UPPSALA
UNIVERSITET

UPTEC I 25008

Degree project 30 credits

May 2025

Enhancing Fantasy Premier League Strategies through Machine Learning and Large Language Models

Emil Notelid & Theo Östlund

Master's Program in Industrial Engineering and Management

Enhancing Fantasy Premier League Strategies through Machine Learning and Large Language Models

Emil Notelid & Theo Östlund

Abstract

Fantasy Premier League (FPL) is the official fantasy football game of the English Premier League, where millions of participants worldwide assemble virtual teams of real football players each season. Players must balance their budget, form, schedule, and injuries each week when selecting their starting lineup and substitutes, earning points based on the players performance on the pitch. The combination of in-depth statistical analysis, strategic team building, and continuous decision-making has made FPL one of the most engaging fantasy sports.

In this thesis, an AI-powered digital assistant with explainable capabilities will be introduced to help FPL users make data-driven decisions to maximize their points each round. By constructing both basic and advanced variables, the models are trained on FPL's scoring components. Two modeling methods were used to achieve the predictions: linear/logistic regression and XGBoost. Both gaining similar results in accuracy and actual fantasy points in the simulation which was run from gameweek 1 through 21 in the 2024/25 season. The AI assistant manager's workflow identifies the optimal starting squad and weekly substitutions during the simulation, with the best results being the linear models, yielding 1293 points which would place the algorithm in the top 12% of all FPL managers.

The predictions were fed into a LLM integrated layer which generate personalized recommendations via an LLM. By translating prediction metadata into clear, understandable explanations, the ”black box” problem is addressed. Evaluation of the full-stack solution shows that the natural language explanations increased AI assistant users trust and understanding compared to just raw score predictions. By integrating predictive analytics, optimization constraints, explainable AI and conversational interfaces, this work offers transparent, human-centric decision support in fantasy sports as well as a proof-of-concept for similar applications in other data-rich settings.

Faculty of Science and Technology

Uppsala University, Uppsala

Supervisor: David Sumpter Subject reader: Kaveh Amouzgar

Examiner: Nina Kivinen

Sammanfattning

Fantasy Premier League (FPL) är det officiella fantasyspelet för engelska Premier League, som engagerar miljontals deltagare globalt varje säsong. Spelet går ut på att sätta ihop virtuella lag med verkliga fotbollsspelare, där deltagarna veckovis måste optimera sin startelva och sina avbytare med hänsyn till budget, spelarform, spelschema och skador. Poäng tilldelas baserat på spelarnas prestationer på planen. Denna kombination av djupgående statistisk analys, strategisk laguppbyggnad och kontinuerligt beslutsfattande har etablerat FPL som en av världens mest fängslande fantasysporter.

Denna avhandling introducerar en AI-driven digital assistent som syftar till att hjälpa FPL-användare fatta datadrivna beslut för att maximera sina poäng i varje spelomgång. Genom att utnyttja data från säsongerna 2021/22 till 2023/24 har modeller tränats på FPLs olike poängkomponenter. Detta har innefattat konstruktion av både grundläggande och avancerade variabler. Två modellmetoder användes för att generera förutsägelserna: linjär/logistisk regression och XGBoost. Båda metoderna uppvisade liknande resultat gällande träffsäkerhet och uppnådda fantasypoäng i en simulering som kördes från omgång 1 till 21 under säsongen 2024/25. AI-assistentens arbetsflöde identifierade den optimala starttruppen och veckovisa byten, vilket resulterade i 1293 poäng – en prestation som skulle placeras bland de topp 12% bästa FPL-användarna globalt.

De genererade förutsägelserna matades in i en LLM som i sin tur genererade personliga rekommendationer, skräddarsydda för varje användares specifika lag. Genom att översätta komplexa data till lättförståeliga förklaringar övervanns utmaningen med svårtolkade algoritmresultat. Utvärderingen av helhetslösningen visar att förklaringarna, presenterade i naturligt språk, markant ökade användarnas förtroende och förståelse, jämfört med att endast presentera förutsägelser i tabellformat. Genom att integrera prediktiv analys, optimeringsbegränsningar, förklarande AI (XAI) och ett chat-gränssnitt, erbjuder detta arbete ett transparent och människocentrerat beslutsstöd inom fantasysport, samt fungerar som ett proof-of-conceptför liknande applikationer inom andra dataintensiva domäner.

Acknowledgement We would like to thank David J.T. Sumpter, professor at Uppsala University and co-founder of Twelve Football, and Dominic Jordan, Chief Data Officer at Twelve Football, for their help during the course of this project.

Contents

Acronyms

AI Artificial Intelligence.

AIC Akaike information criterion.

API Application Programming Interface.

DFS Daily Fantasy Sports.

DSR Design-Science Research.

FPL Fantasy Premier League.

FS Fantasy Sports.

LightGBM Light Gradient Boosting Machine.

LLM Large Language Model.

MAE Mean Absolute Error.

ML Machine Learning.

PPDA Passes Allowed Per Defensive Action.

R^2 Coefficient of determination.

RMSE Root Mean Square Error.

SHAP Shapley Additive Explanations.

xA Expected Assists.

XAI Explainable Artificial Intelligence.

xG Expected Goals.

XGBoost eXtreme Gradient Boosting.

xT Expected Threat.

1 Introduction

Fantasy Sports (FS) trace their roots back to the mid-twentieth century, when fans began competing against one another by drafting teams of real-world players and scoring points from those players actual performances. A prize for the season winner was often part of the fun. The rise of the internet later made FS much more sophisticated, automated, and widely accessible to sports fans [1]. Today it is a global phenomenon with millions of players take part, and the sector has grown into a multi-billion-dollar industry [1, 2, 3]. With prizes or wagers on the line, managers increasingly look beyond gut feeling and lean on more advanced methods. The minority of more advanced users mine the ever-expanding sports data now at their fingertips [4], while the majority of the player bases which lacks the competences can turn to third-party apps that cater directly to fantasy users with predictions and decision support, usually packaged as a subscription or a set of premium features.

As the increasing passion and importance of FS for the millions of users, deploying an effective strategy is critical. Not just for bragging rights, but potentially for substantial financial rewards and social prestige among peers. Despite the enjoyment, camaraderie, and intense competitiveness it fosters, there remains a persistent frustration for fantasy managers: how can one confidently leverage the vast, complex ocean of available player statistics and performance data to make truly informed decisions? This thesis aims to propose a novel tool to help Fantasy Premier League (FPL) users with their strategic decisions.

1.1 Fantasy Sports: Applications and Impact

As FS have evolved from a niche pastime into a mainstream component of sports culture and commerce, engaging tens of millions of participants across various sports leagues. In the United States alone, an estimated 45.9 million people took part in FS in 2019 [5], underscoring the tremendous influence of this phenomenon on the leisure and entertainment market. Additionally, FS represent a dynamic convergence of sports fandom and digital innovation, providing interactive platforms where enthusiasts create and manage virtual teams based on real-world athletes. These platforms not only bolster viewer engagement by fostering deeper connections with the sport but also contribute to revenue generation for leagues through enhanced digital interaction and targeted marketing. Moreover, the rise of FS platforms has democratized access to these analytical tools, providing a unique avenue for fan engagement and offering valuable insights into consumer behavior that further shape the future of the sports industry [6].

From a management perspective, FS significantly influence traditional sports consumption. Evidence suggests that FS participants exhibit higher engagement levels in traditional sports media consumption, such as increased viewership and attendance, compared to non-participants [7, 8]. Consequently, FS platforms serve as complementary tools that amplify fan loyalty, team identification, and brand affinity [7, 8]. FS platforms not only serve as a great second hand influence, bolstering sport revenues, but have also become a revenue engine on its own. Globally, the market is forecast to hit USD 37 billion in 2025 and almost double again by 2030, giving clubs access to a fast-growing digital economy well beyond match-day income [2]. In short, FS ecosystems amplify broadcast, license and retail revenues at every step of the value chain, which is why clubs from Ajax to Manchester City now embed fantasy metrics in their commercial dashboards.

Sports clubs and media companies are not the only ones chasing the FS boom. Betting operators are now weaving Daily Fantasy Sports (DFS) style gamification into their products. DFS mirrors traditional fantasy play but compresses the timeline. Contests can last a week, a single day, or even one slate of games, and its sole purpose is to let players wager real money against each other. The format has surged in popularity in the United States, thriving in a regulatory grey zone within an otherwise restrictive gambling landscape [1]. Additionally, FS in its traditional sense, represent a prime example of gamification strategies employed by professional sports leagues. Gamification (the integration of game-design elements into non-game contexts) has been shown to enhance fan interaction and brand loyalty by creating immersive, competitive and social experiences [5]. FS, characterized by their playful yet competitive nature, utilize gamification to drive this fan engagement, create social communities and stimulate sustained interest across seasons [5]. By embedding gamification, leagues not only attract dedicated fans but also capture casual viewers, transforming them into committed followers through interactive participation.

The FPL, the official fantasy football competition of the English Premier League, exemplifies this surge in popularity. FPL has become the world's largest fantasy football game, with over 11 million managers participating in the 2024/25 season [9]. Among the 11 million users, nine million have opted-in to receive marketing from their favorite teams, creating a GDPR-compliant Customer Relationship Management (CRM) funnel at virtually zero acquisition cost [10]. Once inside that funnel, the high-spend "Golden Fans" identified by Chouaten et al. [11] account for a disproportionate share of merchandising and sponsorship revenue. The authors show that data-driven segmentation can raise a club's operational income and smooth the season-to-season volatility that plagues football finances [11]. This massive user base highlights not only the cultural pervasiveness of FS but also the passion and dedication of fans who invest considerable time in managing their fantasy teams.

1.2 Data Analytics in Fantasy Sports

Amid the growing popularity of FS, analytics and data-driven decision making have become increasingly central to success. The proliferation of data (e.g. player statistics, injury reports, historical performance) has given rise to sophisticated predictive models for fantasy team optimization. In both academic research and the hobbyist community, numerous approaches have been proposed to forecast player points and recommend optimal roster moves [12, 13, 14, 15, 16, 17, 18]. However, while the primary focus of these methods has been on achieving high predictive accuracy through powerful algorithms, there is a notable lack of research addressing the explainability of player-based forecasts. Explainability refers to the capability of a predictive model to provide clear insights into how and why particular predictions are made, thereby enhancing user trust and decision-making effectiveness [19]. Addressing this gap is crucial, as improved transparency and understanding of predictive outcomes could significantly benefit both casual players and advanced users seeking actionable and comprehensible insights. But also to promote the general use and methods behind responsible AI design.

In response to these challenges, this thesis proposes and develops an AI-powered digital assistant for FPL users that emphasizes explainability and personalization. The core contribution of this work is the design of a system that leverages machine learning (ML) techniques, optimization and large language models (LLMs) to offer FPL managers tailored recommendations with clear natural-language explanations. In contrast to conventional prediction engines that might tell a user to "pick Player X" with no context, the envisioned assistant provides a conversational interface that can explain why Player X is a good choice, citing factors that contributed to that ML driven prediction. The assistant integrates predictive modeling to forecast player performance with an explainability layer powered by an LLM, enabling it to justify its suggestions in human-like language. By doing so, the system aims to bridge the interpretability gap: users receive advice that is not only data-driven but also understandable and actionable. Importantly, the recommendations are personalized to each user's fantasy team context and preferences. For example, if two managers ask for transfer suggestions, the assistant's advice will differ based on their unique squads, budgets and more, rather than offering a generic list of top players. This personalized, explainable approach is expected to enhance user trust and engagement, as managers can see the reasoning behind suggestions and more confidently apply them to their decision-making.

By developing this AI assistant, the thesis addresses both an academic gap and a market opportunity. From a research perspective, the project contributes to the literature on FS analytics and explainable AI (XAI). While prior studies have predominantly focused on improving predictive accuracy for fantasy outcomes, relatively little work has examined how to make those predictive models transparent and user-friendly for non-expert decision makers. This thesis thus expands the discourse by demonstrating how ML predictions can be coupled with explanatory narratives, aligning with calls for more interpretable and transparent AI in consumer applications. The result is a framework that balances competitive forecasting performance with human-centric design, illustrating a novel approach to decision support in FS and beyond. From a practical standpoint, there is a clear market need for tools that cater to the millions of everyday fantasy managers who lack the time or expertise to manually analyze data. Existing platforms provide extensive statistics and expert analysis, but few offer an interactive, personalized assistant capable of conversing with users and adapting to their individual needs. The AI-powered FPL assistant developed in this thesis is a proof-of-concept that such a tool is feasible and can enhance the user experience. It demonstrates how recent advances in AI, specifically the emergence of LLMs for natural language understanding and generation, can be applied to make FS analytics more accessible and engaging to the general public. In summary, the introduction of an explainable, ML-driven digital assistant for FPL represents a timely intersection of academic innovation and practical utility: it seeks to empower fantasy football enthusiasts with interpretable insights, thereby enriching their participation, and sets the stage for future research on intelligent, user-centric, transparent systems in the FS domain and beyond.

1.3 Broader Relevance

Modern enterprises increasingly rely on digital support tools to transform complex, high-speed data into actionable insights at scale. That wave of intelligent automation is reflected in global spending on digital transformation, which is projected to grow from USD 1.9 trillion in 2022 to USD near 4 trillion by 2027 [20] and the global sports analytics market alone is forecast to reach USD 15.55 billion by 2030 [21]. The demand for AI solutions that not only deliver accurate forecasts but also explain their reasoning in real time are increasing. This thesis responds to that industrial need by implementing and evaluating an AI-powered assistant for FPL managers that integrates transparent ML predictions with natural language explanations via a conversational interface. By doing so, it showcases practical methods for embedding XAI into consumer-facing platforms, enhancing operational efficiency and transparency, and unlocking new monetization opportunities for digital sports and other data-driven industries.

The FS context explored in this thesis serves as a valuable test bed for XAI applications across various consumer-focused sectors. Recent evidence from multiple sectors underscores this shift. In finance, explainable credit-scoring and peer-to-peer lending models make loan decisions auditable for managers and regulators alike [22, 23]. In healthcare, an interpretable scheduler for psychiatric clinics and a broader survey of clinician information needs show that transparent predictions both improve operations and build trust at the point of care [24, 25]. Manufacturing studies reach the same conclusion: a large-scale Industry 4.0 review finds that human-centred explanations are now a prerequisite for safe, effective automation on the factory floor [26]. Together, these studies highlight a cross-industry trend toward AI systems whose power is matched by clarity. Precisely the design objective pursued for FS in this thesis. The next generation of digital services must combine analytics with explainable, intuitive interfaces. This thesis contributes to that vision by demonstrating how XAI can be implemented and evaluated in a popular consumer-facing application. In doing so, it aligns with the strategic priorities of many industries undergoing digital transformation. By tackling these issues in the fantasy football arena, the research provides broader insights into designing AI systems that consumers not only use, but genuinely embrace – a cornerstone for innovation management in the digital economy.

1.4 Purpose and Research Question

The primary purpose of this thesis is to implement and evaluate an AI-powered digital assistant tailored for FPL managers. This assistant combines ML predictions with XAI, utilizing natural-language explanations generated by an LLM. Through this integration, the thesis aims to transform decision-making processes for FS managers, moving from opaque algorithm-driven recommendations toward an interactive and comprehensible support mechanism. The ultimate objective is to offer transparent, insightful, and user-centric decision support, empowering users to confidently make strategic decisions regarding transfers, lineup selections, and captaincy choices.

This study will address the following research questions:

Research Questions: How can explainable ML methods be used to help FPL managers gain a strategic edge?

Sub-questions

- How does XGBoost and Linear models compare in terms of predictive accuracy and fantasy points yield?
- How can LLMs be integrated with ML predictions to generate natural-language explanations and personalized recommendations, and what is their impact on user experience in the context of FPL?

2 Background

2.1 FPL: Rules & Context

Before reading further, it is important to grasp the basic rules behind FPL to understand the rationale behind the solution. FPL is a competition that runs throughout the English Premier League season each year. Consequently it starts in August when the real football season in England kicks off. This is a crucial time for FPL managers as they have to build their starting squad by selecting real-life Premier League players, adhering to a fixed budget constraint of a fictional 100 million. Each player is assigned a starting price reflecting their perceived value. As the season proceeds, each FPL manager will manage their teams, making transfers and strategic adjustments to maximize their total points. These points are directly linked to the actual performances of the selected players in real Premier League matches, see table 1.

Managers compete not only globally but within private leagues created among friends or communities, often involving informal wagers or prizes. Additionally, the official FPL platform offers substantial awards and recognition to top-performing managers globally. This combination of real-life player performances, strategic management, and competitive social interaction has made FPL a highly engaging game, integrating football analytics, strategy, and community engagement.

Event	Points
Playing up to 60 minutes	1
Playing 60 minutes or more (excluding stoppage time)	2
Goal scored by a goalkeeper	10
Goal scored by a defender	6
Goal scored by a midfielder	5
Goal scored by a forward	4
Each goal assist	3
Clean sheet by a goalkeeper or defender	4
Clean sheet by a midfielder	1
Every 3 shot saves by a goalkeeper	1
Each penalty save	5
Each penalty miss	-2
Bonus points for best players in a match	1–3
Every 2 goals conceded by a goalkeeper or defender	-1
Each yellow card	-1
Each red card	-3
Each own goal	-2

Table 1 FPL point system [9]. Rows shaded in green denote the scoring events that are taken into consideration; rows shaded in red denote events that were not modeled in this thesis.

2.2 Delimitations

While the thesis aims to develop a predictive digital assistant to forecast FPL player's performances, and by extension their FPL points yield, not all aspects of football are easily modeled. Some events exhibit high variability, making future outcomes difficult to predict, so certain delimitations have been imposed. As shown in Table 1, point sources such as penalty saves, penalty misses, red cards, own goals, and goals scored by goalkeepers were excluded. Own goals and goalkeeper goals occur very rarely and in isolation, and although red cards and penalty events might be somewhat easier to predict, they were also omitted to reduce noise in the final predictions.

Similarly, certain delimitations were applied in the optimization layer. The prototype assistant described in this report does not account for the chips available in FPL, as shown in table 2. Although chip usage is a crucial strategic element, especially for high-achieving managers who must time them optimally, the limited injury data and other external factors outside our dataset make it difficult for the assistant to recommend chip plays reliably. Therefore, this strategic decision is deferred to the human manager for this project.

Name	**Effect**
Bench Boost	The points scored by your bench players in the next Gameweek are included in your total.
Free Hit	Make unlimited free transfers for a single Gameweek. At the next deadline your squad is returned to how it was at the start of the Gameweek.
Triple Captain	Your captain points are tripled instead of doubled in the next Gameweek.
Wildcard	All transfers (including those already made) in the Gameweek are free of charge.
Assistant Manager	Add a manager to your team to score additional points over the course of three consecutive Gameweeks.

Table 2 FPL chips and their effects.

3 Related Work

3.1 Data Analytics in football

Data Analytics in football is a relatively new concept, but one that is being used more and more to ensure competitive advantage. Data is commonly used to measure the performance of a player or a team which can be used to create advantages in the next match, or measure the probability of a specific team winning.

A study from 2006 [27] showed that data analytics and ML have unused potential in sports analytics. The study focuses on matches played by a specific team and makes a comparison of different forms of ML techniques to find the most reasonable prediction. The study argues that although there are challenges for data analysis in football, such as small data sets, evolution of teams and inherent noise, there is still potential which has not been achieved yet. Although the data set was small, the study shows that accurate predictions can be made even in domains where data is scarce [27].

Since 2006, much has happened in the sphere of football analytics. Terms such as expected goals (xG) have become mainstream in both academia and media. This term is used to explain and give insight into how a player or team should perform, based on underlying factors such as where shots are taken from and how many players there are between the position of the shot and the goal. One of the first time xG was mentioned was in Barnett and Hilditch's article from 1993 [28], where the authors investigated the impact of artificial pitches on a home team's performance. Another study from 2004 examined twelve factors that might affect the success of a shot [29]. Using data from 37 matches during the 2002 World Cup and logistic regression, the authors identified five factors that had a greater effect on the success of a shot. These factors were distance from the goal, angle from the goal, whether the player taking the shot was at least 1 meter from the nearest defender, whether the shot was immediately preceded by a cross, and the number of players between the shooter and the goal. Such factors are still relevant for xG measurements today [30].

In an article by Cwiklinski et al. [31], they study how data anlytics and ML can help a football team with its team building and player transfer. The study uses three metrics to define a “successful” player transfer and these metrics measure the player’s ability against a set threshold, against the average ability of the team and against the average ability of the starting players. This was done by collecting data from open sources, which included 4700 players in the top European leagues. By filtering and cleaning the data, and using well-known classifiers such as Random Forest, Naive Bayes and AdaBoost, the study obtained an accuracy of about 82%, showing that data analytics and ML can be used by a football team to gain competitive advantage.

A proof that expected goals is a fully functional option in football analytics is presented in a study from 2024 by Roccetti et al [32]. In this study, the authors compare the match results with the alternative outcome based on xG. By applying a non-parametric Kolmogorov-Smirnov test, they investigate whether the ranking of the teams based on the real goals is consistent with the ranking based on xG. The results show that there is no significant difference between the two rankings. The study shows that xG can not only be used to measure shot quality, but can also indicate the probability of the outcome of a match.

A study by Rosli et al. [33] presents a comparative analysis of four data analytics techniques: Decision Trees, Neural Networks, Bayesian Networks, and k-Nearest Neighbors. These were used for predicting football match outcomes (win, draw, or lose). Using raw match data from three seasons of the English Premier League (2013–2016), the study extracts key features such as goals, shots, shots on target, and corners from both home and away matches. The evaluation, based on 10-fold cross-validation and measured primarily through prediction accuracy, reveals that Decision Trees achieve the highest average accuracy of 99.56%, followed by Neural Networks at 96.83%, while Bayesian Networks and k-Nearest Neighbors trail behind with average accuracies of 76.41% and 77.54%, respectively. The same authors argue that the exceptional performance of Decision Trees demonstrates the potential of data analytics techniques to effectively predict football outcomes, which can be valuable for club managers and coaches in strategic planning and match preparation. They also suggest that further improvements could be achieved by incorporating additional match attributes and by transitioning from static, historical data to dynamic, real-time data for enhanced prediction capabilities [33]

There are other areas where customized ML and data analytics can be applied in the football market. One study created a two-phase ML framework to estimate the transfer value of football players [34]. The approach addresses key limitations of existing crowd-sourced methods by automatically detecting the most relevant features for different player roles and by tailoring the regression models accordingly. The study demonstrates that by automatically clustering players according to their roles and then applying a tailored regression approach, one can more accurately estimate their market value. This method not only improves estimation accuracy but also provides a systematic way to handle the complexity and heterogeneity inherent in football data.

3.2 Fantasy Sports Predictive Algorithms

FPL has attracted various predictive modeling approaches in research and practice. Early work by Matthews et al. [12] modeled FPL team selection as a Bayesian reinforcement learning problem, using learned point predictions and optimization for transfers; their automated manager ranked around the top 1% of 2.5 million players, outperforming human team selections in 99% of simulated cases [12]. Other studies have formulated FPL (and similar fantasy games) as optimization problems once predictions are available. For example, Bonomo et al. [35] applied mathematical programming to Argentina's fantasy league: one "prescriptive" model for pre-round team picks and a "descriptive" model to evaluate outcomes, achieving top 0.1% to top 10% finishes in real tournaments [35]. Another example is from Becker and Sun [36], which similarly combined performance prediction with mixed-integer programming for fantasy NFL draft and weekly lineup management, showing promising results in simulation [36].

Modern FS analytics typically involve a two-step approach: *predicting player performance* (usually in terms of fantasy points) using ML models, and then *optimizing team selection* based on those predictions. The prediction step converts historical data (player stats, game outcomes, etc.) into forecasts for upcoming games. The optimization step takes these forecasts and selects the highest-scoring lineup that respects game constraints (budget/salary cap, positional requirements, etc.) [13]. Key ML applications include:

Point Forecasting: Estimating how many fantasy points each player will score in the next game or week. Techniques range from simple regression and feature-based models to advanced neural networks. For example, predicting an PL player's fantasy points might use their recent stats (goals, assists, clean sheets) as features [13].

Player Ranking: Using predicted points to rank players by value. Some studies directly optimize ranking metrics (like Spearman correlation between predicted and actual ranks) to ensure the top picks are truly the best performers [14].

Dynamic Decision-Making: In season-long fantasy formats, ML is used to plan sequences of moves (e.g. transfers or draft picks). This can be framed as a Markov Decision Process and tackled with reinforcement learning (RL) to handle the sequential nature of decisions [12].

By leveraging ML in these ways, automated systems can assist fantasy managers in making data-driven decisions, sometimes rivaling human experts. For instance, one article showed that an AI agent could achieve a ranking in the top 30% of FPL using just sequential planning without full information [15]. A number of recent studies have applied ML to forecast FPL player scores, moving beyond simple heuristics. Bangdiwala et al. [17] compared several simpler models for predicting individual player points and found that a straightforward linear regression outperformed decision trees and random forest algorithms which were prone to overfitting. Others have explored deep learning, for example Lombu et al. [16] who implemented a hybrid of Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks to predict player's next-match points using their last five matches of data. Venter and van Vuuren [18] combined statistical models and machine-learning predictors (including ensembles of top-performing methods) to project player scores each week, and fed these into a optimization model for squad selection. Notably, their system, tested retrospectively on the 2020/21 season, would have ranked in roughly the top 4% worldwide, highlighting the efficacy of a data-driven optimal selection strategy.

This is applicable not only to fantasy football. Other sports such as basketball, baseball and American football also have their respective fantasy games, and with it ways to optimize your picks. In recent NBA (National Basket League) fantasy research, Papageorgiou et al. [37] tested 14 ML models; the best performers were ensemble and tree-based models as well as a linear Bayesian Ridge regression and Elastic Net. After generating predictions for all players, they applied a linear optimization to recommend the best lineup, which was then tested in a real fantasy game. The optimized lineup ranked in the top 18.4% out of 11,764 real user lineups, and many of their generated lineups were profitable in the DFS setting. This demonstrates that combining accurate ML forecasts with optimization can yield lineups that perform well against human competitors, which indicates that a combination of linear models and advanced ensemble methods can produce strong accuracy [37].

Similarly to studies previously mentioned, Beal et al. [38] built an AI system for daily fantasy NFL that uses ML point forecasts and lineup optimization, turning a profit in 81% of game-weeks over four seasons [38]. Across these studies, common ML techniques include regression models (linear, ridge, etc.), tree ensembles (random forest, gradient boosting like XGB/Light Gradient Boosting Machine (LightGBM)), and increasingly neural networks. For instance, a 2024 study by Frees et al. explored deep learning with CNNs (Convolutional Neural Network) to forecast FPL player points, benchmarking against Ridge regression and LightGBM. Their CNN outperformed prior models in literature, while the simpler models still achieved solid performance and highlighted key features as important predictors [13].

Optimization, as mentioned, is an important step in the field of FS analytics. Linear programming, or linear optimization, is one mathematical technique for optimizing a linear objective function subject to a system of linear equality and inequality constraints, such as maximizing profit or minimizing cost [39]. This can be used in FS to optimize lineups. For example, in fantasy NBA, lineups usually require selecting e.g. 8 players under a salary cap, fulfilling position slots (PG, SG, SF, PF, C). Linear programming or integer programming is widely used to ensure the lineup meets all constraints while maximizing predicted points. Researchers often formulate this as a linear optimization problem and solve it with tools such as Gurobi or PuLP [40, 37].

This mix of methods shows that ML is widely applied to predict FPL player scores or optimize fantasy teams, from straightforward regressions to complex reinforcement learning and deep neural networks. However, literature on FS and FPL show that efforts prioritize predictive accuracy and optimization of fantasy squads, often with less, or no, emphasis on model explainability or user-facing interpretation.

3.3 Incorporating Insights from the Literature on XAI

Recent advances in XAI underscore the growing need to interpret the inner workings of predictive models, particularly in high-stakes domains where trust and transparency are crucial. More and more organizations have started to integrate AI and ML into their operations to be used in critical decision-making. Their increased impact has led to growing concerns about the transparency and interpretability of these algorithms. XAI is an effective tool to ensure the algorithms solutions are transparent, trustworthy, responsible and ethical so that all regulatory requirements for AI are properly addressed [41]. For instance, Linardatos et al. [19] offer a comprehensive survey of ML interpretability methods, covering both global and local explanation strategies. Their taxonomy partitions existing approaches by important dimensions such as (a) whether a technique is model-agnostic or model-specific, (b) whether it focuses on local or global explanations, and (c) the type of data on which the technique operates (e.g., tabular, image, or text). The authors [19] also highlight that multiple definitions of ”interpretability” and ”explainability” exist, with many researchers using these terms interchangeably. Despite the lack of a single, universally accepted mathematical definition, the authors emphasize the importance of methods that shed light on how a model transforms inputs into outputs. Their review not only surveys a broad range of approaches, from popular local surrogate models such as LIME to more specialized deep-learning explainers like Grad-CAM, but also clarifies the trade-off between model accuracy and transparency.

The survey by Linardatos et al. [19] supports the claim that linear regression methods, while inherently more interpretable, sometimes lack predictive power compared to black-box methods. Nonetheless, the article also demonstrates how post-hoc explanations can mitigate the opacity of complex models, rendering them at least partially explainable. This thesis build on these findings by focusing on *(a)* linear regression as a transparent baseline for predictive tasks in FS, and *(b)* more advanced but opaque algorithms. For the latter, the post-hoc explanation framework Shapley Additive Explanations (SHAP), as recommended by Linardatos et al. [19], is used to ensure accountability and user trust in model outputs.

XAI remains nascent in FPL prediction models, with most early approaches prioritizing accuracy over interpretability. Recent literature indicates that only a few FPL studies explicitly integrate XAI techniques such as SHAP or LIME. For example, Pokharel et al. [42] built an FPL player performance model using XGBoost (predicting returns on investment) and reported decent accuracy (Root Mean Square Error (RMSE) ≈ 2.05) but paid limited attention to model explainability [42]. In contrast, contemporary sports analytics research increasingly embraces XAI, this is exemplified by Moustakidis et al. [43] who leverage SHAP values to interpret an XGBoost model for team performance, identifying which features drive match outcomes [43]. Similarly, Wang et al. [43] apply LIME to a basketball (NBA) game outcome predictor to illuminate the reasoning behind a neural network's season and monthly win predictions. These cases demonstrate that methods like SHAP and LIME can successfully reveal key factors behind model outputs in sports contexts, aligning model insights with domain intuition and improving stakeholder trust. However, there is a clear gap in translating such explainability into the FPL end-user experience. No peer-reviewed work to date deeply explores integrating XAI into user-facing tools for fantasy managers, in fact, current FPL prediction systems rarely offer any transparency or intuitive explanations for their suggestions. Researchers have called for more human-centered XAI in sports analytics [44], yet the literature has scarcely addressed how explainable models can guide fantasy team selection or transfers in practice. While post 2020 studies in sports analytics increasingly incorporate XAI frameworks, FPL-focused models largely remain black boxes. Embedding explainability into fantasy football decision support, is an emerging frontier, underscored by the need to enhance user trust and decision making for fantasy managers.

3.4 LLMs in Symbiosis with ML

One emerging method to accomplish better explainability and transparency in ML tasks is LLMs. The development of LLMs has created significant opportunities for enhancing the interpretability and explainability of complex ML models. Advanced ML models often produce predictions that are difficult to interpret, which poses a barrier to broader adoption across various domains [45]. Integrating LLMs into ML workflows offers a powerful means to translate intricate model outputs into understandable narratives, bridging the gap between complex model behavior and human interpretability [46]. In particular, using LLMs to explain outputs from ML models addresses the ”black box” problem, where the lack of transparency in model decision-making processes hampers user trust and accountability [46]. This approach aligns well with a human-in-the-loop methodology, wherein the user actively interacts with the explanations provided, especially when ML results deviate from prior knowledge or require deeper scrutiny [47].

Fine-tuning LLMs has emerged as a practical strategy to tailor these models specifically to the task of ML result interpretation. Recent advancements in parameter-efficient methods, such as Low-Rank Adaptation (LoRA) and Quantized LoRA (QLoRA), have made this process significantly more accessible and computationally feasible, even on consumer-grade hardware [48, 49]. Such techniques allow LLMs to dynamically emphasize critical features based on input data, providing context-sensitive explanations that adapt effectively to specific use cases [46].

In the context of interpreting results from regression models and tree-based algorithms using metadata and SHAP values, LLMs offer substantial value. Post-hoc explanation methods, particularly SHAP, benefit from the narrative capabilities of LLMs, allowing complex statistical results to be communicated effectively to end-users without technical backgrounds [46]. This integration promotes trust, ensures accountability, and enhances the practical usability of ML models, positioning LLMs as essential components within XAI frameworks.

3.5 Linear Regression in the Theoretical Framework

Linear regression is widely recognized as a foundational approach in both statistics and ML. It is used to model the relationship between one or more explanatory variables and a continuous outcome, aiming to find a best-fit linear equation that describes how variation in the predictors is associated with variation in the response [50]. Its straightforward and interpretable nature makes it suitable for numerous practical applications in fields such as predictive analytics, risk assessment, and forecasting.

In its most basic form, known as simple linear regression, only one predictor is used.

The relationship can be written as

$$y = \beta_0 + \beta_1 x + \varepsilon,$$

where y represents the outcome, x is the predictor, β_0 is the intercept capturing the value of y when x is zero, and β_1 is the slope indicating how y changes given a one-unit increase in x. When multiple predictors are introduced, the method extends into multiple linear regression. In that situation, the model is

$$y = \beta_0 + \beta_1 x_1 + \ldots + \beta_p x_p + \varepsilon,$$

which generally increases predictive power but also makes interpretation more challenging when predictors correlate strongly. As Nimon and Oswald [51] point out, the presence of multicollinearity can obscure each predictors unique role, especially if standard regression coefficients are examined in isolation.

A key benefit of linear regression is its transparent mechanism for mapping input features to an outcome, which is crucial for the development of XAI systems. In contexts such as an AI assistant for FPL, clear reasoning about how each feature contributes to a prediction fosters trust and helps end-users understand model recommendations. Although more complex algorithms might provide superior raw predictive performance, linear regression offers immediate clarity as it explicitly links each predictor to the result through a coefficient. This explicit linkage aids in diagnosing model behavior and conveying results to non-technical audiences.

Beyond standard regression weights, several other indices help clarify the role of each predictor. Techniques such as relative weights, commonality analysis, and dominance analysis can disentangle correlated predictors by partitioning the overall variance explained (R^2) into unique and shared components. According to Nimon and Oswald [51], these methods can illuminate why certain predictors exhibit higher apparent importance in certain contexts and reveal whether suppression or multicollinearity are substantially affecting the model. Such nuanced interpretive approaches can be integrated into broader modeling workflows for transparent decision-making.

Linear regression remains a fundamental tool in predictive analytics and research methodology. Its capability to clearly represent how predictors influence outcomes, along with its adaptability to nuanced interpretative metrics, underscores its continued relevance for modeling tasks that require explainability. In a master thesis that addresses an AI assistant for FPL, drawing attention to linear regression's strengths, while also noting how advanced techniques address key interpretation challenges, provides a strong theoretical foundation for building predictive models that stakeholders can both trust and understand.

3.6 XGBoost

XGBoost is a efficient and flexible implementation of gradient-boosted decision trees that has become a staple in both academic research and industry applications. In football analysis, XGBoost is a well-known tool that has been used to evaluate individual players [52] and predict match results [53]. At its core, XGBoost builds an ensemble of shallow trees one by one, with each new tree trained to correct the errors made by the previous ensemble. Concretely, if our current models prediction for an observation is $\hat{y}^{(t)-1}$, the next tree seeks to predict the negative gradient of the loss function with respect to $\hat{y}^{(t)-1}$. By using both first- and second-order Taylor expansions of the loss rather than just gradients, XGBoost is able to find optimal splits and leaf weights more accurately and efficiently than many earlier boosting frameworks [54].

What sets XGBoost apart is its carefully designed regularization scheme. Whereas vanilla gradient boosting focuses solely on reducing training error, XGBoost adds a penalty term that discourages overly complex trees. This penalty combines a cost for the number of leaves in each tree with an ℓ_2 penalty on the leaf weights themselves. In practice, this means we can control the trade-off between bias and variance by tuning just a few additional parameters: how heavily we penalize extra leaves, how strongly we shrink leaf weights, and how aggressively we scale each new trees contribution via a learning rate. Together, these mechanisms help prevent overfitting even when dealing with noisy or high-dimensional data [54].

An important aspect of applying XGBoost effectively is the careful tuning of its hyperparameters. Hyperparameters such as tree depth, learning rate, regularization terms, and the number of estimators significantly influence model performance. One widely adopted technique for tuning these parameters is grid search, an exhaustive search approach where a predefined set of values for each hyperparameter is systematically explored to identify the combination that yields the best performance on validation data [55]. Although grid search can become computationally expensive as the number of parameters increases, it remains popular due to its simplicity, transparency, and ease of implementation. Alternative approaches such as random search or Bayesian optimization can provide more efficient parameter tuning but often at the cost of increased complexity and reduced interpretability [56].

3.7 Challenges in Evaluating Model Performance in Football Analytics

In predictive modeling, performance metrics such as Mean Absolute Error (MAE), RMSE, and the coefficient of determination (R^2) are commonly used to assess the accuracy of algorithms like linear regression and XGBoost. These metrics provide quantifiable measures of how closely model predictions align with actual outcomes. However, in the context of sports analytics, particularly football, it is argued that these metrics often fail to capture the true value or utility of a model due to the sport's inherent complexity and unpredictability [57]. Football is characterized by a multitude of interacting factors, many of which are unobservable or difficult to quantify. Tactical decisions, player psychology, team dynamics and random events such as injuries or referee decisions all contribute to match outcomes. This high level of noise and variability in the data naturally limits the predictive power of any statistical model, regardless of its sophistication.

Linear regression, while interpretable and computationally efficient, relies on assumptions such as linearity, independence of errors and constant variance of errors. These assumptions are frequently violated in football contexts. For instance, player performance and match outcomes often exhibit non-linear relationships and complex interdependencies [58]. As a result, linear regression tends to underperform when modeling football data, especially when used without substantial feature engineering or interaction terms.

Furthermore, conventional metrics such as R^2 can be misleading in high-variance domains like sports. A low R^2 value does not necessarily imply that a model is ineffective; rather, it often reflects the complex and unpredictable nature of the underlying phenomena. In sports like football, where outcomes are influenced by numerous random and context-specific factors, a model may still provide valuable insights even if it explains only a small portion of the variance. Models with modest statistical accuracy can be particularly useful for identifying trends, informing strategic decisions, or segmenting players based on performance characteristics, even if their overall predictive precision appears limited when viewed through traditional statistical lenses [59].

3.8 Preprocessing of data

When using ML and large data sets, efficient data preprocessing is very important. This involves transforming raw, often noisy data, into clean data compatible with the chosen analysis model. Data directly extracted from a related source or from the real world is often completely raw, which means there is a risk that it contains errors due to incorrect data entry procedures or that it is missing data in places. Such data may be incomplete, missing attribute values or missing variables that may be of interest. It may also contain noise, errors or outliers, as well as incoherent data containing outliers [60]. García-

Aliaga et al. [61] describe that the quality of one's data depends on three elements: *accuracy*, *completeness* and *consistency*. However, the risk is high that the data is often the opposite to begin with. In such a case, data preparation is required to make the data useful for your purpose.

When extracting data, the first step is to analyze which variables in the data are considered relevant for your purpose. One method to work with is to define variables as dependent or independent [62]. A dependent variable depends on the value of other variables, while an independent variable does not. Independent variables can also be described as predictors because we can use the information from these variables to predict the value of a dependent variable.

Another way to identify which variables should be considered relevant for your purpose is to use a correlation matrix. Such a matrix can be used to identify redundant or overlapping information, both between dependent and independent variables. This is best suited to problems where the relationship between the features is expected to be linear. What the matrix does is to give an indication of how strongly and in which direction two variables are related. The correlation is based on a range between -1 and 1, where a large negative value indicates a strong negative correlation, and a large positive value indicates a strong positive correlation. By looking at the correlation value between different variables, one can choose to ignore features that have low correlation and are therefore redundant, and ignore variables that are highly correlated with others as these are considered to be overlapping and therefore redundant as well [63].

Standardization, or normalization, is also an important step in the preprocessing of data. This method scales a data set, and is important when there is a large difference between the ranges of different variables. The data is then scaled down and placed in a range, usually between 0 and 1 or between -1 and +1, so that all variables have the same range but the values of the specific variables retain the same scale [64]. The choice of which variables to normalize is crucial. Columns with higher standard deviation provide better prediction accuracy after normalization. This underlines the importance of careful selection of variables to optimize ML performance [65].

Careful processing of the data can lead to significant improvements in accuracy, as long as the choice of method is adapted to the specific data set and problem. A careful choice of data pre-processing method can improve model performance, reduce noise and overfitting, and lower the computational cost [66].

3.9 Decision-support artifacts across industries

Beyond sports applications like FPL, similar decision-support pipelines, combining predictive models, optimization, XAI, and even natural language interfaces, have been demonstrated across multiple domains. Below are a few peer-reviewed examples from the past decade that illustrate these components in action.

In the financial sector, researchers have developed explainable decision-support systems for credit risk management. For example, Abbaspour Onari et al. [22] present a framework for loan application decisions that integrates ML with optimization and XAI. Their system uses a predictive model for credit scoring and SHAP to identify which input features most influenced each applicant's outcome. They then formulate a counterfactual optimization problem to suggest minimal changes that a rejected customer could make to become creditworthy [22]. After solving this multi-objective optimization (constrained by a game-theoretic model), the system delivers an interpretable recommendation to the customer or manager, thereby clarifying the decision process [22]. Complementing this work, Bussmann et al. [23] develop an XAI approach for peer-to-peer SME lending: an XGBoost model trained on 15 045 European firms raises out-of-sample AUROC from 0.81 (logistic regression) to 0.93, and borrower-level SHAP values reveal the financial ratios that drive each predicted default probability, meeting regulators demands for transparency in financial technology risk management. Together, these studies show how machine-learning techniques, coupled with explainability and optimization, can support strategic decision-making in banking and fintech.

In healthcare operations, decision support pipelines have leveraged interpretable models and optimization to improve scheduling and resource use. Kasaie and Rajendran [24] developed a data-driven system to reduce patient no-shows and waiting times in a psychiatric clinics schedule. They built several prediction models, including multinomial logistic regression and decision trees for transparency, to forecast patient arrival punctuality [24]. The models outputs were analyzed with SHAP XAI, highlighting key factors (e.g. travel distance, lead time, age, diagnoses) that affect whether a patient will be late [24]. By integrating these interpretable predictions into the appointment booking process, the authors propose a decision-support tool that can adjust scheduling or send targeted reminders. Their prototype demonstrates that combining simple forecasts with an explainability layer can aid operational decisions and improve efficiency in clinical settings [24]. At a broader level, Yang [25] surveys XAI for predictive modeling in healthcare and distinguishes two complementary forms of clinician information needs: information-based explanations and instance-based clarifications. Satisfying both, he argues, is crucial for earning clinical trust and embedding AI tools in routine workflows, thereby reinforcing the value of transparent, easily interrogated predictors such as the Kasaie–Rajendran scheduler. This aligns with the idea of using human-interpretable analytics (rather than complex black-box models) to optimize outcomes in healthcare management.

In manufacturing and the wider Industry 4.0 landscape, Ahmed et al. [26] synthesize insights from more than a thousand publications to map what AI/XAI techniques are available, how they are implemented inside production pipelines and where along the value chain they deliver gains. Their survey outlines the main enabling technologies, including IoT, cyber-physical systems, cloud-computing and big data analytics. It also presents case studies in smart factories, predictive maintenance, quality control and human–machine interaction, showing how interpretable methods such as SHAP and LIME are already used to support optimization tools on the shop floor. The authors argue that human-centered explanations are a prerequisite for trust in high-stakes industrial automation. They also flag open research challenges like data bias, energy-intensive deep models and the absence of domain-specific XAI benchmarks [26]. By framing explainability as an indispensable layer of Industry 4.0 decision support, this survey generalizes the lesson from finance and healthcare further, and concludes that transparent, optimization-aware analytics are fundamental for reliable operational choices in modern production systems.

Emerging research also shows how natural language interfaces can be combined with optimization for decision support in domains like logistics and planning. Lawless et al. [67] introduce a novel interactive pipeline that couples an LLM with a constraint optimization solver [67]. In their case study on meeting scheduling, users can express goals and preferences in natural language, and the LLM translates these into constraints for a scheduling algorithm. The system then finds an optimal or near-optimal schedule using constraint programming, and can explain or adjust the solution through dialog with the user [67]. This human-in-the-loop approach highlights the feasibility of conversational decision-support systems: managers or planners could simply 'ask' the system for an optimal plan under certain criteria and get an answer with explanations. The study demonstrated that a hybrid LLM+optimization framework can iteratively elicit user preferences and produce transparent solutions, pointing to a future of decision-support tools that are both powerful and user-friendly [67].

4 Method

This study explores the integration of data-driven analytics and AI technologies to create an innovative tool aimed at helping FPL player's decision making. The approach is built around several key steps, which include data collection, ML model development, LLM implementation and evaluation of results. These approaches are designed to ensure not only predictive accuracy but also to provide users with insights that are actionable, transparent, explainable and tailored to their individual needs.

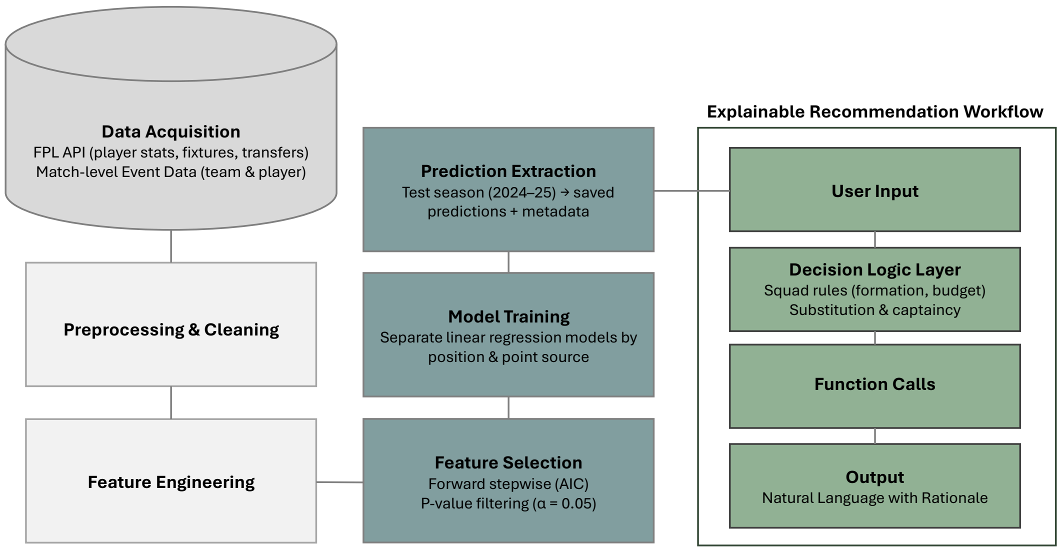

Figure 1 Outline of the workflow of the developed AI assistant, depicting each key stage from initial data processing to the final user output.

4.1 Data collection

To be able to train and validate our predictive models, we assembled a comprehensive dataset combining raw FPL Application Programming Interface (API) outputs with event-level metrics. This multi-source approach ensures that both individual player actions and broader team-level context are captured, along with specific FPL features.

4.1.1 FPL API

In order to make accurate predictions, qualitative and comprehensive data is required. For this study, both historical FPL and Premier League data will be used, integrating player performance metrics with contextual factors such as player form, team dynamics and strength of opposition. This data will be provided to us both through open API and licensing partners. A rigorous process of cleaning, normalizing and aggregating the data will be carried out with an emphasis on the creation of features relevant for predictive modeling. Metrics such as xG and Expected Assists (xA) will be explored to capture the nuances of player performance [32]. External factors, such as venue and opponent team, will also be taken into account.

The data handled in the model included historical FPL data that was retrieved from an open GitHub repository [68]. The repository included a library that contains statistics for each player from all gameweeks between the seasons 2020/2021 to 2024/2025. The variables are listed and described in appendix A, table A.1. The variables are taken directly from the file `merged_gw.csv` and are based on the Premier League's own stats center [69].

The data was processed to be adapted to our task. Missing values and errors were removed. Irrelevant variables were deleted and other variables were transformed to make them easier to apply in the model. The transformation of some of the variables was based on creating a rolling mean from a specific number of matches, instead of only showing statistics from each match. For example, we chose to create a variable called `goals_last5` that showed the average number of goals the player has scored in the last five matches. It was decided that it would be more relevant to look at a player's form over a number of matches instead of just looking at the number of goals a player scored in the last match. The variables included in the processed data can be seen in appendix A, table A.2.

4.1.2 Match-event data

We integrated two licensed match-event datasets: a Player-Match dataset comprising over 60 per-90 metrics and a Team-Match dataset capturing tactical and event-level metrics, see appendix A, table A.3 and A.4. Similar to the variables from the FPL API, an rolling average was created for the variables based on the last five matches. These variables from both the Player-Match and Team-Match datasets were merged with the data from the FPL API to create a final dataset that could be applied to the ML models.

4.1.3 Feature Engineering

In addition to the raw FPL and match-event features, we engineered different metrics that combine multiple underlying actions into single, interpretable efficiency or effectiveness scores, both at the player and team level. For instance, at the player level we capture `progressive_threat_efficiency` by summing a player's per-90 expected threat (xT) contribution from ball progression, ball runs and dribbles, then normalizing by their carries generated xT (with zero carry cases safely handled). Similarly, we compute `finishing_efficiency` as the gap between actual goals per-90 and expected goals per-90, highlighting players who outperform their underlying chance volume. On the team side, we derive measures such as `attacking_efficiency` (the product of box-entry-to-shot rate and shot-conversion rate) and `pressing_intensity_differential` (the difference between opponent PPDA (Passes allowed per defensive action) and own PPDA) to quantify style and defensive pressure in a single glance. These engineered features enrich the model by emphasizing not just volume of activity, but the quality and impact of those actions, improving both predictive accuracy and the transparency of our tool's recommendations. See the engineered features in appendix A, table A.5.

4.2 Model development

The predictive aspect of this project relies on ML models designed to forecast fantasy scores for players based on a range of input variables [70, 71, 72, 73, 38]. Historical FPL data together with performance data specific to each player and team will form the basis for training these models, with validation processes ensuring robustness and reliability. This approach leverages the ability of ML to discover complex patterns and relationships in the data, providing a strategic advantage that cannot be achieved with manual analysis alone. The models are trained on seasons 20/21, 21/22, 22/23 and 23/24 and tested on the season 24/25.

4.2.1 Starting model

Initially, we developed a dedicated binary classifier to predict whether each player would be named in the starting eleven for the upcoming fixture, see table 3. Being part of the starting lineup is the single most important factor in scoring points in the game, and therefore the need for individual models for this type was identified as very important. In addition, results from the other models were considered obsolete as long as the player was not considered to be starting the match. Recognizing that selection criteria vary by role, we trained three separate logistic regression models, one each for defenders, midfielders and forwards. Each classifier ingested features drawn from the raw FPL API and the player and team based datasets. Logistic regression was selected for its simplicity and interpretability in a strictly two-class setting (start vs. non-start) [74]. The model resulted in a probability of whether the player would start or not, which was then converted into binary numbers using a threshold, where 1 describes that the player will start and 0 describes that the player will not start. The threshold is selected based on which gives the best results, and differs based on the separate position models.

Model Name	Position	Description
`xStarted`	DEF, MID, FWD	Predicts the likelihood of a player being in the starting lineup.

Table 3 The models created to anticipate a players performance.

4.2.2 Linear Regression for every source of points

Different point-component models were constructed for each position group, see table 4. For every source of FPL points—goals - assists, minutes played, bonus-points (BPS), conceded goals, saves and yellow cards - we trained a dedicated linear regression model. Segmenting both by point type and by role enables each regression to capture the unique statistical drivers of that contribution (for example, shot creation and key pass metrics for midfielders assists versus tackle and interception counts for defenders clean sheets). All regressions employed the same preprocessing pipeline—missing-value imputation and feature normalization.

Model Name	Position	Description
`xGoals`	DEF, MID, FWD	Predicts the expected number of goals based on player and team statistics specific for chance creation and efficiency.
`xAssists`	DEF, MID, FWD	Predicts the expected number of assists using player creativity and passing metrics.
`xBps`	DEF, MID, FWD	Predicts the Bonus Points System (BPS) score using various performance metrics.
`xConcededGoals`	GK, DEF, MID	Predicts expected goals conceded based on defensive metrics and opponent strength.
`xSaves`	GK	Predicts the expected number of saves a goalkeeper will make in a match.
`xYellow`	GK, DEF, MID, FWD	Predicts the likelihood of receiving yellow cards based on player's disciplinary history.

Table 4 The models created to anticipate a players performance.

Initially, we constructed a series of baseline linear regression models using exclusively the raw data made available through the FPL API. By limiting our feature set to these API-provided variables, such as minutes played, goals, assists, clean sheets and basic match statistics (with rolling averages), we established a controlled experimental framework in which the behavior and performance of the model could be observed without the influence of any external or derived metrics. All preprocessing steps (missing-value imputation, normalization and categorical encoding) were held constant between this baseline and the subsequent, richer models; this ensured that any observed differences in predictive accuracy could be attributed solely to the incorporation of additional data.

Once the baseline models had been trained and evaluated using MAE, RMSE and R^2 on a held out test set, we proceeded to augment our dataset with advanced, derived features (for example player form indices and contextual team-strength metrics). By comparing the out-of-sample performance of the two sets of regressions, we were able to quantify precisely how each layer of added information contributed to our model's ability to predict fantasy points.

For each model and for each playing position, we assembled the full set of candidate predictors and applied a two-step selection procedure. In the first step, we computed the pairwise correlation matrix among all new and existing features and dropped any variable whose correlation with one or more others exceeded a position-specific threshold. This prevented multicollinearity: whenever two features carried essentially the same information and we retained only the one with the stronger individual association relationship to the target and discarded its twin.

In the second step, we ran an automated "Akaike's Information Criterion"-driven (AIC) selection script on the remaining pool of predictors. The script iteratively added individual variables and only those whose inclusion led to a strictly lower AIC for the fitted regression model, thereby ensuring that each retained feature contributed more predictive power than complexity penalty [75]. Because we performed this procedure separately for defenders, midfielders, forwards, and goalkeepers, the final feature sets differ by position. An illustrative subset for the `xGoals` model is shown in table 5. The final variables for every model can be found in appendix B table B.1, B.2, B.3, B.4, B.5 and B.6

Model Name	Position	Variables
xGoals	FWD	threat_last5, relative_team_strength, penalty_chance, cumulative_threat, attack_directness_team_last5, chance_quality_last5, Touches in box per 90_last5, game_control_diff_team_last5, np Goals_team_last5, cumulative_threat
xGoals	MID	selected_last_round, cumulative_threat, relative_team_strength, Linkups per 90_last5, penalty_chance, chance_quality_last5, corner_chance, Dribbles success %_last5, goals_per90_season, Possessions won per 90_last5, np Goals_team_last5, threat_per90_season
xGoals	DEF	relative_team_strength, cumulative_threat, aerial_dominance_last5, corner_chance, Opp. Defensive line height (m)_team_last5, Ball progression (xT) per 90_last5

Table 5 The different variables used in xGoals based on position.

4.2.3 Evaluating of linear regression using XGBoost

We evaluated our linear regression baseline by building a non-linear XGBoost model and comparing their performance. First, we applied the same correlation-based feature selection procedure using a position-based threshold to eliminate highly collinear predictors. Although both algorithms started with the same candidate set, XGBoost's inherent feature importance ranking led us to retain a different subset of inputs for its final fit. The method using XGBoost therefore differed slightly from the one described in Figure 1. As shown in Figure 2, additional steps were taken to tune the model. The hyperparameters were tuned using grid search with cross-validation. The following parameters were included in the grid with varying values: n_estimators, max_depth, learning_rate, subsample, colsample_bytree, gamma, reg_alpha, and reg_lambda. Since the XGBoost models do not include the same metadata as the lin-

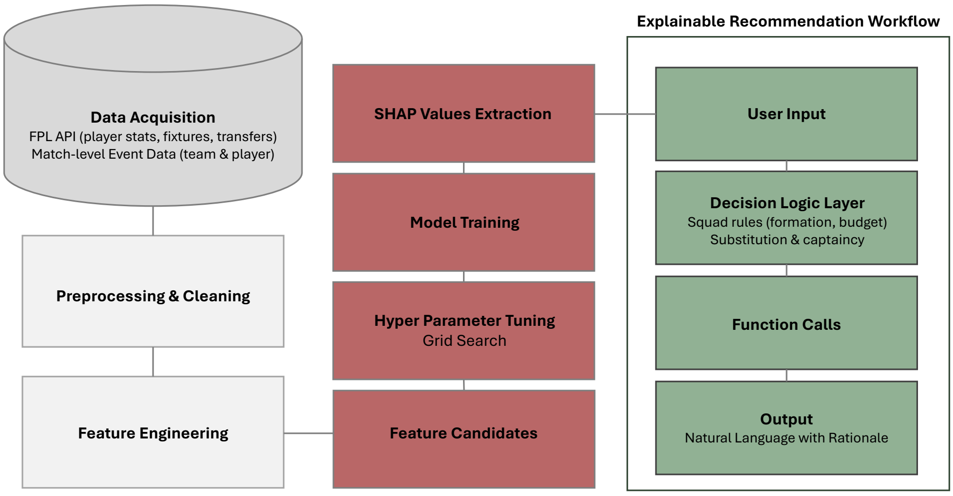

Figure 2 Overview of the process with XGB specific steps taken in the red color.

ear models, an additional step was taken to extract feature importance for explainability. This was accomplished using a SHAP explainer, which was extracted from each model and later used to motivate the projections. Finally, we assessed each model using identical metrics (e.g., RMSE, MAE, R^2) and directly compared their results to determine whether the non-linear approach offered a tangible improvement over ordinary linear regression.

4.3 Squad Optimization

In order to operationalise squad optimization within the FPL decision-support system, a two-tier pipeline was employed. Firstly, the squad selection for gameweek 1 was done though a linear-programming (LP) model. This included two interconnected binary vectors, one for the 15-man squad and one for the eleven starters. Furthermore, constraints in the form of FPL restrictions was embedded directly into the feasible region. The restrictions includes a 100 million budget, a maximum of three players per club, positional boundaries which satisfies the squad restraints seen in 6 as well as formational restrictions also seen in 6. The objective maximize the sum of projected points for the starters, thereby letting the optimizer trade off bench depth against an immediately higher yielding gameweek 1 starting eleven. The optimization problem was solved using the open-source CBC solver. From gameweek 2 an onwards, the starting eleven selection was done similarly to the first gameweek after first choosing an optimal transfer. It scans possible swaps and will make the transfer that gains the highest immediate improvement in terms of projected points. In other words, no long-term planning is utilized beyond the one gameweek horizon. This optimization layer also handles automatic substitutions within the squad, when a player in the starting eleven does not play, an automatic substitution from the bench is deployed. Which players is deployed depends on the bench order which is also taken into consideration and ranked from highest predicted points to lowest.

Constraint type	**Requirement**
15-man squad limits	
Goalkeepers	Exactly 2
Defenders	Exactly 5
Midfielders	Exactly 5
Forwards	Exactly 3
Valid match-day formation	
Goalkeeper	Exactly 1 must start
Defenders	3-5
Midfielders	2-5
Forwards	1-3

Table 6 FPL Squad and Formation Constraints.

4.4 API LLM-integration

To provide a seamless user experience, the tool will integrate APIs for real-time data access and LLMs for natural language generation. An FPL API will be used to provide live updates on team configurations and player performance. In parallel, an LLM (ChatGPT) will be used to produce detailed reports in natural language [76] to translate the explainable insights from the models and SHAP values to natural language insights.

When it comes to the linear models metadata, translating it into natural language explanations is fairly straightforward. By multiplying the model coefficients with the variable values for each observation, it is possible to rank all features according to their impact on the final prediction. These calculations are then passed to the LLM, which generates clear, natural language descriptions. For XGB, this process isn't as direct, so a different approach is required. Here, the saved SHAP explainer is used to calculate SHAP values for each model. Features with positive SHAP values are ranked and grouped together, and the same is done for those with negative values. The larger a feature's positive SHAP value, the greater its impact on increasing the prediction—and the same logic applies in reverse for negative values.

These reports will provide actionable insights, including tailored team recommendations and transfer suggestions. The synergy between predictive analytics and LLM-driven communication ensures that the tool is accessible and user-friendly for FPL players [77, 48, 49].

4.5 Evaluation

The effectiveness of the AI tool will be evaluated by benchmarking against the FPL player base by establishing its percentile position within the wider community. By ensuring that the tool functions competitively and meets the needs of users, its usefulness and relevance shall be thoroughly validated.

The explainability factor and general usability of the digital tool is evaluated through a survey with FPL-users. A survey as an evaluation tool is commonly used, not just when working with human-computer interaction, but in all research fields. The survey will consist of close-ended questions and will be based on a scale, from 1-7 (poor-excellent). The survey was administered to two groups of respondents: one group evaluated the raw data results, while the other used the developed AI assistant. All participants were active FPL users with varying levels of football knowledge and experience in FPL strategy. To ensure a fair evaluation, respondents were assigned to the groups to balance expertise as evenly as possible, so that each group included a comparable range of skill levels. To gain richer insight into user experiences and better contextualize survey results, short semi-structured interviews were conducted with each participant, a common practice in mixed-methods evaluation [78]. The survey can be seen in table 7, also marking what theme each question adhere to.

ID	Question	Theme
1	The data/assistant improved my understanding of how predictions were made.	Interpretability
2	The data/assistant helped me understand why certain players were recommended.	Interpretability
3	The information provided was presented in a clear and interpretable way.	Interpretability
4	I trusted the data/assistant's advice when it conflicted with my own intuition.	Trust
5	Using the data/assistant reduced the time I normally spend on FPL decisions.	Efficiency
6	I found it easy to interact with the assistant/data and get the information I needed.	Usability
7	The information provided increased my confidence in making decisions.	Trust
8	I would use this data/assistant again for future FPL decisions.	Adoption

Table 7 The questions in the survey to evaluate the AI assistant, grouped by theme.

4.6 Research Strategy

The methodology used in this project combines the use of analytics with the accessibility of natural language processing, creating a tool that offers both quantitative predictions and qualitative insights. Although the model's reliance on straightforward linear regression can lead to higher MSE and RMSE, it is argued that this trade-off is justified because it preserves the transparency and interpretability essential in developing responsible AI. Although the method will also be benchmarked against a more advanced method as well, namely XGB. The integration of real-time data improves the tool's ability to provide relevant and timely recommendations.

From the outset, the AI assistant manager was conceived as a one that should remain effective when transplanted from its original problem setting into adjacent decision-support contexts. Our research approach is deliberately framed in the *Exaptation* quadrant of the Design-Science Research (DSR) contribution matrix proposed by [79]. Rather than inventing an entirely new class of artifacts, we transfer a well-established decision-support pattern into a problem space where it has not yet been systematically explored: FPL squad optimization. In DSR terms, scientific value arises when a proven design is shown to function under materially different task characteristics, data semantics and user expectations. The method used in this thesis could be transferred to a different problem and arena, and it would only require a change in configuration, thereby satisfying the external validity criterion stressed by [80]. Thus, this work illustrates how an artifact developed for FPL squad optimization can be effectively repurposed for other decision-support tasks, highlighting the broader applicability and impact of the proposed method within and beyond sports analytics.

5 Results

This section will present the results from the different models created. Initially, the `xStarted`, which predicts whether a player is named in the starting eleven or not, will be presented. The linear regression models which predict the performance of each players will be presented along with the XGBoost models. The models performance metrics will be evaluated and compared with each other, along with a point comparison. Finally, the full-stack AI assistant evaluation will be presented.

5.1 Logistic Regression Models

Table 8 displays the performance of the `xStarted` model when applied to defenders. The model correctly predicted 2,605 non-starts and 1,482 starts out of a total of 4,692 instances. The accuracy was 87.2% for non-starts and 87.0% for starts, indicating consistent performance across both classes. This suggests that the model is well-calibrated in identifying whether defenders are likely to start or not.

	Predicted Non-start	**Predicted Start**	**Total**	**Accuracy**
Non-start	2605	384	2989	0.872
Start	221	1482	1703	0.870

Table 8 Results from the xStarted for defenders.

For midfielders, shown in table 9, the model achieved an accuracy of 88.0% in predicting non-starts and 77.6% in predicting starts. While the overall performance remains solid, the discrepancy between the two classes suggests a slight imbalance, with the model being more conservative and favoring non-start predictions. This may reflect a more variable or competitive selection process in midfield positions.

Table 10 presents the results for forwards, where the model achieved its best performance. The accuracy for non-starts was 91.7%, and for starts, 88.3%. This high performance across both categories indicates that the model can distinguish start patterns

	Predicted Non-start	Predicted Start	Total	Accuracy
Non-start	3838	522	4360	0.880
Start	455	1572	2030	0.776

Table 9 Results from the xStarted for midfielders.

particularly well for forwards, possibly due to more stable or predictable selection behavior in these roles.

	Predicted Non-start	Predicted Start	Total	Accuracy
Non-start	995	90	1085	0.917
Start	53	399	452	0.883

Table 10 Results from the xStarted for forwards.

The `xStarted` model demonstrates strong performance across all positions, with particularly high accuracy for forwards. While defenders also show balanced predictive accuracy, the results for midfielders reveal a tendency to overpredict non-starts. These differences may be influenced by the positional dynamics of team selection, variability in player rotation, or the underlying feature distribution for each role. Overall, the model generalizes well, though slight positional tuning could potentially improve predictions, especially for midfield starts.

Model	Position	Accuracy	F1 Score	AUC	Recall
`xStarted`	FWD	0.9070	0.8480	0.9609	0.8827
`xStarted`	MID	0.8471	0.7633	0.9223	0.7759
`xStarted`	DEF	0.8711	0.8305	0.9388	0.8702

Table 11 Performance Metrics for xStarted by position.

The results in table 11 show that `xStarted` performs strongest when predicting outcomes for forwards, achieving the highest accuracy, F1 score, AUC and recall. Its performance on defenders is also well — with an accuracy of 0.8711, F1 of 0.8305, AUC of 0.9388, and recall of 0.8702 — indicating that the model distinguishes defensive contributions nearly as well as those of forwards. In contrast, midfielders are the most challenging for `xStarted`, with all metrics dipping. Nevertheless, even for the midfield position, the AUC remains robust, suggesting the model still ranks true events ahead of false ones effectively. Overall, `xStarted` demonstrates very strong discriminative power across all field positions, particularly excelling for forwards and defenders.

5.2 Linear Regression Models

The results from the various linear regression models behaved similarly. The more customized a model was for a specific position, the better the model performed. As can be seen in table 12, the model for forwards performs better than the ones for midfielders and defenders. This is because strikers have a greater tendency to score goals, which also makes it easier for the model to predict the number of goals for a striker. The performance metrics for all models can be found in appendix C, table C.1.

Model	**Position**	**MAE**	**RMSE**	$\mathbf{R}^2$
`xGoals`	FWD	0.3739	0.4954	0.0691
`xGoals`	MID	0.1805	0.3374	0.0745
`xGoals`	DEF	0.0645	0.1830	0.0018

Table 12 Performance Metrics for xGoals Model by Position using linear regression.

However, the models prediction of goals differs slightly from the actual number of goals scored, see figure 3. The model does not capture players who actually score two or more goals. The model is centered around a low number of goals, which is usually the case, but fails to take into account players who have a higher chance of scoring more goals.

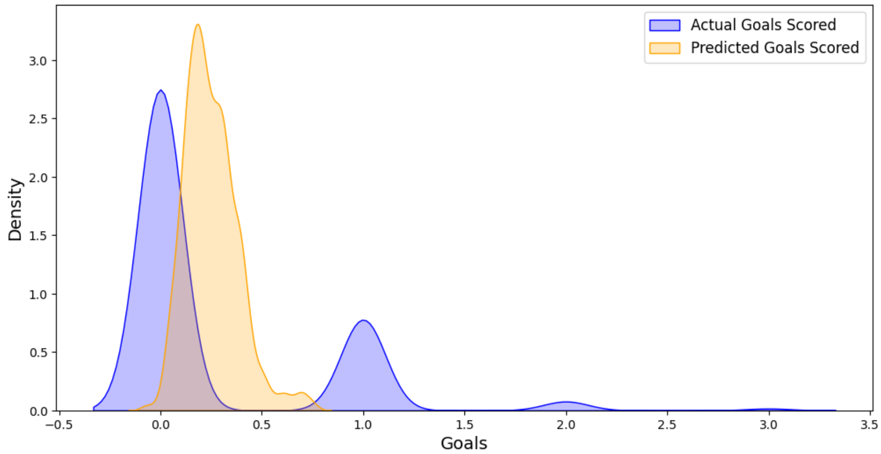

Figure 3 Distribution of actual vs. predicted goals scored for forwards.

Forwards exhibit the greatest spread in actual goals, with a heavy right-hand tail reflecting occasional multi-goal performances. In our density plot, the actual goals curve extends well into one, two, and even three goals per match, whereas the model's predicted density is tightly bunched between roughly 0.1 and 0.4 goals and almost never exceeds 0.5. This tells us that, although the regression has learned the average scoring rate of strikers, it systematically underestimates the likelihood of the high-impact, multi-goal games that characterize top forward performances.

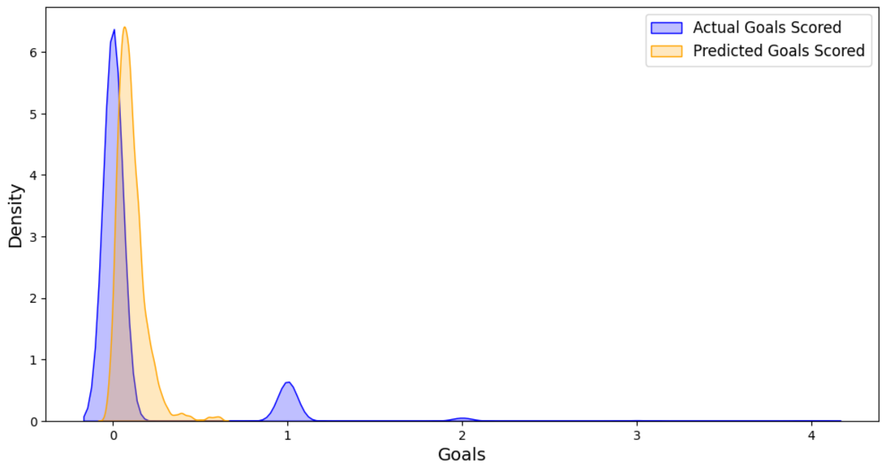

Figure 4 Distribution of actual vs. predicted goals scored for midfielders.

Midfielders produce a narrower range of actual goals than forwards but still show a noticeable secondary hump around one goal and a light tail toward two, see figure 4. The model again concentrates its mass at zero—indeed, it slightly over-peaks there—so it understates the probability that a midfielder will chip in with a goal. In other words, moderate scoring contributions (0.5–2 goals) by midfielders are under-predicted, suggesting missing features that signal when a midfielder pushes forward effectively.

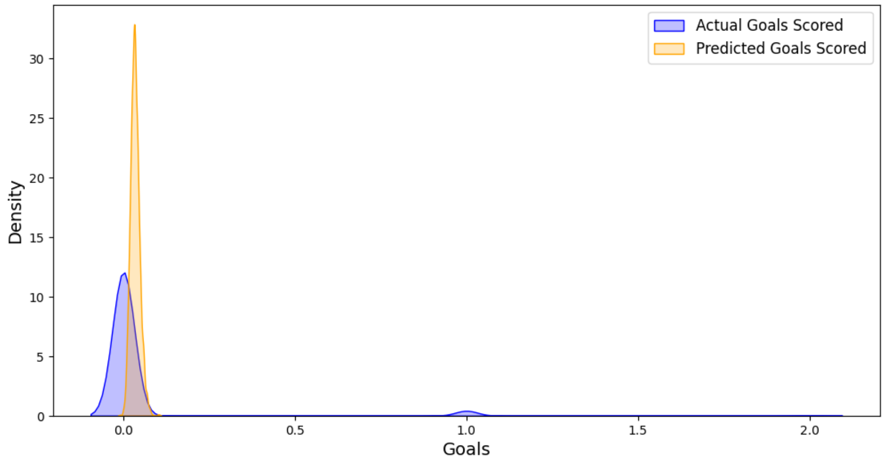

Figure 5 Distribution of actual vs. predicted goals scored for defenders.

Defenders almost always record zero goals, and the model correctly places the bulk of its predicted density at zero, see figure 9. However, the real-world tail of rare one- or two-goal outings is essentially flattened out in the predictions. In practice, this means the model fails to capture the infrequent but important occasions when defenders score—again pointing to an underestimation of low-frequency, high-impact events in its current formulation.

5.3 Explaining XGBoost Predictions using SHAP

Using the model `xGoals_MID` as an example, figures 6 and 7 together paint a picture of which attributes most consistently drive xG predictions for midfielders, and exactly how varying levels of those attributes shift the forecast.

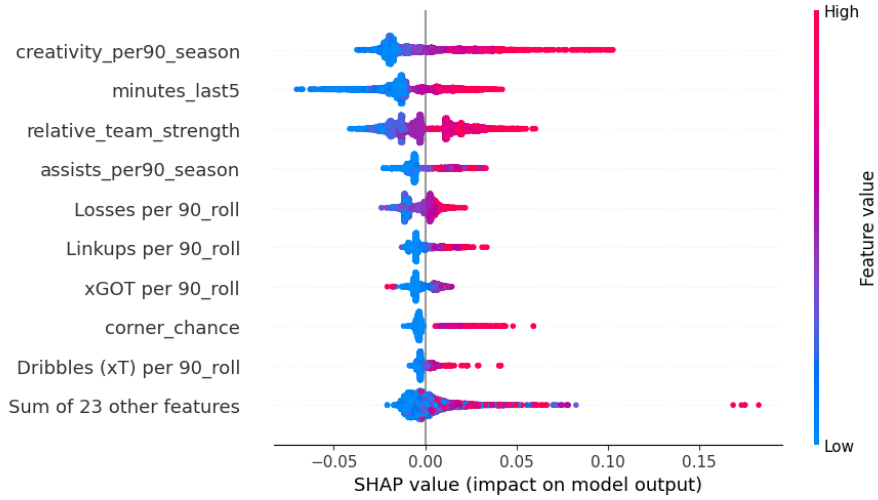

Figure 6 Top ten features for midfielders in terms of `xGoals` based on the mean SHAP values.

In Figure 6, the bar chart ranks features by their average absolute SHAP value, showing that `threat_per90_season` is the dominant factor: midfielders who regularly create dangerous actions (pitch location, shots, dribbles into the box) see the biggest lift in expected goals. Just behind are `relative_team_strength` which captures the quality of the surrounding side compared to your team, and `cumulative_threat`, reflecting sustained attacking involvement over the whole season. The presence of `Linkups_per90_roll` and `goals_per90_season` in the top five underscores that both combination play in the final third and finishing pedigree still materially matter, although to a lesser extent than pure chance creation metrics.

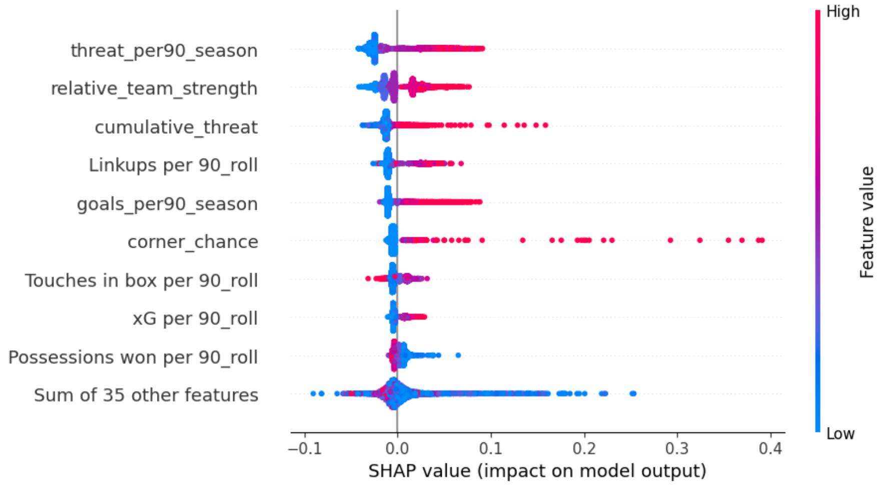

Figure 7 Beeswarm plot over the impact of features for predictions of Assists for midfielders.

Table 13 breaks down the ten most important feature contributions to Chris Wood's predicted xG in gameweek 20 as an example. Overall, these SHAP values sum to Wood's model prediction above the baseline. The largest positive driver is his team's pressing effectiveness over the last five matches (+0.108 xG), closely followed by a the chance of him getting a penalty opportunity (+0.052 xG). Other moderate uplifts come from under-pressure retention and relative team strength. On the negative side, Wood's finishing efficiency roll slightly detracted from his xG (–0.035 xG), as did a lack of deep-run activity (–0.018 xG) and modest carry volumes (–0.012 xG). Together, these feature effects explain why the model forecasted him for approximately + 0.11 xG above its season-long baseline in that week.

Feature	Value	SHAP
high_press_effectiveness_team_avg5	0.1705	+0.1081
penalty_chance	1.0000	+0.0517
Under pressure retention per 90_roll	0.0134	+0.0312
relative_team_strength	0.6667	+0.0282
np_Goals_team_avg5	0.4545	+0.0097
goals_per90_season	0.0222	+0.0164
Passes third (xT) per 90_roll	0.0334	+0.0111
finishing_efficiency_roll	0.4376	–0.0353
Deep runs (xT) per 90_roll	0.0000	–0.0185
Carries (xT) per 90_roll	0.0111	–0.0121

Table 13 Local SHAP feature contributions for Chris Wood in GW 20. Positive values increase his xG forecast above the baseline; negative values decrease it.

5.4 Method Comparison

When comparing the two `xGoals` modeling approaches, simple linear regression and the more complex XGBoost ensemble, we observe a consistent pattern across all positions (see tables 12 and 14, and figure 8). XGBoost generally improves average accuracy, as reflected in lower MAE, but often at the cost of higher variability in its predictions and limited gains in explained variance (R^2). For forwards (FWD), XGBoost significantly reduces the MAE from 0.3739 to 0.2502, showing a notable improvement in average prediction accuracy. However, the RMSE remains nearly unchanged (0.4954 vs. 0.5002), and the R^2 drops slightly from 0.0691 to 0.0511. This suggests that while XGBoost delivers more accurate typical predictions, it does not improve the models ability to explain the variance in forward goal-scoring. A similar trade-off appears with midfielders (MID). XGBoost lowers the MAE from 0.1805 to 0.1182, again indicating better average performance. But this comes with a slightly increased RMSE (0.3374 to 0.3438) and a marginal drop in R^2 from 0.0745 to 0.0609. In effect, XGBoost reduces bias but introduces a bit more variance, especially in occasional misspredictions.

Defenders (DEF), whose goal-scoring is rare and highly variable, show the clearest improvement in MAE, dropping from 0.0645 with linear regression to just 0.0341 with XGBoost. The RMSE remains nearly identical (0.1830 vs. 0.1847), and both models perform similarly in terms of R^2 (0.0018 vs. 0.0023), indicating negligible explanatory power for defender goal output in either case.

Model	Position	MAE	RMSE	R^2
xGoals	FWD	0.2502	0.5002	0.0511
xGoals	MID	0.1182	0.3438	0.0609
xGoals	DEF	0.0341	0.1847	0.0023

Table 14 Performance Metrics for `xGoals` Model by Position using XGBoost.

XGBoost consistently lowers MAE across all positions, making it preferable if the goal is to reduce the average prediction error. However, it does not enhance the models explanatory power (R^2) and may slightly increase the variability of its errors (as seen in RMSE). If interpretability and risk mitigation are important, especially to avoid rare but large errors, linear regression remains a more stable alternative.

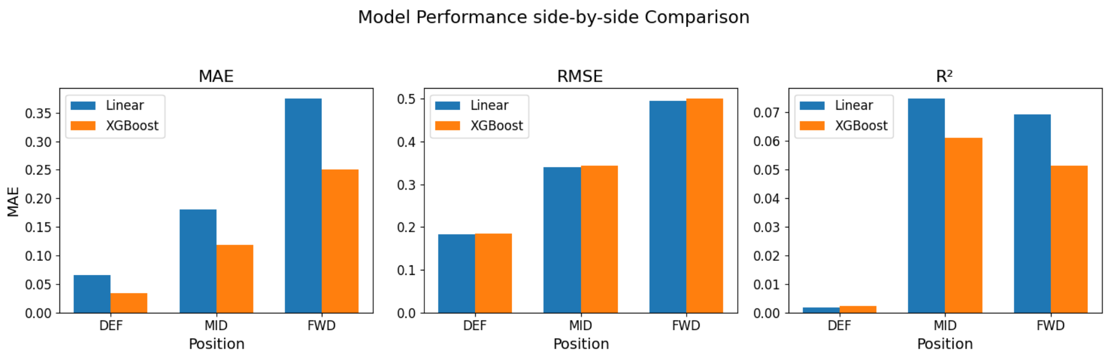

Figure 8 Comparison between the model performances for `xGoals`.

5.4.1 Underfitting

Across all target variables, the comparison of training- and test-set metrics revealed virtually identical MAE, RMSE, and R^2 values for both the linear/logistic regressions and the XGBoost models. This consistency suggests the models exhibit low variance and are therefore not overfitted. However, the modest (and occasionally negative) R^2 scores for certain outcomes indicate the opposite problem: underfitting. In these tasks the models fail to capture the underlying signal, likely because the targets have highly skewed or sparse distributions, influential contextual features are missing and even XGBoost was constrained by limited feature richness rather than algorithmic capacity.

5.4.2 Predicted points

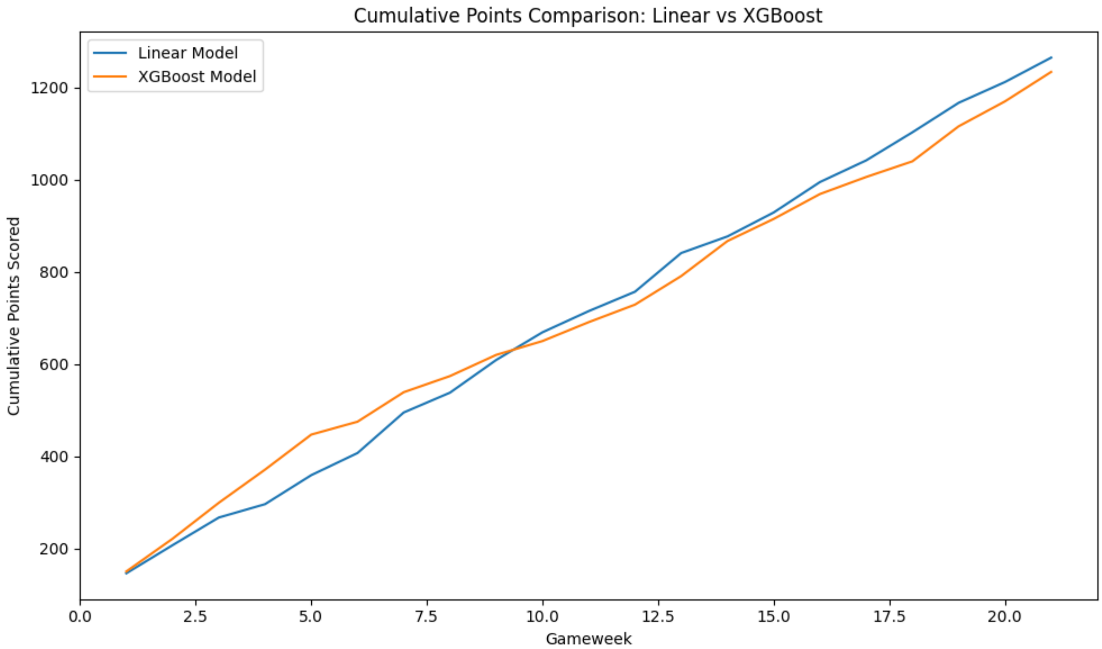

Figure 9 Comparison between simulations done based on two different ML methods.

Looking at the cumulative points scored across the gameweeks, we can see a compelling illustration of the trade offs between the linear regression and XGBoost models. Early on, XGBoost holds a noticeable advantage. Its ability to capture non-linear relationships seems to pay off, as it consistently outperforms the linear regression models in the first 10 gameweeks. This reflects what is said in section 5.4: XGBoost tends to reduce bias and minimizes the average miss, or MAE, by being more flexible in how it learns patterns.

However, as the season progresses, the picture begins to shift. From around gameweek 11 onward, the linear regression model starts to catch up and eventually overtakes XG-Boost. While the XGBoost line dips occasionally, likely due to its tendency to make larger, more erratic mistakes stemming from higher variance, the linear regression models maintains a steadier, more predictable trajectory. This stability is one of the key advantages of linear regression, especially when interpretability and reliability are valued.

By the final gameweek, the linear regression model finishes slightly ahead in total cumulative points, where linear regression tallied a total score of 1293 points and XGBoost 1256 points. The linear regression model would place our AI assistant at the top 12% of all FPL users, if the same point pattern is followed throughout the whole season. Although XGBoost is better at minimizing average error, it does not necessarily translate into better overall performance when large, infrequent errors offset its early gains. Meanwhile, the linear regression models consistency allows it to edge ahead, especially in a setting where total cumulative success is the benchmark.

This comparison highlights the practical implications of the model choice. If the aim is average accuracy and regardless of occasional volatility, XGBoost is a strong choice. But if the priority are a more interpretable, steady-performing model that avoids major prediction pitfalls, the linear regression model proves to be the more dependable option in the long run. To put these results into further perspective, we conducted an additional baseline analysis using a straightforward heuristic of selecting the highest cumulative point scorers from previous gameweeks. This simple cumulative-points heuristic, applied retrospectively from gameweek 2 through gameweek 21, was simulated with the same starting lineup as the top performing team. The baseline heuristic generated a total of 1132 points, considerably below the 1293 points achieved by the linear-model-based optimization approach. This comparison underscores the effectiveness of incorporating predictive analytics and mathematical optimization over purely historical performance-based selection.

5.5 LLM integration

When you open the AI assistant, you will be prompted to enter your team ID, and once that is done your personalized dashboard springs to life: you will see your full squad lineup alongside its current market value and the balance remaining in your transfer bank, as well as how many free transfers you have before penalties apply and a prediction of your teams score for the next gameweek. At a glance, you can also track which of your special "chips" are still available, each one shown in green if it is unused or red once it has been played, so you can plan your transfers and strategize your chip usage with confidence. An overview of the app is shown in figure 10.

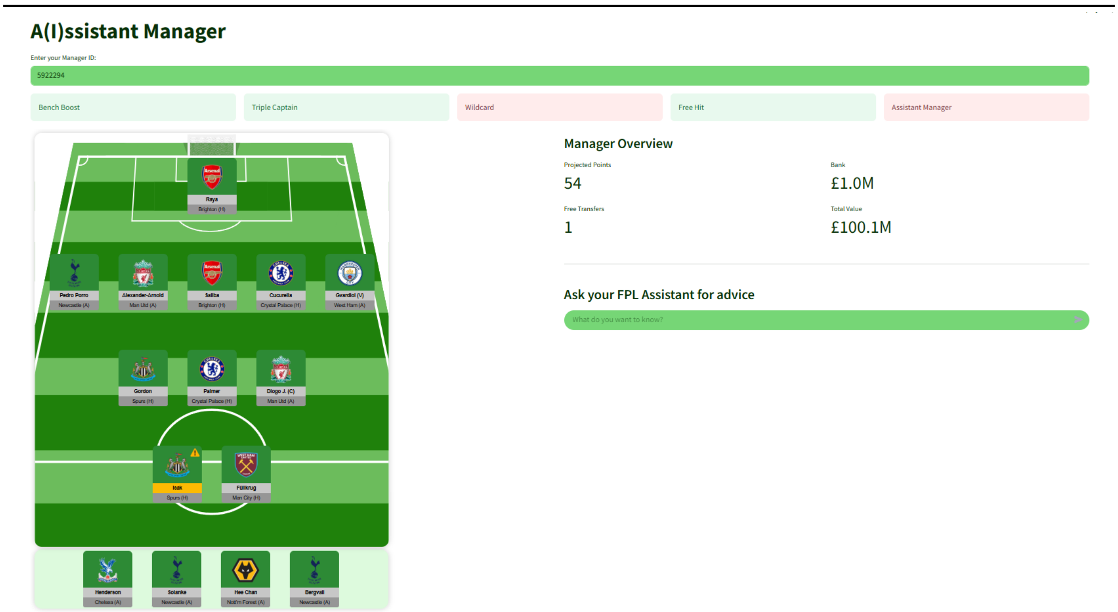

Figure 10 An overview pf the front-end part of the AI-assistant manager and its functions.

The lineup you see mirrors your current FPL squad exactly (see figure 10, arranged with goalkeepers at the top, followed by defenders, midfielders, and forwards. Each player's name is accompanied by your team's emblem, their next opponent, and whether that fixture is A (away) or H (home). To the right of their names, you will also find the markers for captain (C) and vice-captain (V). Below your starting eleven, your bench appears with the same details. If a player is shaded in yellow, like "Isak" in this example, that indicates FPL has flagged them as doubtful for the upcoming match; this status comes directly from FPL's own data, not our predictive model.

Figure 11 Screenshot of the AI-assistant Manager pitch interface.

As illustrated in figure 10, a dedicated input field serves as the interface to our predictive framework. Here, users may submit free-form queries concerning their personal squad, at which point the system retrieves their current roster and bank balance and applies our suite of forecasting models. Rather than exposing raw numerical outputs, the assistant translates its analysis into a structured, explanatory narrative: it identifies which players to remove and which to introduce, quantifies the anticipated point differential relative to alternative options and translates it to explanatory answer which specifies the particular model driving each recommendation, and elucidates the most influential factors underpinning its guidance. See an example in figure 12.

Figure 12 Screenshot of the AI-assistant Manager "Chat Prompt" interface, illustrating how prompts are configured in the front end.

The LLM-powered assistant transforms raw FPL data and complex model outputs into a single, intuitive conversational experience. By entering your manager ID you instantly see your exact squad, and "Ask your FPL Assistant for advice" triggers the behind-the-scenes pipeline: up-to-the-minute API retrieval, ensemble forecasting (linear regression models), and a LLM-style narrative that tells you precisely which transfers or substitutions to make, by how many predicted points you'll gain, and which statistical factors drove the recommendation. This seamless integration ensures that managers not only get optimal, data-backed guidance but also understand the "why" behind every suggestion.

5.5.1 Evaluation

The usability and explainability of the AI tool were evaluated by comparing how useful users found the raw data versus the personalized recommendations provided by the AI assistant. Both groups were given a test team that was identical for all test subjects. Five rounds were simulated, with the test subjects having to make changes (if needed) before each new round. One group received only the raw data, which included predictions, performance metrics and the models meta data. The other group received tailored recommendations from the full-stack AI assistant which is illustrated in figure 10, based on their understanding of what was needed to improve the teams performance. The models used for the evaluation was the linear models. A survey was then conducted where each test subject was given the chance to rate their experience based on some statements. These statements can be found in table 15 for the group using the AI assistant manager and 16 for the group using the raw data.

Subject	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8
1	6	4	7	3	3	6	4	6
2	7	7	7	6	7	6	7	7
3	3	2	4	1	3	5	2	3
4	5	6	6	5	6	7	5	6
5	6	6	7	6	6	7	7	6
6	5	5	2	4	5	6	2	4
Avg.	**5.83**	**6.17**	**6.33**	**5.33**	**6.17**	**6.50**	**6.00**	**6.50**

Table 15 Survey responses from users with access to the full-stack AI assistant manager (n=6). (Scale: 1-7, poor-excellent; Avg. = average per question)

Subject	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8
7	2	2	4	3	5	2	4	3
8	1	1	2	1	3	1	2	2
9	4	3	5	3	4	4	5	5
10	2	3	3	2	3	2	3	3
11	5	4	5	3	5	3	4	4
Avg.	**2.8**	**2.6**	**3.8**	**2.4**	**4.0**	**2.4**	**3.6**	**3.4**

Table 16 Survey responses from users with access only to raw data and predictions (n=5). (Scale: 1-7, poor-excellent; Avg. = average per question)

The survey results reveal a clear distinction between the experiences of users who interacted with an AI assistant and those who only had access to raw data, see figure 13. Subjects 1 through 6, who were supported by the assistant, consistently rated their experience more positively across all aspects evaluated. In contrast, subjects 6 through 11, working solely with raw data, gave lower scores, suggesting a less effective and more challenging experience.

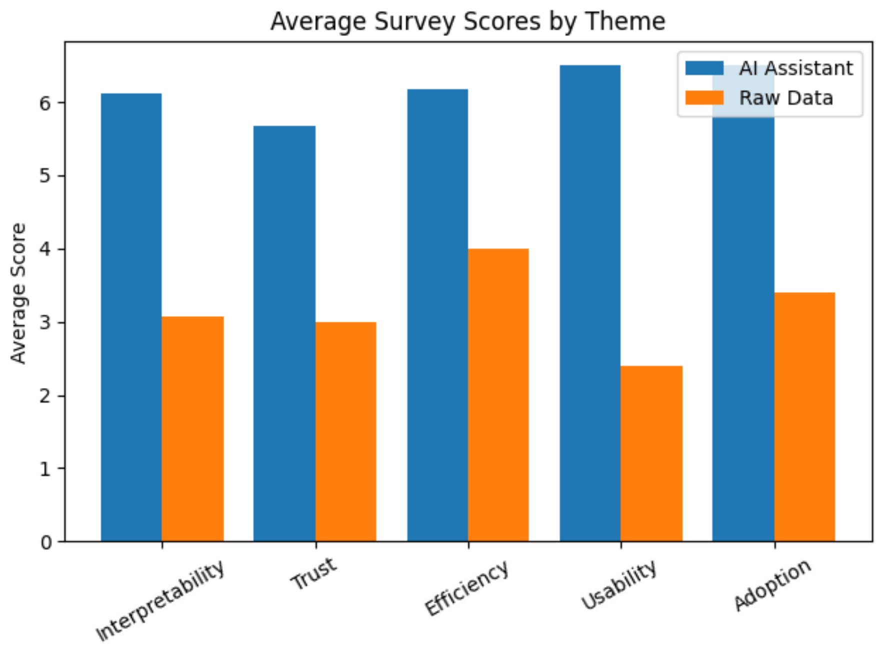

Figure 13 Comparison of survey results when divided into different themes.

One of the most striking differences appears in users understanding of the recommendations. Those with the AI assistant felt it significantly enhanced their grasp of the predictions, with ratings an average rating of 5.83 on the question assessing overall understanding. This effect was even more pronounced when participants were asked whether they understood why certain players were recommended: the assistant supported users again gave high ratings, while those with raw data gave some of the lowest scores across the survey. This highlights the assistants strength in translating complex information into meaningful explanations—something the raw data alone failed to do effectively.

Question 3 followed a similar pattern. Users with the assistant found the information easier to understand, with an average score of 6.33. Meanwhile, the raw-data group struggled to make sense of the same material, with an average score of 3.8. This suggests that the AI assistant not only delivered information but also helped frame it in a way that was accessible and user-friendly.

Trust in the assistant's recommendations showed once again an evident difference. The assistant supported group expressed a higher degree of trust, especially when the assistants advice conflicted with their own intuition. Although both groups were somewhat cautious in this regard, the assistant seemed to provide just enough confidence to shift user perspectives, at least slightly.

Perhaps most practically, the assistant made the decision-making process more efficient. Participants who used the assistant reported that it saved them time. For the raw-data users, however, time savings occurred but not to the same extent, suggesting that the burden of interpretation slowed them down.

The same pattern was observed for both usability and adoption. As expected, users of the assistant found it significantly easier to handle the data when it was presented in a more understandable language. They also said that they would continue to use the tool to a greater extent if it were available, unlike users who only used the raw data, where their results in the survey can instead be interpreted as indicating that there were interesting aspects, but that it was too complex for continued use.

The group that only had access to the raw data quickly noted that interpreting the information was challenging. A key issue was the difficulty of identifying worthwhile players among the vast pool available. Participants said they relied on their football knowledge to filter out players they already knew would not fit their team and focused on traditionally high-performing players to assess their scores more closely. Ultimately, they used the data in much the same way they normally played FPL, namely through a combination of intuition and familiarity.

The second group, which had access to the AI assistant, found that many of the suggested players were already on their radar. However, the AI tool made it easier to directly compare those options, streamlining the decision-making process. While a few players surfaced that they might not have considered otherwise, most were familiar names. The combination of AI insights and their own football knowledge helped participants feel more confident and efficient in finalizing their team choices—something they viewed as a definite advantage. The questions are based on the question ID's in table 7 section 4.5.

6 Discussion

6.1 Performance Metrics Evaluation

The performance metrics presented in table C.1 reflect the challenges of predicting football outcomes using linear regression. Across all target variables and player positions, we observe modest results, with low R^2 values and relatively high error metrics (MAE and RMSE). These outcomes may initially appear underwhelming when compared to theoretical expectations for model performance. However, when contextualized within the inherent complexity of football data, they become more understandable, as mentioned in section 3.7.

Football is a highly dynamic sport influenced by numerous interacting and often unobservable factors. Tactical decisions, player psychology, team cohesion, and random in-game events all contribute to match outcomes, yet are difficult, if not impossible, to capture fully in structured datasets. This high level of noise and variability in the data naturally limits the predictive power of any statistical model, particularly one as constrained as linear regression. In our residual analysis, the largest errors occurred when the model attempted to predict extreme outcomes, such as a strikers unusually high number of expected goals in a single match or a deep-lying midfielders exceptional passing performance. Because these extreme cases are rare in the training set, the linear model tends to regress toward the mean rather than extrapolate correctly, leaving large residuals in the tails of the distribution and dragging down overall R^2.

An additional factor influencing the R^2 scores is the role-specific distribution of the target variables. For example, defenders typically score and assist very infrequently, resulting in a narrow range of values and low variance in their target outputs. In such cases, even small prediction errors can lead to disproportionately low R^2 values because the model is unable to explain much of the limited variation present in the data. This contrasts with forwards, whose match-to-match outputs vary more widely, providing the model with a broader distribution to learn from and explain. As a result, R^2 is expected to be higher for forwards and midfielders in terms of goals and assists, not necessarily because the model performs better for these positions in absolute terms, but because their underlying data allows more variance to be captured and predicted. In this light, low R^2 scores for defenders reflect the statistical structure of the problem more than they do model inadequacy.

Linear regression, while interpretable and computationally efficient, relies on the assumption of linear relationships between input features and target outcomes. In the football domain, such relationships are rarely purely linear. For example, the impact of a forwards shooting position or a midfielders pass completion rate on expected goals or assists is likely to be non-linear and highly contextual. These limitations contribute to the observed performance, especially in the R^2 scores. Nonetheless, these results should not be interpreted as a failure. Rather, they highlight the difficulty of modeling such a complex domain and point to potential avenues for future work. More sophisticated modeling techniques, could capture non-linear interactions and improve predictive performance, although our experiments show that this is not the case. Furthermore, enriching the dataset with additional contextual features (e.g possession based features and tracking data features) could provide the model with deeper insight into the game dynamics.

Our comparative results, where linear regression sometimes edges out XGBoost in R^2 yet falls short in MAE (see figure 9 in section 5.4), highlight a dual question: Is a linear model simply too simplistic to capture the true intricacies of FPL point generation, or are we hamstrung by feeding both models an incomplete set of features? The fact that XGBoost consistently produces lower MAE suggests that its nonlinear splits and regularization indeed lower average prediction errors, picking up on interactions that a purely additive model cannot. Yet, when a linear regression, unburdened by complexity penalties, still explains a slightly larger share of the total variance for midfielders and forwards, it hints that a surprising amount of the predictable signal in our dataset behaves in a roughly linear fashion. In essence, while the linear form underfits subtle, interacting effects it nevertheless "soaks up" most of the mid-range variation because basic features (`selected_last_round`, `relative_team_strength`, `cumulative_threat`, etc.) dominate the bulk of the distribution.

At the same time, both models demonstrate marked improvements when we enrich our feature space with engineered variables, showing that neither linear nor non-linear algorithms can create predictive power from missing information. When key signals are absent, even XGBoost's complex tree splits cannot fill the gap. The linear model's occasional R^2 edge simply reflects that much of the predictive signal in the raw FPL API data is essentially linear. Yet the nearly identical train- and test-set errors, paired with poor R^2 scores for targets such as expected yellow cards and bonus points, point to underfitting driven by high bias. Two issues appear decisive: Firstly, these targets are sparse and heavy-tailed with most players record zero events while a small minority account for nearly all positives, so a model centered on the mean misses rare but crucial cases. Secondly, the current feature set omits contextual variables, underscoring the need for richer feature engineering or more specialised architectures.

While the models used in this study yielded limited predictive accuracy, the results align with expectations given the complexity of the problem. They serve as a useful foundation for further exploration and refinement, and underscore the importance of selecting appropriate modeling strategies for complex real-world phenomena like football.

6.2 Underestimation of high performing players

All of our models had a similar behavior with difficulty to capture players that over perform during a match based on the models purpose. The models tend to underestimate those rare matches in which players significantly outperform their typically low scoring averages, see figures 3, 4 and 9 in section 5.2. Take `xGoals` as an example, scoring in any given Premier League match is uncommon, and netting multiple goals is even rarer. In fact, the 2023/24 season saw an average of 3.28 goals per match, the highest on record since the season 1964/65 [81], but these goals are shared among at least 20 outfield players, with forwards naturally shouldering the bulk of the scoring. Given this distribution, it is unsurprising that our models gravitate toward predicting low scoring probabilities across the board and rarely assign high probabilities to multiple-goal performances.

However, we realized early on that this inherent bias does not have to undermine our overall approach. Rather than relying on raw probability outputs alone, we implemented a threshold-based scheme. As illustrated in figure 3 section 5.2, the model's continuous predictions resemble a normal distribution centered around 0.3 for `xGoals_FWD`. We leveraged this by defining discrete cutoff points: predictions above a certain value indicate a “chance of a goal,” while lower values denote a “low chance.” We then introduced additional tiers, such as “high chance” and “very low chance”, to capture finer gradations. When integrated into our LLM framework, these thresholds enable the system to translate numerical probabilities into descriptive categories instead of presenting a raw percentage, as shown in figure 12 section 5.5. This thresholding strategy helps mitigate the models tendency to under represent rare, high-scoring events while preserving the interpretability of the output.

6.3 Comments on the AI Assistant’s Performance

We let our AI assistant pick the squad and handle substitutions based on our predictions and logic on the Premier League season of 2024-25. Since the season was still ongoing during the project, the total test span was limited to the 21 first gameweeks of the season. During this span, the respective models gained average point values of 61.6 and 59.8 points per week with the linear method proving to yield the most points. If we assume that this mean would hold up for the rest of the season, it would place our AI assistant in the top 12% of all FPL managers. However, it is important to note that this does not fairly represent the performance of our AI assistant due to the following:

1. Our assistant did not take *chips* into consideration. There are multiple chips you can use 1 or 2 times a season and which are vital tools for top prospect managers to effectively maximize their points yield. See table 2 in section 2.2 for more detailed information.

2. The assistant handled the team through gameweek 1-21 which involved no *double gameweeks*. Double gameweeks happen mostly in the end of the season, and occurs due to teams having international competitions or cup fixtures which interfere

with the Premier League schedule. As a result, some games will be postponed to later gameweeks resulting in some teams having two matches in one gameweek. If a manager takes this into consideration they will have a good opportunity to maximize their points total for that gameweek. This can also be proved statistically given that the average points yield for all managers during these gameweeks often are much higher than the average gameweek.

3. All managers have access to *current information* on injuries, suspensions and other factors that play a part in if a player starts or plays minutes on the pitch. Given that our AI assistant manager did not have access to that information, it had to guess wether a players starts or not based on the predictive modeling. Even though our `xStarted` models had an average accuracy of 0.875, it will have a major effect on its final performance. Considering the thought out usage of the AI assistant, with a human-in-the-loop element, this factor will be minimized, if not eliminated entirely, when you have a human manager alongside the AI assistant manager figuring out the squad choices.

4. The AI assistant did not take long-term planning into account. Planning for future gameweeks is a crucial element that must be taken into account when playing FPL competitively. Knowing future matchups, double and blank gameweeks, international competitions etc. is an important factor in choosing players. Even though they might struggle in one gameweek, they might have a really easy stretch of games coming up. Which makes it worth holding on to them since you only have a limited amount of free substitutions each week.

Based on these four factors and the actual performance of the AI assistant, the resulting points yield was arguably much better than what the results show and our conclusion is three-fold:

- The AI assistant manager is, by a good margin, better than the average human manager at predicting player performances and thus, getting a higher points yield in the FPL. It was also shown to outperform a baseline heuristic approach of substituting in top point scorers each week.

- The AI assistant manager performs better in collaboration with the human manager, combining the predictive accuracy of the AI assistant manager and the up-to-date knowledge of the human manager, a better performance will be achieved.
- The AI assistant manager could be improved upon to get better results if incorporating factors that this thesis project did not take into account. Such as point sources, chips-usage optimization, long-term planning, return-on-investment strategies and more up-to-date data from multiple sources that covers injuries, cup and international competitions etc.

Additionally, our evaluation of the explainable layer of our AI assistant further shows that the explainability and transparency factor will further increase the final performance of managers using the tool.

6.4 Broader Implications of the Method

The AI assistant, which couples transparent predictive models with an explanatory, conversational LLM interface, exemplifies a decision-support architecture with applicability beyond fantasy sports. As we already highlighted, sectors from banking to manufacturing are confronting vast, complex datasets and regulatory pressures that demand both accuracy and interpretability. By transforming straightforward linear model based predictions to translated conversational rationales through an LLM, our method shows that it is possible to deliver reliable forecasts without leaving users in the dark, a capability valuable in other problem settings as well. [22, 23, 26].

Moreover, the natural-language interface developed allows users to articulate specific constraints and preferences like “I need three midfielders under 6 millions” and immediately see how those trade-offs affect expected points. This human-in-the-loop paradigm mirrors recent advances in scheduling and resource-allocation systems, where conversational prompts steer optimization engines to generate and refine solutions in domains as varied as hospital patient triage or workforce rostering [24, 25]. By wrapping solvers in chat, the barrier can be lowered, to be customized for non-technical stakeholders, whether they manage a fantasy roster or allocate clinical staff across shifts.

The modular pipeline described in section 3.9 is eminently reusable. For example, in peer-to-peer lending, an XGBoost+SHAP pipeline originally built for corporate risk analysis was adapted to retail credit marketplaces with minimal adaptation, boosting both performance and transparency [22, 23]. Likewise, a tele-triage system could employ a simple logistic regression to handle straightforward cases and invoke a more complex ML layer only when uncertainty is high, all while an LLM articulates its reasoning to patients and clinicians in clear terms [24]. The same architectural pattern that helps a fantasy manager optimize a squad under budget constraints, forecast fixture difficulty and understand the why behind each suggestion can be redeployed across industries. An explainable, LLM-mediated decision-support engine offers a universal framework for combining predictive power with human-centered transparency.

7 Conclusions

To address the research questions, we trained an transparent, position-specific linear regression models on both basic FPL API features and engineered team- and player-level metrics, establishing clear, interpretable links between each statistic and the predicted contribution to points. The linear regression model were benchmarked against the non-linear XGBoost alternatives. This model were tuned via grid search and regularized to prevent overfitting, and extracts feature importance explanations using SHAP values to reveal which factors most strongly drive predictions. Both models had their advantages, but linear regression performed better in terms of pure predicted points vs. actual points. The linear regression was integrated into a conversational LLM layer that translates prediction scores into natural-language rationales for each recommendation, grounding each transfer or lineup suggestion in human-readable reasoning.

The linear regression model was simulated for the first 21 rounds of the 2024/25 season and players were selected for each round, it had accumulated 1,293 points while the XGB models gained similar results, accumulating 1,256 points during the same test period. The average per round placed the model among the top 12% of users. Across these retrospective tests on the 2024/25 season, this hybrid approach provided end users with transparent insight into why each player was selected. When evaluated with active FPL managers, our survey found that users of the AI assistant scored significantly higher on measures of interpretability, trust and confidence in decision making compared to those who only had access to the raw data with predictions.

By easily audited linear models with an LLM-mediated narrative interface, we demonstrate a practical path toward decision support tools that are both data-driven and genuinely transparent. Users are no longer faced with a “black box” but are empowered to see exactly which variables, be it recent expected-goals over-performance or rising pressing intensity, were most influential in each recommendation. This clarity not only fosters trust but also enables managers to learn and refine their own strategies over time. Nonetheless, certain limitations remain. Our models currently omit rare events, like penalties and own goals, and do not yet integrate the strategic chip mechanics of FPL, which may affect optimal timing considerations. Moreover, while our evaluation shows promising gains in user trust and understanding, further work is needed to assess real–world leaderboard impact over multiple seasons and across diverse user skill levels.

This thesis also provides a proof-of-concept for an explainable, LLM-augmented AI assistant in fantasy football. The tool balances competitive predictive performance with human-centric clarity. By showing how ML, optimization, and natural-language generation can be woven into a seamless workflow, we offer a blueprint for transparent decision support systems in sports and beyond.

7.1 Future Work

To further enhance the proposed methodology, future studies on the optimization of the FPL strategy should consider incorporating additional factors. This could involve utilizing a wider range of data sources, such as scraping news sites and social media, to provide the AI assistant manager with more comprehensive and up-to-date information. Another important area for future work is chip optimization, which would further refine the projections. Closely related to this is the need for long-term planning. A strategic AI assistant manager should be able to plan beyond a single gameweek, accounting for anticipated player value changes, upcoming double or blank gameweeks as well as optimal use of free transfers. One approach to this thesis' current constraints could be to leverage methods like dynamic programming or rolling-horizon planning. Dynamic programming would allow the AI manager to recursively evaluate the best future gameweek scenarios by breaking the optimization problem down into overlapping subproblems. These subproblems could, for instance, be based on potential player returns, fixture difficulties and anticipated value fluctuations. Employing this sort of strategy could result in a more comprehensive strategy that extends beyond immediate transfers. Alternatively (or simultaneously), rolling-horizon planning would enable recalibration of decisions weekly by continuously updating forecasts as new data becomes available, ensuring flexibility in response to changing conditions such as injuries, suspensions, or fixture rescheduling. Optimizing chip strategies such as Wildcard, Bench Boost, Free Hit, or Triple Captain would significantly benefit from these approaches, as they require assessing multiple future scenarios, including double or blank gameweeks. These methods could systematically quantify the expected benefit of deploying chips at various stages of the season, ultimately improving the assistant's long-term strategic performance in FPL.

Beyond algorithmic optimization, feature engineering is the most promising avenue for further progress. Incorporating richer, more varied data should strengthen predictive performance by capturing signals overlooked in the present study. Contextual variables tied to defenders' bonus-point allocation and yellow-card incidence across all positions were areas where our models struggled and could be improved upon in future work. A broader feature set in general would also make it feasible to include FPL scoring events omitted here, such as red cards, penalty saves, and missed penalties, thereby widening the model's scope and boosting its predictive power. To strengthen the dataset, we recommend three additions. First, include features that capture more of the team-level characteristics such as tactical styles and head-2-head match history. Second, add variables derived from possession chains to describe the build-up context leading to each event. Third, integrate tracking-data metrics to provide richer spatial and temporal context. Together, these enhancements should give the models a more complete picture of on-pitch dynamics and improve predictive accuracy.

Additionally, future research building on the proposed AI assistant manager should experiment with a broader set of methods and machine learning models, particularly in the context of XAI. Exploring different model architectures such as neural networks, ensemble methods or hybrid approaches, could potentially enhance predictive accuracy and the tool's adaptability to complex decision-making scenarios. Furthermore, incorporating recent advancements in XAI techniques, like counterfactual explanations or feature attribution methods beyond SHAP, may improve the interpretability and transparency of recommendations provided to users. Comparative studies assessing how various models and explanation strategies impact user trust, understanding, and decision quality would also contribute valuable insights. Ultimately, such research could help refine the balance between predictive performance and explainability, ensuring the AI assistant manager remains both powerful and user-friendly for a wide range of FPL participants.

References

[1] J. Crelin, “Daily fantasy sports and gambling: Overview,” 2021.

[2] Mordor Intelligence, “Fantasy sports — market share analysis, industry trends & statistics: Growth forecasts (2025–2030),” https://www.mordorintelligence.com/industry-reports/fantasy-sports-market, Mar. 2025, accessed 17 May 2025.

[3] L. Wilkins, “A bibliometric analysis of fantasy sports research,” *Entertainment Computing*, vol. 48, p. 100613, 2024.

[4] M. B. Haugh and R. Singal, “How to play fantasy sports strategically (and win),” *Management Science*, vol. 67, no. 1, pp. 72–92, 2021.

[5] M. Buser, H. Woratschek, and B. D. Ridpath, “Gamification through fantasy sports–empirical findings from professional sport leagues,” *Sport, Business and Management: An International Journal*, vol. 11, no. 5, pp. 575–597, 2021.

[6] J. Campbell, “Artificial intelligence could transform football: So what might the future look like?” *The New York Times*, January 2025, accessed: 2025-02-26. [Online]. Available: https://www.nytimes.com/athletic/5954943/2025/01/08/artificial-intelligence-could-transform-football-so-what-might-the-future-look-like/

[7] A. J. Karg and H. McDonald, “Fantasy sport participation as a complement to traditional sport consumption,” *Sport Management Review*, vol. 14, no. 4, pp. 327–346, 2011.

[8] T. Nesbit and K. King, “The impact of fantasy sports on television viewership,” *Journal of Media Economics*, vol. 23, no. 1, pp. 24–41, Jan. 2010.

[9] P. League, “Fantasy premier league,” https://fantasy.premierleague.com/, 2025, accessed: 2025-01-24.

[10] L. Campbell-Guthrie. (2025, May) How fantasy premier league became a global obsession. Published 15 May 2025. Accessed 17 May 2025. Paywalled. [Online]. Available: https://www.ft.com/content/f822a11c-1dde-4678-9d7d-10921bb81013

[11] K. Chouaten, C. Rodriguez Rivero, F. Nack, and M. Reckers, “Unlocking high-value football fans: unsupervised machine learning for customer segmentation and lifetime value,” *Frontiers in Sports and Active Living*, vol. 6, p. 1362489, 2024.

[12] T. Matthews, S. Ramchurn, and G. Chalkiadakis, "Competing with humans at fantasy football: Team formation in large partially-observable domains," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 26, no. 1, 2012, pp. 1394–1400.

[13] D. Frees, P. Ravella, and C. Zhang, "Deep learning and transfer learning architectures for english premier league player performance forecasting," *arXiv preprint arXiv:2405.02412*, 2024.

[14] S. Wolf, M. Schmitt, and T. Pawlowski, "A football player rating system," *Journal of Sports Analytics*, vol. 6, no. 3, pp. 157–165, 2020. [Online]. Available: https://doi.org/10.3233/JSA-200411

[15] The Alan Turing Institute. (2023) Airsenal: Fantasy football meets machine learning. Accessed: 2025-04-02. [Online]. Available: https://www.turing.ac.uk/news/airsenal

[16] A. S. Lombu, I. V. Paputungan, and C. K. Dewa, "Predicting fantasy premier league points using convolutional neural network and long short term memory," *Jurnal Teknik Informatika (JUTIF)*, vol. 5, no. 1, pp. 263–272, February 2024.

[17] M. Bangdiwala, R. Choudhari, A. Hegde, and A. Salunke, "Using ml models to predict points in fantasy premier league," in *2022 2nd Asian Conference on Innovation in Technology (ASIANCON)*. IEEE, August 2022, pp. 1–6.

[18] V. Venter and J. H. van Vuuren, "An optimisation approach towards soccer fantasy premier league team selection," *ORiON*, vol. 40, no. 1, pp. 69–107, 2024.

[19] P. Linardatos, V. Papastefanopoulos, and S. Kotsiantis, "Explainable ai: A review of machine learning interpretability methods," *Entropy*, vol. 23, no. 1, p. 18, 2020.

[20] International Data Corporation (IDC), "Worldwide Spending on Digital Transformation Is Forecast to Reach Almost $4 Trillion by 2027, According to New IDC Spending Guide," International Data Corporation, Needham, MA, Press Release prUS52305724, May 2024. [Online]. Available: https://my.idc.com/getdoc.jsp?containerId=prUS52305724

[21] Spherical Insights. (2022) Sports analytics market report. Accessed: 2025-05-19. [Online]. Available: https://www.sphericalinsights.com/reports/sports-analytics-market

[22] M. A. Onari, M. J. Rezaee, M. Saberi, and M. S. Nobile, "An explainable data-driven decision support framework for strategic customer development," *Knowledge-Based Systems*, vol. 295, p. 111761, 2024.

[23] N. Bussmann, P. Giudici, D. Marinelli, and J. Papenbrock, "Explainable ai in fintech risk management," *Frontiers in Artificial Intelligence*, vol. 3, p. 26, 2020.

[24] A. Kasaie and S. Rajendran, "Integrating machine learning algorithms and explainable artificial intelligence approach for predicting patient unpunctuality in psychiatric clinics," *Healthcare Analytics*, vol. 4, p. 100242, 2023.

[25] C. C. Yang, "Explainable artificial intelligence for predictive modeling in healthcare," *Journal of healthcare informatics research*, vol. 6, no. 2, pp. 228–239, 2022.

[26] I. Ahmed, G. Jeon, and F. Piccialli, "From artificial intelligence to explainable artificial intelligence in industry 4.0: a survey on what, how, and where," *IEEE Transactions on Industrial Informatics*, vol. 18, no. 8, pp. 5031–5042, 2022.

[27] A. Joseph, N. E. Fenton, and M. Neil, "Predicting football results using bayesian nets and other machine learning techniques," *Knowledge-Based Systems*, vol. 19, no. 7, pp. 544–553, 2006.

[28] V. Barnett and S. Hilditch, "The effect of an artificial pitch surface on home team performance in football (soccer)," *Journal of the Royal Statistical Society Series A: Statistics in Society*, vol. 156, no. 1, pp. 39–50, 1993.

[29] J. Ensum, R. Pollard, and S. Taylor, "Applications of logistic regression to shots at goal in association football: Calculation of shot probabilities, quantification of factors and player/team," *Journal of Sports Sciences*, vol. 22, no. 6, pp. 500–20, 2004.

[30] A. Rathke, "An examination of expected goals and shot efficiency in soccer," *Journal of Human Sport and Exercise*, vol. 12, no. 2, pp. 514–529, 2017.

[31] B. Ćwiklinski, A. Giełczyk, and M. Choraś, "Who will score? a machine learning approach to supporting football team building and transfers," *Entropy*, vol. 23, no. 1, p. 90, 2021.

[32] M. Roccetti, F. Berveglieri, and G. Cappiello, "Football data analysis: The predictive power of expected goals (xg)," *ResearchGate Preprint*, 2024.

[33] C. M. F. C. M. Rosli, M. Z. Saringat, N. Razali, and A. Mustapha, "A comparative study of data mining techniques on football match prediction," in *Journal of Physics: Conference Series*, vol. 1020, no. 1. IOP Publishing, 2018, p. 012003.

[34] I. Behravan and S. M. Razavi, "A novel machine learning method for estimating football players' value in the transfer market," *Soft Computing*, vol. 25, no. 3, pp. 2499–2511, 2021.

[35] F. Bonomo, G. Duran, and J. Marenco, “Mathematical programming as a tool for virtual soccer coaches: a case study of a fantasy sport game,” *International Transactions in Operational Research*, vol. 21, no. 3, pp. 399–414, 2014.

[36] A. Becker and X. A. Sun, “An analytical approach for fantasy football draft and lineup management,” *Journal of Quantitative Analysis in Sports*, vol. 12, no. 1, pp. 17–30, 2016.

[37] G. Papageorgiou, V. Sarlis, and C. Tjortjis, “An innovative method for accurate nba player performance forecasting and line-up optimization in daily fantasy sports,” *International Journal of Data Science and Analytics*, pp. 1–24, 2024.

[38] R. Beal, T. J. Norman, and S. D. Ramchurn, “Optimising daily fantasy sports teams with artificial intelligence,” *International Journal of Computer Science in Sport*, vol. 19, no. 2, pp. 21–35, 2020.

[39] R. Whittle. (2016) Towards an initial understanding of linear programming. Accessed: 2025-05-19. [Online]. Available: https://www.firstpr.com.au/linear-programming/

[40] V. Veluru, T. Xiao, S. Addagudi, S. Kumar, and G. Mohanraj, “Machine learning optimization model to predict fantasy basketball teams,” in *2024 International Conference on Computing and Data Science (ICCDS)*. IEEE, 2024, pp. 1–4.

[41] A. Bhattacharya, *Applied Machine Learning Explainability Techniques: Make ML models explainable and trustworthy for practical applications using LIME, SHAP, and more*. Packt Publishing Ltd, 2022.

[42] P. Pokharel, A. Timalsina, S. Panday, B. Acharya, and P. Campus, “Fantasy premier league-performance prediction,” 10 2022.

[43] S. Moustakidis, S. Plakias, C. Kokkotis, T. Tsatalas, and D. Tsaopoulos, “Predicting football team performance with explainable ai: leveraging shap to identify key team-level performance metrics,” *Future Internet*, vol. 15, no. 5, p. 174, 2023.

[44] A. Procopiou and A. Piki, “The 12th player: Explainable artificial intelligence (xai) in football: Conceptualisation, applications, challenges and future directions,” in *Proceedings of the 11th International Conference on Sport Sciences Research and Technology Support*, vol. 1. Science and Technology Publications, Lda, 2023, pp. 213–220.

[45] E. Mariotti, J. M. Alonso, and A. Gatt, “Towards harnessing natural language generation to explain black-box models,” in *2nd Workshop on interactive natural language technology for explainable artificial intelligence*, 2020, pp. 22–27.

[46] A. Bilal, D. Ebert, and B. Lin, "Llms for explainable ai: A comprehensive survey," *arXiv preprint arXiv:2504.00125*, 2025.

[47] A. Pang, H. Jang, and S. Fang, "Generating descriptive explanations of machine learning models using llm," in *2024 IEEE International Conference on Big Data (BigData)*, 2024, pp. 5369–5374.

[48] H. Santos and A. Khalil, "Unleashing the potential of llm in ml: Techniques for fine-tuning, adaptation, and practical deployment with chatgpt," *Baltic Multidisciplinary Journal*, vol. 2, no. 2, pp. 179–184, 2024.

[49] A. Cook and O. Karakuş, "Llm-commentator: Novel fine-tuning strategies of large language models for automatic commentary generation using football event data," *Knowledge-Based Systems*, vol. 300, p. 112219, 2024.

[50] D. Maulud and A. M. Abdulazeez, "A review on linear regression comprehensive in machine learning," *Journal of applied science and technology trends*, vol. 1, no. 2, pp. 140–147, 2020.

[51] K. F. Nimon and F. L. Oswald, "Understanding the results of multiple linear regression: Beyond standardized regression coefficients," *Organizational research methods*, vol. 16, no. 4, pp. 650–674, 2013.

[52] A. T. Yigit, B. Samak, and T. Kaya, "An xgboost-lasso ensemble modeling approach to football player value assessment," *Journal of Intelligent & Fuzzy Systems*, vol. 39, no. 5, pp. 6303–6314, 2020.

[53] Z. Li, G. Wang, S. Gao, L. Xu, and S. Su, "Research on prediction of football match results based on xgboost and lstm," in *2024 14th International Conference on Information Technology in Medicine and Education (ITME)*. IEEE, 2024, pp. 769–772.

[54] O. Sagi and L. Rokach, "Approximating xgboost with an interpretable decision tree," *Information Sciences*, vol. 572, pp. 522–542, 2021.

[55] J. Bergstra and Y. Bengio, "Random search for hyper-parameter optimization," *The journal of machine learning research*, vol. 13, no. 1, pp. 281–305, 2012.

[56] P. Probst, A.-L. Boulesteix, and B. Bischl, "Tunability: Importance of hyperparameters of machine learning algorithms," *Journal of Machine Learning Research*, vol. 20, no. 53, pp. 1–32, 2019.

[57] M. Cavus and P. Biecek, "Explainable expected goal models for performance analysis in football analytics," in *2022 ieee 9th international conference on data Science and advanced analytics (DSAA)*. IEEE, 2022, pp. 1–9.

[58] K. Lawson. (2023) Regretting regression: Arsenal, spurs, and the limits of regression analysis in football. Accessed: 2025-05-19. [Online]. Available: https://statsbomb.com/articles/soccer/regretting-regression-arsenal-spurs-and-the-limits-of-regression-analysis-in-football/

[59] R. Nau. (2020) What's a good value for r-squared? Accessed: 2025-05-19. [Online]. Available: https://people.duke.edu/~rnau/rsquared.htm

[60] S. Zhang, C. Zhang, and Q. Yang, "Data preparation for data mining," *Applied Artificial Intelligence*, vol. 17, no. 5–6, pp. 375–381, 2003.

[61] S. García, J. Luengo, F. Herrera *et al.*, *Data preprocessing in data mining*. Springer, 2015, vol. 72.

[62] S. H. Simpson, "Creating a data analysis plan: What to consider when choosing statistics for a study," *The Canadian journal of hospital pharmacy*, vol. 68, no. 4, p. 311, 2015.

[63] H. D. Vinod, "Matrix algebra topics in statistics and economics using r," in *Handbook of Statistics*. Elsevier, 2014, vol. 32, pp. 143–176.

[64] P. J. M. Ali, R. H. Faraj, E. Koya, P. J. M. Ali, and R. H. Faraj, "Data normalization and standardization: a technical report," *Mach Learn Tech Rep*, vol. 1, no. 1, pp. 1–6, 2014.

[65] J.-M. Jo, "Effectiveness of normalization pre-processing of big data to the machine learning performance," *The Journal of the Korea institute of electronic communication sciences*, vol. 14, no. 3, pp. 547–552, 2019.

[66] H. S. Obaid, S. A. Dheyab, and S. S. Sabry, "The impact of data pre-processing techniques and dimensionality reduction on the accuracy of machine learning," in *2019 9th annual information technology, electromechanical engineering and microelectronics conference (iemecon)*. IEEE, 2019, pp. 279–283.

[67] C. Lawless, J. Schoeffer, L. Le, K. Rowan, S. Sen, C. St. Hill, J. Suh, and B. Sarrafzadeh, ""i want it that way": Enabling interactive decision support using large language models and constraint programming," *ACM Transactions on Interactive Intelligent Systems*, vol. 14, no. 3, pp. 1–33, 2024.

[68] V. Anand, "FPL Historical Dataset," Retrieved April 2025 from https://github.com/vaastav/Fantasy-Premier-League/, 2022.

[69] Premier League. (2025) Premier league statistics. Accessed: 2025-04-02. [Online]. Available: https://www.premierleague.com/stats/

[70] J. R. Landers and B. Duperrouzel, “Machine learning approaches to competing in fantasy leagues for the nfl,” *IEEE Transactions on Games*, vol. 11, no. 2, pp. 159–172, 2018.

[71] A. Baughman, M. Forester, J. Powell, E. Morales, S. McPartlin, and D. Bohm, “Deep artificial intelligence for fantasy football language understanding,” *arXiv preprint arXiv:2111.02874*, 2021. [Online]. Available: https://doi.org/10.48550/arXiv.2111.02874

[72] A. Baughman, S. Hammer, G. Cannon, M. Forster, J. Powell, C. Jason, and S. Gudimetla, “Deep artificial intelligence for fantasy football multimodal media understanding,” *arXiv preprint*, 2023.

[73] M. Bhatnagar, “Fancric: Multi-agentic framework for crafting fantasy 11 cricket teams,” *arXiv preprint arXiv:2410.01307*, 2024. [Online]. Available: https://doi.org/10.48550/arXiv.2410.01307

[74] J. S. Cramer, “The origins of logistic regression,” Tinbergen Institute, Technical Report, 2002. [Online]. Available: https://doi.org/10.2139/ssrn.360300

[75] K. P. Burnham, D. R. Anderson, K. P. Burnham, and D. R. Anderson, *Practical use of the information-theoretic approach*. Springer, 1998.

[76] Google AI. (2025) Google ai: Tools, models, and resources. Accessed: 2025-05-19. [Online]. Available: https://ai.google.dev/

[77] H. Naveed, A. U. Khan, S. Qiu, M. Saqib, S. Anwar, M. Usman, and A. Mian, “A comprehensive overview of large language models,” *arXiv preprint arXiv:2307.06435*, 2023. [Online]. Available: https://doi.org/10.48550/arXiv.2307.06435

[78] J. W. Creswell and C. N. Poth, *Qualitative inquiry and research design: Choosing among five approaches*. Sage publications, 2016.

[79] S. Gregor and A. R. Hevner, “Positioning and presenting design science research for maximum impact,” *MIS quarterly*, pp. 337–355, 2013.

[80] A. R. Hevner, S. T. March, J. Park, and S. Ram, “Design science in information systems research,” *MIS quarterly*, pp. 75–105, 2004.

[81] WorldFootball.net. (2025) Premier league statistics. Accessed: 2025-05-19. [Online]. Available: https://www.worldfootball.net/stats/eng-premier-league/1/

Appendix

A Variables

Variable	Description	Variable	Description
name	Player's name.	position	Player's playing position.
team	Team identifier.	xP	Expected fantasy points.
assists	Number of assists.	bonus	Fantasy bonus points awarded.
bps	Bonus Points.	clean_sheets	Matches without conceding goals.
creativity	ICT Creativity metric.	element	Unique player ID.
expected_assists	Expected assists (xA).	expected_goal_involvements	Sum of expected goals and assists.
expected_goals	Expected goals (xG).	expected_goals_conceded	Expected goals conceded (xGC).
fixture	Fixture identifier.	goals_conceded	Goals conceded by team.
goals_scored	Goals scored by player.	ict_index	Combined ICT Index score.
influence	ICT Influence metric.	kickoff_time	Match kickoff date and time.
minutes	Minutes played.	modified	Last data update timestamp.
opponent_team	Opponent's team ID.	own_goals	Own goals scored.
penalties_missed	Penalties missed.	penalties_saved	Penalties saved (GK).
red_cards	Red cards received.	round	Gameweek or round number.
saves	Goalkeeper saves.	selected	Number of managers selecting player.
starts	Matches started by player.	team_a_score	Away team's goals.
team_h_score	Home team's goals.	threat	ICT Threat metric.
total_points	Total fantasy points earned.	transfers_balance	Net transfers (in minus out).
transfers_in	Transfers in by managers.	transfers_out	Transfers out by managers.

value	Player's fantasy value.	was_home	Home match indicator (true/-false).
yellow_cards	Yellow cards received.	GW	Gameweek number.

Table A.1: The variables from the FPL API.

Variable	Description	Variable	Description
name	Player's name.	position	Player's position.
team	Team identifier.	assists	Assists in match.
bonus	Bonus points.	bps	Bonus Points score.
goals_conceded	Goals conceded.	goals_scored	Goals scored
kickoff_time	Kickoff time.	minutes	Minutes played.
opponent_team	Opponent's team ID.	red_cards	Red cards.
saves	Goalkeeper saves.	total_points	Fantasy points.
was_home	True/false.	yellow_cards	Yellow cards.
GW	Gameweek number.	starts	Number of starts.
season	Season identifier.	team_id	Team's unique ID.
team_strength	Team strength.	opponent_strength	Opponent's strength.
avg_min_season	Avg. minutes per match (season).	avg_starts_season	Avg. starts per match (season).
cumulative_goals	Total goals.	cumulative_minutes	Total minutes played.
cumulative_saves	Total saves.	cumulative_threat	Total Threat.
cumulative_assists	Total assists.	cumulative_creativity	Total Creativity.
goals_per90_season	Goals per 90 min (season).	yellow_season	Total yellow cards (season).
prev_transf_delta	Transfers in-out last round.	starts_last5	Starts in last 5 matches.
yellow_last5	Yellow cards in last 5 matches.	red_last5	Red cards in last 5 matches.
ict_last5	ICT score last 5 matches.	points_last5	Points in last 5 matches.
clean_sheets_last5	Clean sheets last 5 matches.	minutes_last5	Minutes played last 5 matches.
team_goals_last5	Team goals last 5 matches.	opp_goals_last5	Opponent goals last 5 matches.
opp_diff_last5	Goal diff. (team-opponent) last 5.	goals_conceded_last5	Goals conceded last 5 matches.
saves_last5	Saves last 5 matches.	goals_last5	Goals scored last 5 matches.

threat_last5	Threat last 5 matches.	assists_last5	Assists last 5 matches.
creativity_last5	Creativity last 5 matches.	goals_season	Total goals (season).
threat_season	Total threat (season).	assists_season	Total assists (season).
creativity_season	Total creativity (season).	saves_season	Total saves (season).
goals_per90_last5	Goals per 90 min (last 5).	threat_per90_last5	Threat per 90 min (last 5).
assists_per90_last5	Assists per 90 min (last 5).	creativity_per90_last5	Creativity per 90 min (last 5).
saves_per90_last5	Saves per 90 min (last 5).	threat_per90_season	Threat per 90 min (season).
assists_per90_season	Assists per 90 min (season).	creativity_per90_season	Creativity per 90 min (season).
saves_per90_season	Saves per 90 min (season).	yellow_per_game_career	Avg. yellow cards per game (career).
red_per_game_career	Avg. red cards per game (career).	selected_last_round	Ownership percent last round.
fk_taker	Free kick taker (yes/no).	pk_taker	Penalty kick taker (yes/no).
corner_taker	Corner taker (yes/no).	opp_kickoff_time	Opponent's previous match kick-off.
opponents_goal_form	Opponent's goal form.	opponents_conceded_form	Opponent's goals conceded form.

Table A.2: The variables from FPL API after feature engineering.

Variable	**Description**
Aerials per 90	Aerial duels attempted per 90 minutes.
Aerials won %	Percentage of aerial duels won.
Aerials won per 90	Number of aerial duels won per 90 minutes.
Assists per 90	Assists provided per 90 minutes.
Attacking aerials won %	Percentage of attacking aerial duels won.
Attacking aerials won per 90	Attacking aerial duels won per 90 minutes.
Ball progression (xT) per 90	Expected Threat (xT) from ball progression per 90 minutes.
Ball recoveries per 90	Number of ball recoveries per 90 minutes.
Ball runs (xT) per 90	Expected Threat generated from ball carries per 90 minutes.
Box entries per 90	Entries into the opposition's box per 90 minutes.

`Carries (xT) per 100 receptions`	Expected Threat generated per 100 ball receptions.
`Carries (xT) per 90`	Expected Threat from ball carries per 90 minutes.
`Counterpressing per 90`	Counterpressing actions performed per 90 minutes.
`Creative passes per 90`	Creative or key passes made per 90 minutes.
`Crosses (xT) per 90`	Expected Threat from crosses per 90 minutes.
`Deep completions per 90`	Completed passes deep into opposition territory per 90 minutes.
`Deep runs (xT) per 90`	Expected Threat from deep runs per 90 minutes.
`Defending 1v1 %`	Percentage of successful defensive 1v1 duels.
`Defensive actions per 90`	Number of defensive actions performed per 90 minutes.
`Defensive aerials won %`	Percentage of defensive aerial duels won.
`Defensive aerials won per 90`	Defensive aerial duels won per 90 minutes.
`Defensive duels won %_x`	Percentage of successful defensive duels.
`Defensive duels won per 90`	Defensive duels won per 90 minutes.
`Defensive intensity per 90`	Intensity of defensive actions per 90 minutes.
`Dribbles (xT) per 90`	Expected Threat from dribbles per 90 minutes.
`Dribbles success %`	Percentage of successful dribbles.
`Goals per 90`	Goals scored per 90 minutes.
`Goals per box touch`	Goals scored per touch inside the box.
`Headed plays per 90`	Number of headed actions per 90 minutes.
`High recoveries per 90`	Ball recoveries made high up the pitch per 90 minutes.
`High turnovers per 90`	Turnovers forced high up the pitch per 90 minutes.
`High turnovers per low reception`	High turnovers relative to low receptions.
`Interceptions per 90`	Interceptions made per 90 minutes.
`Key passes per 90`	Passes leading directly to goal attempts per 90 minutes.
`Linkups per 90`	Number of link-up play actions per 90 minutes.
`Long ball receptions per 90`	Long balls successfully received per 90 minutes.
`Losses per 90`	Ball losses per 90 minutes.
`Passes (xT) per 100 receptions`	Expected Threat generated per 100 passes received.
`Passes (xT) per 90`	Expected Threat from passes per 90 minutes.
`Passes into third (xT) per 90`	Expected Threat from passes into the attacking third per 90 minutes.

Passes third (xT) per 90	Expected Threat from passes in the final third per 90 minutes.
Penalty area receptions per 90	Ball receptions in the penalty area per 90 minutes.
Playmaking passes per 90	Passes contributing significantly to chance creation per 90 minutes.
Possessions won per 90	Possessions regained per 90 minutes.
Second assists per 90	Passes preceding an assist per 90 minutes.
Shot conversion %	Percentage of shots converted into goals.
Successful 1v1 per 90	Successful offensive one-on-one actions per 90 minutes.
Tackle success %	Percentage of successful tackles.
Touches in box per 90	Touches inside opposition box per 90 minutes.
Touches per 90	Total ball touches per 90 minutes.
True tackles won per 90	Successful tackles clearly regaining possession per 90 minutes.
Under pressure retention per 90	Ball retention under pressure per 90 minutes.
xA per 90	Expected Assists per 90 minutes.
xG per 90	Expected Goals per 90 minutes.
xG per box touch	Expected Goals per touch inside the box.
xG per shot_x	Expected Goals per shot.
xGBuildup per 90	Expected Goals contribution from buildup play per 90 minutes.
xGChain per possession	Expected Goals in possession chains per possession.
xGCreated per 90	Expected Goals created per 90 minutes.
xGDribble per 90	Expected Goals contribution from dribbles per 90 minutes.
xGOT per 90	Expected Goals on Target per 90 minutes.

Table A.3: Player-Based metrics from the match-event dataset.

Variable	**Description**
np Shots	Non-penalty shots taken by team.
np Goals	Non-penalty goals scored by team.
np xG	Non-penalty expected goals.
xG	Expected goals including penalties.
Shot conversion	Percentage of shots converted into goals.
xG per shot_y	Expected goals per shot.
High opportunity shots	Number of high-quality scoring opportunities.
xT	Expected Threat generated by team actions.
Penalty area touches	Total touches within the opposition's penalty area.
Box entries	Entries into the opposition penalty box.
Possessions to final third	Team possessions progressing to the final third.
Possessions to final third %	Percentage of possessions reaching the final third.
Final third to box %	Percentage of final third possessions reaching the box.

`Box entries to shot %`	Percentage of box entries resulting in a shot.
`Defensive line height (m)`	Average defensive line height in meters.
`Defensive duels won %_y`	Percentage of defensive duels won.
`Long ball %`	Percentage of long balls played.
`Pass tempo`	Average pace or speed of passing.
`High recoveries`	Ball recoveries in advanced areas.
`Red cards`	Number of red cards received.
`Yellow cards`	Number of yellow cards received.
`Corners`	Number of corners awarded.
`Penalties`	Number of penalties awarded.
`Final third throw-ins`	Throw-ins taken in the final third.
`Match duration`	Total duration of the match.
`Opp. np Shots`	Opponent non-penalty shots.
`Opp. np Goals`	Opponent non-penalty goals.
`Opp. np xG`	Opponent non-penalty expected goals.
`Opp. Goals`	Goals scored by opponent.
`Opp. xG`	Opponent's expected goals including penalties.
`Opp. Shot conversion`	Opponent's shot conversion rate.
`Opp. xG per shot`	Opponent expected goals per shot.
`Opp. High opportunity shots`	Opponent's high-quality shots.
`Opp. xT`	Opponent's Expected Threat.
`Opp. Penalty area touches`	Opponent's penalty area touches.
`Opp. Box entries`	Opponent's box entries.
`Goal difference`	Difference between team's goals and opponent's goals.
`Ball possession %`	Team's percentage of ball possession.
`Opp. Ball possession %`	Opponent's percentage of ball possession.
`xPoints`	Expected points based on match performance.
`Points - xPoints`	Actual points minus expected points.
`Turnovers`	Total turnovers conceded.
`Defensive intensity`	Intensity of team's defensive actions.
`PPDA`	Passes per defensive action by the team.
`Opp. PPDA`	Opponent's passes per defensive action.
`Win probability %`	Probability of winning the match.
`Loss probability %`	Probability of losing the match.
`Draw probability %`	Probability of match ending in a draw.
`Ball-in-play minutes`	Total active playtime minutes.
`Territorial dominance`	Team's territorial control of the game.
`Opp. Territorial dominance`	Opponent's territorial control of the game.

Table A.4: Team-Based metrics from the match-event dataset.

Variable	**Description**
Player-based Metrics	
`aerial_dominance`	Player's effectiveness in aerial duels.
`off_aerial_threat`	Player's aerial threat in attacking situations.
`def_aerial_stability`	Player's stability in defensive aerial duels.
`progressive_threat_efficiency`	Efficiency of player's ball progression actions measured by threat.

`creative_passing_threat`	Threat generated by player's creative passes.
`finishing_efficiency`	Player's effectiveness at converting chances into goals.
`chance_quality`	Quality of scoring opportunities involving the player.
`assist_efficiency`	Efficiency in converting passes into assists.
`shot_involvement_efficiency`	Efficiency of player's involvement in shot actions.
`defensive_duel_effectiveness`	Effectiveness of player's defensive duels.
`effective_pressing_actions`	Player's effective pressing actions per match.
`true_tackle_efficiency`	Efficiency of tackles clearly regaining possession.
`ball_retention_under_pressure`	Player's ability to retain the ball under pressure.
`ball_security`	Player's effectiveness in maintaining possession.
Team-based Metrics	
`attack_directness`	Team's directness in attacking play.
`attacking_efficiency`	Team's overall efficiency in creating and converting chances.
`chance_creation_quality`	Quality of team's created scoring opportunities.
`defensive_compactness`	Team's compactness and organization in defense.
`pressing_intensity_diff`	Difference in pressing intensity compared to opponents.
`high_press_effectiveness`	Team's effectiveness when pressing high.
`quick_recovery_retention`	Proportion of recoveries where possession is retained within 5 seconds.
`transition_attack_quality`	Quality of team's attacking transitions.
`transition_def_stability`	Stability in defensive transitions.
`control_of_territory`	Team's control over different areas of the pitch.
`final_third_control`	Team's effectiveness in controlling the final third.
`game_control`	Overall control of the match by the team.
`game_control_diff`	Difference in game control compared to opponents.
`recovery_effectiveness`	Effectiveness of team's ball recovery actions.
`opp_retention_pressure`	Pressure applied affecting opponent's ball retention.

Table A.5: Created variables based on player and team metrics.

B Variables in models

Model Name	Position	Variables
xGoals	FWD	selected_last_round, relative_team_strength, penalty_chance, threat_last5, attack_directness_team_last5, chance_quality_last5, Touches in box per 90_last5, game_control_diff_team_last5, np Goals_team_last5, cumulative_threat
xGoals	MID	selected_last_round, cumulative_threat, relative_team_strength, Linkups per 90_last5, penalty_chance, chance_quality_last5, corner_chance, Dribbles success %_last5, goals_per90_season, Possessions won per 90_last5, np Goals_team_last5, threat_per90_season
xGoals	DEF	relative_team_strength, cumulative_threat, aerial_dominance_last5, corner_chance, Opp. Defensive line height (m)_team_last5, Ball progression (xT) per 90_last5

Table B.1 The different variables used in xGoals based on position.

Model Name	Position	Variables
XAssists	FWD	relative_team_strength, progressive_threat_efficiency_last5, assists_per90_season, Territorial dominance_team_last5, corner_chance
XAssists	MID	relative_team_strength, Ball progression (xT) per 90_last5, minutes_last5, corner_chance, Linkups per 90_last5, Losses per 90_last5, Passes into third (xT) per 90_last5, creativity_per90_season, attack_directness_team_last5, xA per 90_last5, xGOT per 90_last5
XAssists	DEF	creativity_per90_season, Passes third (xT) per 90_last5, relative_team_strength, corner_chance, ball_security_last5, Crosses (xT) per 90_last5, Box entries per 90_last5, xA per 90_last5, assists_per90_season, progressive_threat_efficiency_last5, Losses per 90_last5, np Goals_team_last5, Territorial dominance_team_last5

Table B.2 The different variables used in xAssists based on position.

Model Name	Position	Variables
xBPS	FWD	avg_min_season, relative_team_strength, cumulative_goals, penalty_chance, np Goals_team_last5, game_control_team_last5, fk_taker, Box entries to shot % _team_last5, chance_quality_last5, progressive_threat_efficiency_last5
xBPS	MID	avg_min_season, relative_team_strength, ict_last5, Passes (xT) per 90_last5, corner_chance, fk_taker, Touches in box per 90_last5, penalty_chance, goals_conceded_last5, avg_starts_season, opponents_conceded_form, Dribbles success %_last5, ball_security_last5, cumulative_goals, minutes_last5, Box entries to shot % _team_last5, off_aerial_threat_last5, opponents_goal_form, was_home, chance_creation_quality_team_last5, transition_attack_quality_team_last5, high_press_effectiveness_team_last5, progressive_threat_efficiency_last5, prev_transf_delta
xBPS	DEF	relative_team_strength, minutes_last5, Passes (xT) per 90_last5, quick_recovery_retention_team_last5, fk_taker, aerial_dominance_last5, opp_goals_last5, opponents_goal_form, Passes third (xT) per 90_last5, cumulative_creativity, Passes into third (xT) per 90_last5, high_press_effectiveness_team_last5, ict_last5, final_third_control_team_last5, Box entries to shot % _team_last5, Possessions won per 90_last5, game_control_diff_team_last5, Under pressure retention per 90_last5, shot_involvement_efficiency_last5, off_aerial_threat_last5, corner_chance, assists_per90_season, assist_efficiency_last5, transition_def_stability_team_last5, true_tackle_efficiency_last5, Tackle success %_last5, recovery_effectiveness_team_last5, control_of_territory_team_last5, goals_conceded_last5

Table B.3: The different variables used in xBPS based on position.

Model Name	Position	Variables
xCon.Goals	MID	relative_team_strength, goals_conceded_last5, ict_last5, Tackle success %_last5, Opp. Shot conversion_team_last5, Defensive aerials won %_last5, clean_sheets_last5, opponents_goal_form, true_tackle_efficiency_last5, Under pressure retention per 90_last5, was_home, Opp. Long ball %_team_last5, Opp. np Goals_team_last5, Opp. Final third to box %_team_last5, Opp. Defensive duels won %_team_last5, PPDA_team_last5, defensive_compactness_team_last5, Opp. xG per shot_team_last5, Opp. np xG_team_last5, selected_last_round, Defensive actions per 90_last5, aerial_dominance_last5, defensive_duel_effectiveness_last5, attack_directness_team_last5, Opp. Defensive line height (m)_team_last5, control_of_territory_team_last5, Opp. High recoveries_team_last5, Turnovers_team_last5, Opp. Corners_team_last5
xCon.Goals	DEF	relative_team_strength, goals_conceded_last5, opponents_goal_form, Opp. Shot conversion_team_last5, clean_sheets_last5, attack_directness_team_last5, Defensive aerials won %_last5, Opp. Final third to box %_team_last5, ict_last5, team_cards_avg, def_aerial_stability_last5, Opp. xG per shot_team_last5, Opp. np xG_team_last5, Opp. np Goals_team_last5, ball_security_last5, was_home, opp_retention_pressure_team_last5, PPDA_team_last5, defensive_compactness_team_last5, Recoveries within 5s_team_last5, Opp. Corners_team_last5, control_of_territory_team_last5, Opp. Territorial dominance_team_last5, Turnovers_team_last5, recovery_effectiveness_team_last5
xCon.Goals	GK	relative_team_strength, Opp. Final third to box %_team_last5, opponents_goal_form, clean_sheets_last5, defensive_duel_effectiveness_last5, prev_transf_delta, saves_season, Opp. Shot conversion_team_last5, Opp. Defensive intensity_team_last5, goals_conceded_last5

Table B.4: The different variables used in xConcededGoals based on position.

Model Name	Position	Variables
xSaves	GK	relative_team_strength, saves_season, was_home, saves_per90_last5, Opp. np Goals_team_last5, goals_conceded_last5

Table B.5 The different variables used in xSaves based on position.

Model Name	Position	Variables
xYellow	ALL	minutes, Defensive duels won per 90 r5, yellow_per_game_career, cumulative_minutes, yellow_season, team_strength, Under pressure retention per 90 r5, Long ball receptions per 90 r5, points_last5, opponents_goal_form, Attacking aerials won % r5, was_home, opponents_conceded_form, prev_transf_delta, High recoveries per 90 r5, selected_last_round, Touches in box per 90 r5, minutes_last5, avg_min_season, Tackle success % r5, red_per_game_career, Attacking aerials won per 90 r5, Headed plays per 90 r5

Table B.6 The different variables used in xYellow based on position.

C Performance metrics of all models

Model	Position	MAE	RMSE	R^2
xGoals	FWD	0.3739	0.4954	0.0691
xGoals	MID	0.1805	0.3374	0.0745
xGoals	DEF	0.0645	0.1830	0.0018
xAssists	FWD	0.1994	0.3444	0.0057
xAssists	MID	0.1995	0.3416	0.0047
xAssists	DEF	0.1037	0.2340	-0.0046
xBPS	FWD	10.86	15.13	0.1030
xBPS	MID	7.046	10.17	0.1663
xBPS	DEF	9.571	11.12	-0.0817
xCon.Goals	MID	0.8002	1.006	0.1342
xCon.Goals	DEF	0.8840	1.109	0.1400
xCon.Goals	GK	0.9060	1.127	0.1357
xSaves	GK	1.520	1.997	0.0362
xYellow	ALL	0.2360	0.3544	-0.0043

Table C.1 Performance Metrics for the models.