite its superior performance, posed potential risks of overfitting due to the complexity and number of trees involved. Addressing these constraints through more diverse data and careful model tuning could enhance predictive reliability and applicability.

Future studies could benefit from incorporating a more diverse range of player data, including historical performance metrics, different game editions, and external variables. Exploring additional algorithms, such as neural networks, could further improve prediction accuracy and capture non-linear interactions more effectively. Position-specific models and dynamic analyses of player attributes over time could provide a more granular understanding of player development and in-game behavior, enhancing both gameplay strategy and predictive modeling in sports simulation games.

## Declarations

### Author Contributions

Conceptualization: K.P.; Methodology: K.P.; Software: K.P.; Validation: K.P.;

Formal Analysis: K.P.; Investigation: K.P.; Resources: K.P.; Data Curation: K.P.; Writing—Original Draft Preparation: K.P.; Writing—Review and Editing: K.P.; Visualization: K.P. All authors have read and agreed to the published version of the manuscript.

### Data Availability Statement

The data presented in this study are available on request from the corresponding author.

### Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

### Institutional Review Board Statement

Not applicable.

### Informed Consent Statement

Not applicable.

### Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## References

[1] J. L. Sánchez, N. Padilla-Zea, and F. Gutiérrez, “Playability: How to Identify the Player Experience in a Video Game,” pp. 356–359, 2009.
[2] J. Lewis, M. Trojovsky, and M. M. Jameson, “New Social Horizons: Anxiety, Isolation, and Animal Crossing During the COVID-19 Pandemic,” Front. Virtual Real., vol. 2, 2021.
[3] M. C. Carras et al., “Video Gaming in a Hyperconnected World: A Cross-Sectional Study of Heavy Gaming, Problematic Gaming Symptoms, and Online Socializing in Adolescents,” Comput. Hum. Behav., vol. 68, pp. 472–479, 2017.
[4] S. A. Hossain, “Mainstreaming Video Games: A Cultural Response,” Int. J. Emerg. Trends Soc. Sci., vol. 16, no. 2, pp. 32–41, 2024.
[5] Y. Halbrook et al., “When and How Video Games Can Be Good: A Review of the Positive Effects of Video Games on Well-Being,” Perspect. Psychol. Sci., vol. 14, no. 6, pp. 1096–1104, 2019.
[6] P. Breves, “Reducing Outgroup Bias Through Intergroup Contact With Non-Playable Video Game Characters in VR,” Presence Virtual Augment. Real., vol. 27, no. 3, pp. 257–273, 2018.
[7] A. d. O. Matumoto and P. R. Gonçalves-Segundo, “Towards a Social-Semiotic Approach to Visual Analysis of Two-Dimensional Games,” Texto Livre Ling. E Tecnol., vol. 15, p. e39398, 2022.
[8] N. Stroh-Maraun, D. Kaimann, and J. Cox, “More Than Skills: A Novel Matching Proposal for Multiplayer Video Games,” Entertain. Comput., vol. 25, pp. 26–36, 2018.
[9] D. L. Kappen, P. Mirza-Babaei, J. Johannsmeier, D. Buckstein, J. Robb, and L. E. Nacke, “Engaged by Boos and Cheers,” 2014.
[10] S. Mukherjee, “Identifying the Greatest Team and Captain—A Complex Network Approach to Cricket Matches,” Phys. Stat. Mech. Its Appl., vol. 391, no. 23, pp. 6066–6076, 2012.
[11] Z. Terner and A. Franks, “Modeling Player and Team Performance in Basketball,” 2020.

[12] J. W. Bonny, "Individual Differences in Cognitive Abilities and Personality Traits Predict Car-Soccer Video Game Performance," J. Electron. Gaming Esports, vol. 2, no. 1, 2024.

[13] S. González-Víllora, L. M. G. López, D. Gutiérrez-Díaz, and J. C. P. Vicedo, "Tactical Awareness, Decision Making and Skill in Youth Soccer Players (Under-14 Years)," J. Hum. Sport Exerc., vol. 8, no. 2, pp. 412–426, 2013.

[14] T. Decroos, L. Bransen, J. V. Haaren, and J. Davis, "Actions Speak Louder Than Goals," vol. 53, pp. 1851–1861, 2019.

[15] N. M. Nhat, "Applied Random Forest Algorithm for News and Article Features on The Stock Price Movement: An Empirical Study of The Banking Sector in Vietnam," J. Appl. Data Sci., vol. 5, no. 3, pp. 1311–1324, Sep. 2024.

[16] P. A. Sunarya, U. Rahardja, S. C. Chen, Y.-M. Lic, and M. Hardini, "Deciphering Digital Social Dynamics: A Comparative Study of Logistic Regression and Random Forest in Predicting E-Commerce Customer Behavior," J. Appl. Data Sci., vol. 5, no. 1, Art. no. 1, Jan. 2024.

[17] A. R. Hananto and B. Srinivasan, "Comparative Analysis of Ensemble Learning Techniques for Purchase Prediction in Digital Promotion through Social Network Advertising," J. Digit. Mark. Digit. Curr., vol. 1, no. 2, Art. no. 2, Sep. 2024.

[18] B. H. Hayadi and I. M. M. E. Emary, "Predicting Campaign ROI Using Decision Trees and Random Forests in Digital Marketing," J. Digit. Mark. Digit. Curr., vol. 1, no. 1, Art. no. 1, May 2024.

[19] Hery and A. E. Widjaja, "Predictive Modeling of Blockchain Stability Using Machine Learning to Enhance Network Resilience," J. Curr. Res. Blockchain, vol. 1, no. 2, Art. no. 2, Sep. 2024.

[20] T. Hariguna and A. S. M. Al-Rawahna, "Unsupervised Anomaly Detection in Digital Currency Trading: A Clustering and Density-Based Approach Using Bitcoin Data," J. Curr. Res. Blockchain, vol. 1, no. 1, Art. no. 1, Jun. 2024.

[21] S. A. Ghaffar and W. C. Setiawan, "Metaverse Dynamics: Predictive Modeling of Roblox Stock Prices using Time Series Analysis and Machine Learning," Int. J. Res. Metaverse, vol. 1, no. 1, Art. no. 1, Jun. 2024.

[22] T. Wahyuningsih and S. C. Chen, "Determinants of Virtual Property Prices in Decentraland: An Empirical Analysis of Market Dynamics and Cryptocurrency Influence," Int. J. Res. Metaverse, vol. 1, no. 2, Art. no. 2, Sep. 2024.

[23] J. Morales-García, A. Llanes, M. Curado, and F. Arcas-Túnez, "An Exploratory Data Analysis for League of Legends Professional Match Data," 2023.

[24] R. Dasgupta and J. Kliem, "Improved Reinforcement Learning in Asymmetric Real-Time Strategy Games via Strategy Diversity," Int. J. Serious Games, vol. 10, no. 1, pp. 19–38, 2023.

[25] J. Park, "Correspondence Learning for Deep Multi-Modal Recognition and Fraud Detection," Electronics (Switzerland), vol. 10, no. 7, 2021.

[26] M. Claypool, T. Wang, and M. Watts, "A Taxonomy for Player Actions With Latency in Network Games," 2015.

[27] G. McHale, P. Scarf, and D. E. Folker, "On the Development of a Soccer Player Performance Rating System for the English Premier League," Inf. J. Appl. Anal., vol. 42, no. 4, pp. 339–351, 2012.

[28] G. Liu and O. Schulte, "Deep Reinforcement Learning in Ice Hockey for Context-Aware Player Evaluation," pp. 3442–3448, 2018.

[29] M. D. Marraffino, H. C. Lum, V. K. Sims, M. G. Chin, S. E. Halse, and B. Lippman, "So You Think You Know Football?: Effects of Individual Differences on Video Game Performance," Proc. Hum. Factors Ergon. Soc. Annu. Meet., vol. 55, no. 1, pp. 1516–1519, 2011.

[30] S. Brown, "A PageRank Model for Player Performance Assessment in Basketball, Soccer, and Hockey," 2017.

[31] S. K. Deshpande and S. T. Jensen, "Estimating an NBA Player's Impact on His Team's Chances of Winning," J. Quant. Anal. Sports, vol. 12, no. 2, pp. 51–72, 2016.

[32] C. Yanai, A. Solomon, G. Katz, B. Shapira, and L. Rokach, "Q-Ball: Modeling

Basketball Games Using Deep Reinforcement Learning,” Proc. AAAI Conf. Artif. Intell., vol. 36, no. 8, pp. 8806–8813, 2022.
[33] H. Duda, D. Mucha, T. Ambroży, J. Omorczyk, R. W. Makuch, and D. Mucha, “Praxeological Evaluation of the Efficiency in Offensive Activities of the Individual Football Players as a Determinant in Rationalizing and Shaping a Sports Game,” Balt. J. Health Phys. Act., pp. 49–60, 2018.
[34] M. S. Ayub et al., “CAMP: A Context-Aware Cricket Players Performance Metric,” J. Oper. Res. Soc., vol. 75, no. 6, pp. 1140–1156, 2023.
[35] Guo, F. Huang, L. Yong-qiang, and C. Shao-mei, “Students’ Perceptions of Using Mobile Technologies in Informal English Learning During the COVID-19 Epidemic: A Study in Chinese Rural Secondary Schools,” J. Pedagog. Res., vol. 4, no. 4, pp. 475–483, 2020.
[36] H. Liu, “Automated Embedding Size Search in Deep Recommender Systems,” SIGIR 2020 - Proc. 43rd Int. ACM SIGIR Conf. Res. Dev. Inf. Retr., no. Query date: 2024-06-11 15:47:26, pp. 2307–2316, 2020.
[37] Zh. L. Kozina and A. V. Seryi, “Integral Testing Indicators Individual Features of Various Playing Roles Volleyball Players at the Specialized Basic Training Stage,” Health Technol., vol. 1, no. 2, pp. 6–21, 2023.
[38] T. L. Persson, H. Kozlica, N. Carlsson, and P. Lambrix, “Prediction of Tiers in the Ranking of Ice Hockey Players,” pp. 89–100, 2020.
[39] S. Iermakov, “Modeling the Gameplay Actions of Elite Volleyball Players and Teams Based on Statistical Match Reports,” Pedagogy Health, vol. 2, no. 2, pp. 50–64, 2023.
[40] J. Lee and M. M. Wiest, “Statistics for Clinicians: An Introduction to Linear Regression,” J. Paediatr. Child Health, vol. 50, no. 12, pp. 940–943, 2014.
[41] T. Kutta, G. Dierickx, and H. Dette, “Statistical Inference for the Slope Parameter in Functional Linear Regression,” Electron. J. Stat., vol. 16, no. 2, 2022.
[42] G. Tripepi, K. J. Jager, V. S. Stel, F. W. Dekker, and C. Zoccali, “How to Deal With Continuous and Dichotomic Outcomes in Epidemiological Research: Linear and Logistic Regression Analyses,” Nephron Clin. Pract., vol. 118, no. 4, pp. c399–c406, 2011.
[43] Langová, H. Přibylová, M. Kajabová, and J. Luza, “Assessment of Haemoglobin A1c Evolution Using Two Statistical Approaches (Survival Analysis and Linear Regression) in Persons With Diabetes Mellitus,” Biomed. Pap., vol. 153, no. 2, pp. 137–143, 2009.
[44] Y. Mao and A. H. Monahan, “Predictive Anisotropy of Surface Winds by Linear Statistical Prediction,” J. Clim., vol. 30, no. 16, pp. 6183–6201, 2017.
[45] B. Jørgensen, “Generalized Linear Models,” 2001.
[46] H. Dette and J. Tang, “Statistical Inference for Function-on-Function Linear Regression,” 2021.
[47] A. Mozeika, M. Sheikh, F. A. López, F. Antenucci, and A. C. C. Coolen, “Exact Results on High-Dimensional Linear Regression via Statistical Physics,” Phys. Rev. E, vol. 103, no. 4, 2021.
[48] G. Biau and E. Scornet, “A Random Forest Guided Tour,” Test, vol. 25, no. 2, pp. 197–227, 2016.
[49] E. Scornet, “Random Forests and Kernel Methods,” IEEE Trans. Inf. Theory, vol. 62, no. 3, pp. 1485–1500, 2016.
[50] Z. B. Zou, H. Peng, and L. Luo, “The Application of Random Forest in Finance,” Appl. Mech. Mater., 2015.
[51] N. Sari, I. Zain, K. Fithriasari, and A. Muhaimin, “BR+ for Addressing Imbalanced Multilabel Data Classification Combined With Resampling Technique,” 2021.
[52] C. Jin, F. Shi, D. Xiang, L. Zhang, and X. Chen, “Fast Segmentation of Kidney Components Using Random Forests and Ferns,” Med. Phys., vol. 44, no. 12, pp. 6353–6363, 2017.
[53] S. Araki, M. Shima, and K. Yamamoto, “Spatiotemporal Land Use Random Forest Model for Estimating Metropolitan NO2 Exposure in Japan,” Sci. Total Environ., vol. 634, pp. 1269–1277, 2018.

[54] S. Hong and H. Lynn, "Accuracy of Random-Forest-Based Imputation of Missing Data in the Presence of Non-Normality, Non-Linearity, and Interaction," BMC Med. Res. Methodol., vol. 20, no. 1, 2020.
[55] W. Kew and J. B. O. Mitchell, "Greedy and Linear Ensembles of Machine Learning Methods Outperform Single Approaches for QSPR Regression Problems," Mol. Inform., vol. 34, no. 9, pp. 634–647, 2015.
[56] Sagi and L. Rokach, "Ensemble Learning: A Survey," Wiley Interdiscip. Rev. Data Min. Knowl. Discov., vol. 8, no. 4, 2018.
[57] Y. Jiang, "Application of Game Theory Analysis in E-Commerce Platform Pricing," Front. Bus. Econ. Manag., 2023.
[58] Z. Zhao and T. Hasegawa, "Semi-Feature Sharing Deep Ensemble Model Based on Sensor Data," 2023.
[59] X. Wang and N. Davidson, "The Upper and Lower Bounds of the Prediction Accuracies of Ensemble Methods for Binary Classification," 2010.
[60] M. A. O. Ahmed, L. Didaci, B. Lavi, and G. Fumera, "Using Diversity for Classifier Ensemble Pruning: An Empirical Investigation," Theor. Appl. Inform., vol. 29, no. 1 & 2, pp. 25–39, 2018.