Fantasy Draft Optimizer 2025: Enhanced PPR Projections Using 2024 Data

By, Jonathan Siegel and Isabelle Bernal

***Abstract*—Fantasy football participants rely on a scoring system or historical rankings which may not account for nuanced performance patterns or trends in the 2024 season. Current projections can lead to consequential draft decisions due to a lack of precision in evaluating player potential based on their roles, consistency, and point contribution under PPR standards. The primary goal of this project is to create a data-driven system for drafting the best fantasy football team for the 2025 season. We will use an analysis of 2024 player performance data, specifically the project aims to refine the PPR scoring system to produce more accurate player projections and improve team structure strategies.**

I. Introduction

Fantasy football is a game that allows participants to create their own virtual team of players from the NFL. Each participant is also known as an owner and is a part of a particular league with an average of ten teams. Participating in fantasy football is exciting and engaging but also has hardships since the performance of individual players tends to vary largely. Some players who were expected to have high success may underperform while players who were not expected to perform well might surprise participants. Overall, these uncertainties are what make fantasy football addicting and rewarding.

Our project is important and would benefit fantasy football participants since our project aims to address challenges such as time-consuming research when selecting players and relying on inaccurate projections to create their team. Most people who play fantasy football often work full-time and lack the resources and time to make data-driven decisions when choosing players for their team. While there are websites out there offering general advice, our project is unique in the sense that it utilizes unique datasets and advanced machine-learning algorithms to provide more accurate player projections and recommendations. By using these machine learning techniques, we aim to enhance the fantasy football experience and allow for a better decision-making process for participants.

II. Literature Review

There are other machine learning projects out there such as https://github.com/jamesterrell/Fantasy_Football_Machine_Learning which utilizes libraries and machine learning approaches such as pandas, requests, and forecasting models to predict player performances. In comparison, this project only uses forecasting models (autoregressive models) to

predict player performance meanwhile our project uses a wide range of machine learning models, including linear regression, regularization, KMeans clustering, and gradient boosting. The variety of machine learning models for our project allows the performance and prediction to be more effective compared to the GitHub project. In addition, our project uses multiple datasets, a refined PPR scoring system that adapts to player and league trends, along with using evaluation metrics of R-Squared and AUC-ROC to help fine-tune our models' predictions. However, this GitHub project only uses one dataset, uses different evaluation metrics, and only focuses on a preexisting PPR scoring system which distinguishes our projects.

There is another machine learning project such as https://github.com/asimw4/fantasy-football-point-predictor which uses libraries such as pandas, scikit-learn, and matplotlib/seaborn, along with multiple datasets, and implements a Random Forest Regressor that achieves an accuracy of 67.5%. In comparison to our project where we use a variety of models, multiple datasets as well, and evaluation metrics in which this GitHub project does not specify that they use evaluation metric which could be the reason their accuracy is not as high.

Therefore, our project is unique and more resourceful compared to other fantasy football machine learning projects since our project is based on multiple machine learning techniques including gradient boosting, linear regression, classification, and clustering along with integrating multiple datasets. The use of the evaluation metrics of AUC-ROC and R-squared ensures that our predictions are more dependable and accurate compared to other fantasy football player predictors.

III. Data Collection and Preprocessing

In our project, we used two different datasets found in https://github.com/nflverse/nflverse-data/releases?page=2. The first dataset that we used is called player_stats_2024.csv which contains information about every NFL player except kicker, their position, and their weekly stats such as rushing yards, receptions, receiving yards, and touchdowns. This is important for our project because to create our projections, we need all of these stats to be able to compute anything. The third dataset we used is called player_stats_kicking_2024.csv which contains information about all NFL kickers and their weekly stats such as field goals made, missed, and blocked. This is especially important when it comes to calculating projected points for each kicker because these stats are what contribute to the PPR point system calculation.

Data cleaning and preprocessing mostly involved renaming the columns that stored the player ID to player_id so that there would not be any issues when merging all the datasets. Having a player ID in our merged dataset is important because in the event there are two players with the same name, we can distinguish them by their ID. In addition, we made sure that the inputs in the column all had the same type and were all uppercase. Next, we handled missing values and incorrect data types in all the datasets by excluding them since this would be considered incorrect input. Lastly, we checked that in all the datasets we used, there were no more missing or invalid values in the dataset to ensure proper merging when we started manipulating the data.

IV. Approach

The primary goal of this project is to optimize fantasy football team selections for the 2025 season by using enhanced Points Per Reception projections based on 2024 data. We broke this down into seven steps: data collection, data processing and cleaning, exploratory data analysis, feature engineering, machine learning models (Gradient Boosting, XGBoost, and Linear regression), position-based analysis, and evaluation metrics to analyze the predictive accuracy and outcomes.

Diving into each of the steps, for data collection we used two datasets that included player statistics and kicking performance to analyze each NFL player's performance. These datasets included metrics such as rushing yards, receptions, touchdowns, field goals, and receiving yards which are all crucial when calculating the PPR scoring. For data processing and cleaning, we addressed any inconsistencies, and missing values, and standardized the column names to ensure smooth data merging. This was essential to avoid any biased or inaccurate predictions.

Next, is the exploratory data analysis which allows us to identify any trends and patterns in the data such as player performance, key features, and outliers. The fourth step is featuring engineering which allows us to enhance the accuracy of the predictions of the machine learning models. We created new features that allowed our models to capture player performance better. The fifth step is machine learning models which were Gradient Boosting, XGBoost, and Linear regression where we were able to compare the performances of these models and select the most accurate predictor of the PPR scores.

For the sixth step, we implemented position-based analysis where we predicted the top-performing players in the position of quarterback, running back, wide receiver, tight end, flex, and kicker which depended on the machine learning models. Lastly, the seventh step was the evaluation metrics where Root Mean Squared Error and R-squared were used to identify the best-performing approach and which models captured the variance in PPR scores the best.

V. Data Analysis

The data utilized for our Fantasy Football project was sourced from GitHub. The primary dataset, player_stats_2024.csv, contains comprehensive statistics for offensive players, including receiving yards, receptions, passing yards, and other performance metrics. Another significant dataset, player_stats_kicking_2024.csv, focuses on kicking statistics, such as field goals made/attempted and points after touchdowns (PATs).

Upon reviewing the datasets, we observed that they were already well-structured, requiring minimal cleaning. For missing data, we applied techniques like pandas.fillna to fill gaps with placeholder values. Additionally, we renamed certain columns, such as gsis_id to player_id, for clarity and to enhance feature engineering and dataset organization.

Before implementing machine learning algorithms, we leveraged the raw dataset to identify the best fantasy team by filtering and ranking players using functions like sort_values(by='fantasy_points_ppr', ascending=False). This allowed us to assemble an optimal team based purely on existing data without applying predictive modeling. Similarly, we calculated the average fantasy score for each player as of Week 12 to gain further insights. The results from both methods were largely consistent, featuring standout players such as Barkley from the Eagles and Chase from the Bengals.

For the predictive phase, we trained two different regression models to project an ideal fantasy team for the 2025 season. Our target variable, fantasy_points_ppr, represented a player's fantasy performance. During feature engineering, we numerically encoded player positions (creating position_encoded) and developed additional metrics, such as total_touches, yards_per_touch, and tds_per_game, which are critical for maximizing fantasy points. Key performance indicators, including passing, rushing, and receiving yards, were selected as predictors. The dataset was then split into predictors (X) and the target variable (y).

To ensure robust model evaluation and selection, we implemented a 5-fold cross-validation scheme, reducing variance and improving reliability. We trained and evaluated two machine learning models, Gradient Boosting and XGBoost Regressors, using metrics like RMSE and R^2 to compare their performance.

Additionally, a simple Linear Regression model was retrained on the entire dataset to make final predictions, maximizing the use of all available data for accuracy. This model was also trained on the predictors (X) and target variable (y) identified earlier.

Both modeling approaches produced similar results. The final outputs were structured to include players for positions QB, RB, RB, WR, WR, TE, and FLEX. Players were ranked based on their predicted fantasy points, and the top-performing players for each position were selected to fulfill the team structure requirements.

VI. Evaluation and Results

The evaluation of this project is based on the performance of the various machine learning models, analysis of the predictive accuracy, and the use of the results to create an optimal fantasy football team for the 2025 season. The machine learning models that were used are Gradient Boosting, XGBoost, and Linear Regression where Gradient Boosting is the best-performing model out of the three. This model was fine-tuned using 5-fold cross-validation to optimize the parameters such as max_depth, learning_rate, and n_estimaters. The Gradient Boosting model resulted in an average RMSE of 1.1585 and an R-squared score of 0.97999 which indicated there is high accuracy and strong predictability. In comparison to the XGBoost model where the performance was not as strong with an RMSE of 1.2063 and an R-squared score of 0.9944. The Linear Regression model is in between XGBoost and Gradient boosting with the results of their RMSE being 1.1118 with an R-squared score of 0.9799.

Using the predictions from the Gradient Boosting model, we were able to identify the best players for each position in fantasy football. Jalen Hurts (PHI) was projected as the top quarterback with a PPR score of 34.65. For running backs, Alvin Kamara (NO) and Saquon Barkley (PHI) were listed as top running backs with PPR scores of 45.57 and 45.27. Among wide receivers, Ja'Marr Chase (CIN) and Jauan Jennings (SF) were selected as top wide receivers with PPR scores of 54.06 and 46.44. Taysom Hill (NO) was projected as the top tight end with a PPR score of 41.25. Lastly, for the FLEX position, CeeDee Lamb (DAL) was projected as the highest-scoring player among wide receivers, running backs, and tight ends with a PPR score of 40.31.

The kicker position was evaluated separately from the other positions due to the separate dataset. A Random Forest Regression was trained to predict the kicker's performance based on the metrics of field goals made, attempts, and extra points after touchdowns. The top kicker predicted was Ka'imi Fairbairn (HOU) with the projection of total season points being 110.98.

```
Mean Squared Error: 0.07366724137931037
R^2 Score: 0.9943847778306483

Best Kicker:
player_id              00-0032726
player_name           K.Fairbairn
team                          HOU
predicted_points           110.98
```

The results were visually shown through tables and charts that showcased the predicted PPR scores for each position and the performance metric for all the models. Out of the three models, the Gradient Boosting model outperformed the other models consistently and demonstrated the reliability and accuracy of predicting the top players. Being able to use various metrics such as RMSE and R-squared allowed the comparison of the various machine learning models to be based on accuracy and reliability. In addition, this allowed for the uncovering of breakout players and optimizing the positional balance to create the best team for the 2025 season.

Despite the models delivering strong predictive accuracy, there were still limitations in our project. The analysis was restricted to data only available up to Week 12 of the 2024 season which excludes any late-season data up to Week 18 and any unforeseen factors. In addition, other influences that were not accounted for were changes in the coaching staff and team dynamics such as player trades. In the future, we could incorporate some time-series analysis or additional datasets to address these limitations that could impact the predictions. In all, the Gradient Boosting model provided the most accurate and reliable predictions which demonstrate the positive impact of machine learning in fantasy football analytics and predictions.

	player_id	player_name	player_display_name	position	position_group	headshot_url	recent_team
0	00-0036389	J.Hurts	Jalen Hurts	QB	QB	https://static.www.nfl.com/image/upload/f_auto...	PHI
1	00-0033906	A.Kamara	Alvin Kamara	RB	RB	https://static.www.nfl.com/image/private/f_aut...	NO
2	00-0034844	S.Barkley	Saquon Barkley	RB	RB	https://static.www.nfl.com/image/upload/f_auto...	PHI
3	00-0036900	J.Chase	Ja'Marr Chase	WR	WR	https://static.www.nfl.com/image/upload/f_auto...	CIN
4	00-0036259	J.Jennings	Jauan Jennings	WR	WR	https://static.www.nfl.com/image/upload/f_auto...	SF
5	00-0033357	T.Hill	Taysom Hill	TE	TE	https://static.www.nfl.com/image/upload/f_auto...	NO
6	00-0036358	C.Lamb	CeeDee Lamb	WR	WR	https://static.www.nfl.com/image/upload/f_auto...	DAL

Above is a picture of Gradient Boosting results with their prediction of the best team for the 2025 season.

	player_id	player_name	player_display_name	position	position_group	headshot_url	recent_team
0	00-0034796	L.Jackson	Lamar Jackson	QB	QB	https://static.www.nfl.com/image/upload/f_auto...	BAL
1	00-0033906	A.Kamara	Alvin Kamara	RB	RB	https://static.www.nfl.com/image/private/f_aut...	NO
2	00-0034844	S.Barkley	Saquon Barkley	RB	RB	https://static.www.nfl.com/image/upload/f_auto...	PHI
3	00-0036900	J.Chase	Ja'Marr Chase	WR	WR	https://static.www.nfl.com/image/upload/f_auto...	CIN
4	00-0036259	J.Jennings	Jauan Jennings	WR	WR	https://static.www.nfl.com/image/upload/f_auto...	SF
5	00-0033357	T.Hill	Taysom Hill	TE	TE	https://static.www.nfl.com/image/upload/f_auto...	NO
6	00-0038543	J.Smith-Njigba	Jaxon Smith-Njigba	WR	WR	https://static.www.nfl.com/image/upload/f_auto...	SEA

Above is a picture of the linear regression model results with their prediction of the best team for the 2025 season

VII. Conclusion

In conclusion, our fantasy football draft optimizer project showcases how utilizing different machine learning techniques provides a more accurate and data-driven player projection. Using the three datasets based on the 2024 NFL season and various machine learning techniques, we created a predictive model that refined the PPR scoring system and picked the best team to draft for the 2025 fantasy football season. Out of the three machine learning models, Gradient Boosting was the most accurate and reliable with an RMSE of 1.1585 and an R-squared score of 0.9799. The project identified top-performing players for each position which offered insights into who to draft for the 2025 fantasy football season.

However, despite the major success of our project, there were also some limitations in our project as well. This includes late-season data, coaching changes, and player trades. These limitations can be implemented for future work such as incorporating time-series analysis and potentially the addition of other datasets to improve the predictions. The project is still successful though and this validates the application of machine learning in football analytics and allows future fantasy football predictions. Therefore, this project enhances the decision-making process for fantasy football participants allowing for a more rewarding and engaging experience.

REFERENCES

[1] NFLverse Data Repository, "NFLverse data releases," [Online]. Available: https://github.com/nflverse/nflverse-data/releases?page=2. [Accessed: Dec. 17, 2024].USA: Abbrev. of Publisher, year, ch. *x*, sec. *x*, pp. *xxx–xxx*.

[2] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," [Online]. Available: https://github.com/dmlc/xgboost. [Accessed: Dec. 17, 2024].

[3] GeeksforGeeks, "ML | Gradient Boosting," [Online]. Available: https://www.geeksforgeeks.org/ml-gradient-boosting/. [Accessed: Dec. 17, 2024].

[4] J. Brownlee, "XGBoost for regression," [Online]. Available: https://machinelearningmastery.com/xgboost-for-regression/. [Accessed: Dec. 17, 2024].

Examples:

[5] GeeksforGeeks, "R-squared and adjusted R-squared in regression analysis," [Online]. Available: https://www.geeksforgeeks.org/r-squared-and-adjusted-r-squared-in-regression-analysis/. [Accessed: Dec. 17, 2024].

[6] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and É. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825-2830, 2011. [Online]. Available: https://scikit-learn.org/stable/documentation.html. [Accessed: Dec. 17, 2024]..

[7] Bleacher Nation, "What is PPR in fantasy football?" [Online]. Available: https://www.bleachernation.com/fantasy-football/2024/07/17/what-is-ppr-in-fantasy-football/#:~:text=PPR%20Scoring:%20Points%20Per%20Reception,each%20week%20to%20your%20lineup. [Accessed: Dec. 17, 2024].