FPL-POINTS-PREDICTOR *Predicting Real-World Player Performance*

Abstract

Fantasy Premier League (FPL) is a popular online game in which participants select real Premier League players and earn points based on their on-field performances. Predicting these points is challenging due to the high variability of football outcomes and the influence of multiple contextual factors. This project proposes a data-driven framework to predict player-level FPL points for upcoming gameweeks using historical performance data.

The analysis relies on multi-season player–gameweek datasets and compares several predictive approaches, ranging from simple baselines to supervised learning models, including linear regression and gradient boosting. Feature engineering focuses on recent player form through lagged performance variables, while model evaluation is conducted out of sample using both accuracy-based and ranking-oriented metrics.

The results show that machine learning models—particularly gradient boosting—consistently outperform baseline methods in terms of predictive accuracy and ranking stability. The project delivers a reproducible prediction pipeline, interpretable evaluation outputs, and automated reporting tools. Overall, the findings illustrate how historical FPL data can support informed decision-making in sports analytics setting, while highlighting the inherent limits imposed by the volatility of football performance.

Keywords: sports analytics, Fantasy Premier League, athlete performance, machine learning, predictive modeling, Python

1. Introduction

Fantasy Premier League (FPL) is a popular online football game in which millions of participants select squads composed of real English Premier League players. Points are awarded to each player based on their actual match performances, including goals, assists, clean sheets, and other in-game events. As a result, FPL represents a complex decision-making environment under uncertainty, where participants must continuously evaluate player performance, form, and risk when selecting their own 15 players squad (11 starters, 4 substitutes). I personnaly am enjoying this game for several seasons now and to make the game even more exciting, for the last two seasons, I also created a mini-league in which each member (mainly friends) must commit to a certain amount, which will then be redistributed to the best players in the mini-league.

Predicting FPL points is challenging due to the stochastic nature of football outcomes and the multitude of factors influencing individual performances, such as player form, team strength, opposition quality, injuries, and tactical changes. Despite this uncertainty, large amounts of historical performance data are publicly available, making FPL an attractive setting for applying data-driven and machine learning techniques. Accurate predictions of player-level points can provide valuable insights for decision support in team selection and transfer strategies. It may be interesting to know that at the beginning of each season, a player's price is set. The people who set the prices most likely use the same method (i.e. prediction models) to determine them. The more promising a player is in terms of points scored, the more expensive he will be.

This project addresses the question of whether it is possible to develop a predictive framework capable of estimating Fantasy Premier League points at the player–gameweek level using historical data. Several modelling approaches are implemented and compared, ranging from simple baseline methods to supervised machine learning models. The project focuses on evaluating predictive accuracy, consistency and interpretability, while ensuring reproducibility through a modular and well-structured Python codebase. An important and interessant aspect of the project is to compare whether these predictions can compete with the odds offered by bookmakers.

The remainder of this report is organised as follows. (Section 1 is the introduction) Section 2 reviews related work and existing approaches in sports performance prediction. Section 3 presents the data, modelling methodology, and evaluation strategy. Section 4 reports and analyses the empirical results. Section 5 discusses key findings and limitations, and Section 6 concludes with directions for future work.

2. Literature Review

The prediction of player performance is a well-established topic in sports analytics, supported by the growing availability of detailed historical data. In football, predictive modelling has been applied to tasks such as match outcome forecasting, player valuation, and performance evaluation, typically using historical statistics, contextual variables, and machine learning techniques. While quantitative approaches provide valuable insights, qualitative assessment—the so-called "eye test"—remains relevant, as statistical models cannot fully capture all aspects of player performance.

Existing research and applied projects commonly rely on publicly available match- or player-level datasets, including Fantasy Premier League records and large football databases. Football, and particularly the English Premier League, offers an unusually rich statistical environment compared to other sports. Simple methods such as moving averages or linear regression are often used as interpretable baselines, while more advanced machine learning approaches have shown superior performance by capturing non-linear effects. Hubáček, Šourek, and Železný (2019), for example, show that gradient boosted tree models outperform simpler approaches by modeling higher-order interactions between team strength, recent form, and contextual variables. Although the feature set used in this project is more limited than in some existing studies, restricting model complexity remains an appropriate and robust modelling choice.

Within the specific context of Fantasy Premier League, much of the existing work originates from practitioner communities, including data science blogs and Kaggle competitions. These approaches typically rely on player-level historical features such as minutes played, recent points, and team indicators, but often lack systematic evaluation and reproducibility. Berrar et al. (2019) highlight that the absence of standardized evaluation protocols remains a common limitation in sports analytics research.

This project builds on existing methodologies by implementing and comparing multiple predictive approaches within a unified and reproducible framework. By evaluating baseline models alongside supervised learning methods using consistent performance metrics, it assesses the extent to which historical FPL data can support reliable player-level predictions and how these compare with alternative benchmarks, including bookmaker-based expectations.

3. Methodology

This section describes the data sources, modelling approach, and evaluation strategy used to predict Fantasy Premier League points at the player–gameweek level. The methodology is designed to ensure comparability between models, robustness of results, and full reproducibility.

3.1 Data Description

Two publicly available datasets hosted on Kaggle are used in this project: one providing historical Fantasy Premier League player data, and one providing football match data with bookmaker odds.

Player-level Fantasy Premier League data are obtained from the Kaggle dataset *Fantasy Premier League Player Data (2016–2024)* by Reeve Barreto. This dataset consists of a single CSV file containing player–gameweek–level observations. In this project, only seasons from 2016/17 to 2022/23 are retained; the 2023/24 season is excluded due to incomplete observations. The raw file is used to construct cleaned and processed player–gameweek datasets, which form the basis for all predictive models.

Bookmaker data are obtained from the Kaggle dataset *Club Football Match Data (2000–2025)* by Adam Gbor. From this dataset, only the file *matches.csv* is used. The raw bookmaker odds are extracted and subsequently transformed into normalized implied probabilities (home win, draw, away win), which are then used as an external benchmark for model comparison.

The resulting dataset comprises several hundred thousand player–gameweek observations, each described by multiple features, including player and team identifiers, playing position, minutes played, and FPL points. To capture short-term performance dynamics, lagged features based on past points and playing time are constructed as proxies for recent form and availability.

Basic data quality checks are applied prior to modelling. Observations with missing or inconsistent key variables are removed, while remaining missing values are handled conservatively. Outliers are retained, as extreme performances are an inherent and relevant characteristic of football data. To avoid look-ahead bias and reflect realistic forecasting conditions, a time-based train–test split is used, ensuring that only past information is available at prediction time.

3.2 Modelling Approach

Several predictive models are implemented and compared to evaluate the trade-offs between simplicity, interpretability, and predictive performance. Simple baseline models based on historical averages are first used to establish an interpretable reference point and to contextualise the gains achieved by more complex approaches.

More advanced supervised learning methods are then applied, including linear regression and gradient boosting. Linear regression provides a transparent benchmark, enabling direct interpretation of the relationship between historical player features and future FPL points. Gradient boosting is employed to capture non-linear relationships and interaction effects that are common in football performance data and poorly represented by linear models. All models share the same input features and are evaluated on identical out-of-sample test sets to ensure fair comparison. Hyperparameters are kept close to default values to prioritise robustness and reproducibility over marginal performance improvements.

Model performance is evaluated using complementary metrics. Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) assess point-level accuracy, with RMSE penalising large errors more heavily. The coefficient of determination (R^2) measures variance explained relative to a naïve baseline, while rank-based correlation metrics evaluate the preservation of player ordering within a gameweek, a key criterion for Fantasy Premier League decision-making.

3.3 Evaluation Strategy

The project is implemented entirely in Python, using a modular and reproducible codebase. Core libraries include pandas and NumPy for data manipulation, scikit-learn for model implementation and evaluation, and matplotlib for visualisation and reporting.

The system architecture follows a clear separation of concerns. Data loading and preprocessing are handled in dedicated modules, ensuring that raw data are transformed consistently across experiments. Modelling components are implemented as reusable classes and functions, allowing multiple algorithms to be evaluated using the same data pipeline. Evaluation and reporting modules compute performance metrics and generate tables and figures in a standardised format.

This modular structure allows the full experimental pipeline—from raw data to final results—to be executed in a reproducible manner. All steps, including preprocessing, model training, prediction, and evaluation, can be reproduced using the same configuration and codebase, facilitating transparency and comparability across modelling approaches.

4. Results

This section presents the empirical results obtained from the evaluation of the different predictive models. All results are reported on out-of-sample test data and are based on a consistent experimental setup across models to ensure fair comparison.

4.1 Experimental Setup

All experiments are conducted in a CPU-based environment without GPU acceleration, which is sufficient given the moderate dataset size and the use of classical machine learning models. The project is implemented in Python (version 3.10), relying primarily on pandas and NumPy for data processing, scikit-learn for modelling and evaluation, and matplotlib for visualisation. A single, consistent software environment is used across all experiments to ensure reproducibility.

Model training follows a time-based train–test split to replicate a realistic forecasting setting in which future gameweeks are predicted using only past information. Accordingly, no cross-validation is performed, and each model is trained once per configuration and evaluated on held-out future observations. Hyperparameters are kept close to scikit-learn defaults to prioritise robustness and reproducibility, with gradient boosting models using a fixed number of trees and limited depth to control overfitting.

4.2 Quantitative Performance Evaluation

Model performance is evaluated using multiple complementary metrics, including Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), the coefficient of determination (R^2), and Spearman rank correlation, computed on a hold-out test season (2022/23). Additional diagnostics such as prediction bias, median absolute error (MedAE), and the 90th percentile of absolute error (P90_AE) are reported to provide a more complete view of error distribution and robustness.

Table 1 reports the out-of-sample performance of four models: a seasonal gradient boosting model, a seasonal linear regression model, and two linear regression models based on rolling lag features. Models are ranked according to MAE, which serves as the primary accuracy criterion.

model_key	mae	rmse	r2	spearman	bias	medae	p90_ae	rank_mae
gbm_seasonal	1.122	2.025	0.263	0.693	0.036	0.462	2.772	1
linear_seasonal	1.217	2.083	0.22	0.666	0.045	0.46	2.918	2
linear_anytime_lag5	1.219	2.086	0.22	0.666	0.045	0.46	2.923	3
linear_anytime_lag3	1.248	2.105	0.204	0.67	0.057	0.566	2.938	4

Figure 1: Predictive Performance across models via metrics.

The gradient boosting seasonal model achieves the best overall performance across nearly all metrics, with an MAE of 1.122 and an RMSE of 2.025, indicating a substantial reduction in prediction error compared to linear baselines. It also explains a larger share of variance ($R^2 = 0.263$) and exhibits the strongest ranking ability, with a Spearman correlation of 0.693, highlighting its effectiveness in preserving player ordering within gameweeks.

Linear regression models perform consistently but remain inferior to the gradient boosting approach. The seasonal linear model yields an MAE of 1.217 and an R^2 of 0.220, while the lag-based linear models show slightly weaker performance, particularly for shorter lag windows. These results suggest that linear models capture part of the signal contained in historical player data but struggle to fully model the non-linear dynamics underlying Fantasy Premier League point outcomes.

Across all models, prediction bias remains low (below 0.06 in absolute value), indicating that errors are largely symmetric and that no systematic over- or under-prediction is present. Median absolute errors are substantially lower than mean errors, reflecting the heavy-tailed nature of the point distribution, where a small number of high-variance performances drive a significant share of total error. This observation is further supported by the P90_AE values, which range between approximately 2.77 and 2.94 points.

Overall, these results confirm that machine learning models, and gradient boosting in particular, provide measurable and consistent improvements over linear and heuristic baselines, both in terms of point prediction accuracy and ranking quality. The superior performance of the gradient boosting model supports the hypothesis that non-linear interactions between historical player features play an important role in predicting future Fantasy Premier League performance.

4.3 Qualitative Analysis and Visualisation

To complement the numerical evaluation, several visual diagnostics are used to assess model behaviour and error structure. These include plots of predicted versus actual points, as well as residual plots.

Predicted-versus-actual scatter plots illustrate how closely model predictions align with realised Fantasy Premier League points. Models with higher predictive accuracy exhibit tighter clustering around the diagonal, indicating better calibration. Residual plots provide insight into the distribution of prediction errors and reveal patterns such as heteroskedasticity, where larger errors tend to occur for extreme point outcomes.

Across these visual diagnostics, the gradient boosting model consistently shows improved alignment with observed values and reduced dispersion relative to simpler models. Nevertheless, all models exhibit increased uncertainty for high-variance outcomes, reflecting the inherently stochastic nature of football performance. It does indeed seem very difficult to predict statistically unusual performances such as a hat trick, even the best players in the world rarely achieve such a performance.

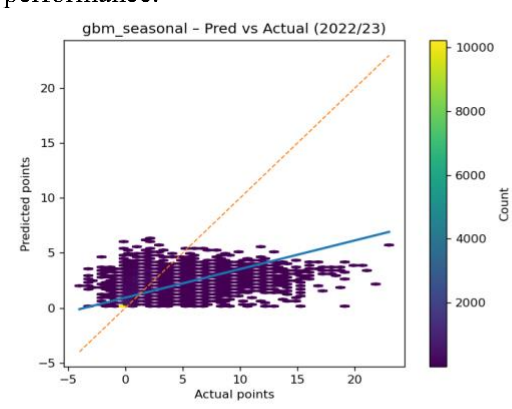

Figure 2: Predictions vs actual points

The predicted-versus-actual scatter plot highlights a clear positive relationship between model predictions and realised Fantasy Premier League points, indicating that the model captures the overall ordering of player performances. Most observations lie within a relatively narrow prediction range, revealing a strong compression effect. While this behaviour limits the model's ability to reproduce extreme outcomes, it contributes to stable rankings across players. The systematic underprediction of high realised scores, reflected by the regression line lying below the 45-degree reference, is characteristic of a regression-to-the-mean effect. This suggests that the model prioritises robustness and ranking consistency over exact point forecasts, which is desirable in a decision-making context such as Fantasy Premier League team selection.

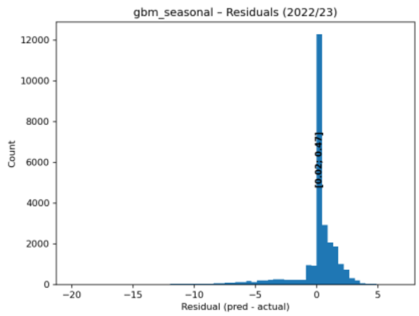

Figure 3: Residuals with GBM

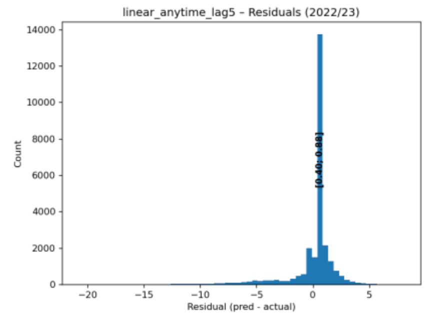

Figure 4: Residuals with linear model

Figures 3 and 4 compare the residual distributions of the gradient boosting model and the linear regression model with lag-5 features, highlighting how model complexity affects the magnitude and concentration of prediction errors.

For the gradient boosting model (Figure 3), residuals are strongly concentrated within a narrow interval, with the highest density observed between 0.02 and 0.47 points. This tight clustering indicates that, for the majority of observations, prediction errors remain small and centred close to zero, reflecting effective calibration across typical Fantasy Premier League outcomes. The relatively limited dispersion suggests improved robustness, particularly for players with stable playing time and consistent roles.

In contrast, the linear regression model (Figure 4) exhibits a substantially wider residual distribution. Its highest-density region lies between 0.40 and 0.88 points, indicating more frequent and larger deviations between predicted and realised scores. This shift toward higher residual values reflects systematic underestimation of realised points, especially for above-average performances, and highlights the difficulty of linear specifications in capturing non-linear relationships and interaction effects present in player performance dynamics. While both models display residual distributions centred slightly above zero, the magnitude of this positive bias is markedly reduced for the gradient boosting model. Overall, the comparison demonstrates that increased model flexibility leads to a clearer contraction of residuals toward smaller errors, resulting in more stable and reliable point predictions.

player_id	name	team	pos	GW	pred	actual
68	Ben Pearson	AFC Bournemouth	MID	16	0.15	1
333	Bruno Borges Fernandes	Manchester United	MID	37	3.51	8
66	Dominic Solanke	AFC Bournemouth	FWD	20	2.79	0
154	James McArthur	Crystal Palace	MID	31	1.16	0
532	Kevin Mbabu	Fulham	DEF	26	0.15	0
496	Luke Plange	Crystal Palace	FWD	10	0.15	0
72	Mark Travers	AFC Bournemouth	GK	18	2.75	2
653	Omari Hutchinson	Chelsea	MID	28	0.15	0
586	Pervis Estupiñán	Brighton & Hove Albion	DEF	27	3.16	1
329	Raphaël Varane	Manchester United	DEF	34	0.75	0

Figure 5: 10 players points predictions

gw	pid	name	team	pos	pred_points	act_points
35	356	Callum Wilson	Newcastle	FWD	5.13	2
24	465	Jarrod Bowen	West Ham	MID	4.66	2
3	357	Kieran Trippier	Newcastle	DEF	4.32	6
21	7	Martin Ødegaard	Arsenal	MID	5.37	5
15	369	Miguel Almirón Rejala	Newcastle	MID	4.33	8
33	40	Ollie Watkins	Aston Villa	FWD	4.63	2
26	31	Emiliano Martínez Rom...	Aston Villa	GK	5.37	7
4	225	Rodrigo Moreno	Leeds United	FWD	5.28	2
9	318	Erling Haaland	Man City	FWD	5.64	23
30	13	Bukayo Saka	Arsenal	MID	4.25	1

Figure 6: Best player of 10 GWs

Figures 5, 6 and 7 illustrate the behaviour of the model at the player level across different sampling schemes, ranging from randomly selected players to top-ranked players within specific Gameweeks.

In the random player sample (Figure 5), predicted scores display a pronounced concentration at very low values, with several players receiving identical predictions of 0.15 points. These cases correspond to players with minimal recent playing time, for whom the model primarily estimates the probability of appearance rather than expected on-field contribution. Conversely, high-impact players within the same sample are assigned substantially higher predicted values. For instance, Bruno Fernandes, who records a high realised score, is also associated with a markedly higher prediction, indicating that the model successfully differentiates between low-involvement and high-involvement profiles.

This calibration pattern persists when predictions are aggregated at the Gameweek level. In Figure 6, which reports the best predicted player across ten randomly selected Gameweeks, predicted scores cluster within a relatively narrow interval of approximately 4.3 to 5.6 points. In many cases, realised outcomes are of comparable magnitude (e.g. Ødegaard: 5.37 predicted vs 5 realised; Trippier: 4.32 vs 6; Martínez: 5.37 vs 7), suggesting reasonable accuracy for typical performances. Extreme deviations remain present, as illustrated by Haaland, whose realised score of 23 points vastly exceeds his predicted value of 5.64, underscoring the inherent difficulty of forecasting exceptional match-level events.

A similar compression of predictions is observed in the fixed Gameweek analysis (Figure 7). For demo GW21, the model assigns predicted scores between approximately 4.1 and 5.4 points to the top-ranked players. While several realised outcomes closely match predictions (e.g. Ødegaard: 5.37 vs 5; Bruno Fernandes: 4.15 vs 5), notable discrepancies persist (e.g. Trippier: 4.32 vs 8; Rashford: 4.18 vs 7). These deviations reflect match-specific volatility rather than systematic miscalibration.

Taken together, these results confirm that the model produces conservative and tightly clustered point estimates, which limits its ability to predict extreme performances but ensures stable and informative player rankings across different Gameweeks and selection schemes.

Figures 8, 9 and 10 extend the analysis from the player level to team-level outcomes and benchmark the model's predictions against bookmaker probabilities, providing an external validation of the estimated team strengths.

gw	pid	name	team	venue	opponent	pred_points	act_points
21	7	Martin Ødegaard	Arsenal	H	Man United	5.37	5
21	427	Harry Kane	Spurs	A	Fulham	4.79	6
21	107	Solly March	Brighton	A	Leicester City	4.61	2
21	332	Luke Shaw	Man United	A	Arsenal	4.39	0
21	357	Kieran Trippier	Newcastle	A	Palace	4.32	8
21	407	James Ward-Prowse	Southampton	H	Aston Villa	4.19	2
21	335	Marcus Rashford	Man United	A	Arsenal	4.18	7
21	303	Riyad Mahrez	Man City	H	Wolves	4.16	6
21	333	Bruno Borges Fern...	Man United	A	Arsenal	4.15	5
21	628	Serge Aurier	Nott'm Forest	A	Bournemouth	4.11	2

Figure 7: Top 10 players of a single GW

gameweek	home_team	away_team	p_home_model	p_home_b365	abs_error_cal
38	Arsenal	Wolverhampton Wanderers	0.4812	0.6988	0.2175
29	Manchester City	Liverpool	0.3088	0.586	0.2772
4	Aston Villa	West Ham United	0.5249	0.4177	0.1072
25	Fulham	Wolverhampton Wanderers	0.4216	0.3794	0.0422
18	Nottingham Forest	Chelsea	0.3563	0.1843	0.172
18	Manchester City	Everton	0.5104	0.8436	0.3332
27	Southampton	Brentford	0.3619	0.3221	0.0398
4	Manchester City	Crystal Palace	0.5489	0.8094	0.2606
9	West Ham United	Wolverhampton Wanderers	0.4619	0.4904	0.0286
32	Liverpool	Nottingham Forest	0.6186	0.8144	0.1958

Figure 8: 10 random team strength

team	top11_sum	top11_avg	rows	n_players_used	top_n_players_used
Man City	56.34	5.12	1145	31	11
Liverpool	55.07	5.01	1331	36	11
Brighton	53.46	4.86	1295	41	11
Arsenal	53	4.82	1230	38	11
Newcastle	51.91	4.72	1229	36	11

Figure 9: Top 5 teams according to model

In the random match sample (Figure 8), the model assigns home win probabilities that typically fall within a moderate range of approximately 0.30 to 0.70, depending on the relative strength of the teams involved. For most fixtures, the absolute difference between the model's predictions and Bet365 implied probabilities remains limited, often below 10 percentage points, indicating a satisfactory level of calibration. Larger discrepancies tend to occur in matches involving top teams, where bookmakers assign very high probabilities, while the model remains more conservative.

This conservative behaviour is consistent with the team-level rankings shown in Figure 9. The model clearly differentiates teams based on the aggregated predicted strength of their top-11 players. Manchester City ranks first with a Top-11 sum of 56.34 points (average 5.12 per player), followed closely by Liverpool (55.07, 5.01) and Brighton (53.46, 4.86). The relatively narrow range between teams—approximately 51.9 to 56.3 points—suggests a competitive upper tier, while still capturing meaningful differences in squad quality derived from individual player predictions. Brighton's position may appear counterintuitive, but it aligns with a squad-building strategy focused on depth rather than star power. The club's practice of selling its best performers and reinvesting in promising players results in a high number of contributors. This is further supported by the use of 41 distinct players in the analysis, indicating strong internal competition and rotation.

The comparison with bookmakers over a larger sample of 182 matches (Figure 10) confirms these patterns quantitatively. The model achieves a mean absolute error of 0.153 relative to Bet365 probabilities, with a positive correlation of 0.267, indicating partial but non-trivial alignment with market expectations. The largest absolute errors, reaching approximately 0.40–0.44, systematically involve matches featuring Manchester City, where bookmakers assign extreme home-win probabilities (0.84–0.89), while the model predicts substantially lower values (0.42–0.46). Conversely, near-perfect agreement is observed in more balanced fixtures, with absolute errors as low as 0.001, typically when both the model and bookmakers estimate probabilities in the 0.40–0.46 range.

- **MAE:** 0.153
- **Corr:** 0.267
- **n:** 182

Worst 5 predictions (largest |Bet365 – model|)

gw	home_team	away_team	p_home_win_b365	p_home_win_model	abs_error
27	Crystal Palace	Manchester City	0.106	0.55	0.443
5	Manchester City	Nottingham Forest	0.894	0.461	0.432
15	Manchester City	Fulham	0.852	0.421	0.431
10	Manchester City	Southampton	0.861	0.453	0.408
16	Manchester City	Brentford	0.842	0.443	0.4

Best 5 predictions (smallest |Bet365 – model|)

gw	home_team	away_team	p_home_win_b365	p_home_win_model	abs_error
5	Leeds United	Everton	0.451	0.452	0.001
21	Leeds United	Brentford	0.405	0.403	0.001
33	Wolverhampton Wanderers	Crystal Palace	0.382	0.381	0.001
31	Everton	Fulham	0.463	0.462	0.001
30	Leeds United	Crystal Palace	0.453	0.459	0.005

Figure 10: Comparison with bookmaker

Overall, these results indicate that the model captures relative team strength effectively and aligns well with bookmaker assessments for typical matchups, while deliberately avoiding extreme probability assignments. This conservative stance reflects a data-driven approach that prioritises robustness over market sentiment, particularly in fixtures involving dominant teams. For our predictive model to be financially viable, i.e. for our implied odds to 'beat' the odds offered by bookmakers, the big teams would need to be regularly upset by the smaller ones.

5. Discussion

5.1 What worked well

The proposed framework successfully captures meaningful predictive signals from historical FPL data, with the gradient boosting model consistently outperforming linear baselines. At the player level, the model demonstrates strong ranking performance, with a clear positive relationship between predicted and realised points and well-controlled residual dispersion. Although point predictions are conservative, the model reliably distinguishes high-impact from low-involvement players and produces stable rankings across Gameweeks—an essential property in an FPL decision-making context.

At the team level, aggregating player predictions into team strength measures yields coherent and interpretable rankings. Match-level probability estimates align well with bookmaker odds for most fixtures, and the relatively low mean absolute error against Bet365 probabilities supports the external validity of the approach

5.2 Challenges, limitations, and insights

A central challenge was modelling the high intrinsic variability of football performance, driven by unobserved factors such as injuries, tactical decisions, line-ups, substitutions, and rare match events. Early linear models struggled to capture non-linear relationships and interaction effects, which motivated the adoption of gradient boosting techniques. Careful control of model complexity proved essential to improve predictive accuracy while avoiding overfitting, ultimately resulting in a more robust and stable framework.

The results largely confirm the initial hypothesis that machine learning models outperform simpler baselines by exploiting richer feature interactions. However, the degree of prediction compression and the systematic underestimation of extreme outcomes were more pronounced than initially expected. This behaviour reflects a deliberate trade-off between robustness and sensitivity and is consistent with regression-to-the-mean effects. Despite this limitation, the model consistently identifies high-potential players and dominant teams, supporting its intended use as a decision-support tool rather than a point-exact forecasting engine.

Several limitations of the approach must be acknowledged. The model relies exclusively on historical performance data and does not incorporate real-time contextual information such as injuries or line-up announcements, which constrains short-term predictive accuracy for players with volatile playing time. Its conservative design further leads to systematic underprediction of rare but high-impact events. Comparisons with bookmaker odds also indicate that betting markets embed additional information and sentiment not captured by the model, particularly in fixtures involving dominant teams. A notable and somewhat surprising finding is that conservative predictions still yield strong comparative performance. By prioritising relative strength over extreme forecasts, the model remains well aligned with realised outcomes and bookmaker assessments. In addition, the relatively narrow dispersion of team strength estimates among top teams highlights the highly competitive nature of the Premier League, where small differences in squad composition and player form can have meaningful effects despite appearing limited in absolute terms.

6. Conclusion

This project investigated the extent to which historical Fantasy Premier League data can be leveraged to predict player- and team-level performance in an inherently uncertain sporting environment. By combining structured data pipelines with supervised learning models, the study assessed both the predictive potential and the practical limits of data-driven approaches to Fantasy Premier League decision-making.

The primary contribution of this project is the development of a robust and interpretable framework for predicting Fantasy Premier League points. The results demonstrate that machine learning models—particularly gradient boosting—substantially improve predictive accuracy and error control relative to linear baselines. While exact point prediction remains difficult, the model effectively captures relative player performance and produces stable rankings across Gameweeks.

At the player level, predictions are intentionally conservative, reflecting a trade-off between robustness and sensitivity to extreme outcomes. Despite this compression, the model reliably identifies high-impact players and distinguishes low-involvement profiles, aligning well with FPL decision-making focused on relative comparisons rather than precise forecasts.

At the team level, aggregating player predictions into team strength measures yields coherent and interpretable rankings. Match-level probability estimates align meaningfully with bookmaker odds for typical fixtures, while remaining more cautious in games involving dominant teams, thereby limiting overconfident predictions driven by market sentiment. Achieving financial viability (across sports betting markets) with these predictions would require the presence of persistent market inefficiencies, such as a systematic underestimation of smaller teams leading to more frequent-than-expected upsets.

Overall, the project meets its objectives by showing that historical performance data contains exploitable predictive signals at both player and team levels, while clearly acknowledging the limits imposed by the inherent stochasticity of football outcomes.

6.2 Future Work

Several directions for future work emerge from this study. From a methodological standpoint, incorporating real-time contextual information—such as injuries, line-ups, fixture congestion, or tactical factors—could improve short-term predictive accuracy and mitigate the underprediction of extreme performances. Probabilistic or Bayesian models may also better capture predictive uncertainty, particularly for high-variance players and matchups.

Further work could explore loss functions focused on ranking quality rather than point accuracy, aligning more closely with Fantasy Premier League decision-making. Ensemble approaches combining multiple model classes may additionally improve robustness.

From an applied perspective, the framework could be extended into an automated decision-support tool providing weekly rankings, transfer recommendations, and risk-adjusted metrics. The methodology is also transferable to other fantasy sports and performance-based forecasting problems in finance and analytics, where ranking and uncertainty management are central.

Finally, while football outcomes remain inherently unpredictable, this project shows that well-designed data-driven models can deliver interpretable and actionable insights by explicitly accounting for uncertainty.

References

1. Fantasy Premier League. (2024). **Official Fantasy Premier League Rules and Scoring System*.* Retrieved from https://fantasy.premierleague.com

2. Kaggle. (2023). **Fantasy Premier League Dataset*.* Retrieved from https://www.kaggle.com

3. Hubáček, Šourek and Železný (2019) **Gradient boosting for soccer prediction lecture**

4. Berrar et al. (2019) **Domain knowledge in soccer prediction**

5. https://www.youtube.com/@LetsTalkFPL **YouTube FPL channel**

6. https://www.youtube.com/@FPLHarry **YouTube FPL channel**

Appendices

Appendix A: Code Repository

The complete source code for this project is available on GitHub and is structured to ensure clarity, modularity, and reproducibility.

GitHub Repository:
https://github.com/JulesProgg/fpl-points-predictor

Appendix B: Install instructions

This project uses a Conda environment to ensure reproducibility.

git clone https://github.com/JulesProgg/fpl-points-predictor cd fpl-points-predictor conda env create -f environment.yml conda activate fpl-points-predictor

Alternatively, if Conda is not available, the dependencies listed in environment.yml can be installed manually in an existing Python environment (Python ≥ 3.10).

Appendix C: Reproducing Results

All results reported in this paper can be reproduced by running the main.py.

Appendix D: Project structure

FPL-POINTS-PREDICTOR/

— main.py

— README.md

— environment.yml

— src/ │ – init.py │ – data_loader.py │ – evaluation.py │ – models.py │ – reporting.py

— data/ │ – raw/ │ – processed/

— results/ │ – figures/ │ – metrics/ │ – predictions/ │ – tables/

— tests/ │ – test_data_loader.py │ – test_evaluation.py │ – test_models.py

— report/ │ - final_report.pdf

— .gitignore

— AI_USAGE.md

— download_data_instructions.md

— PROPOSAL.md

— pyproject.toml