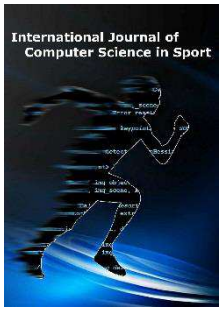
International Journal of Computer Science in Sport

Volume 24, Issue 1, 2025

Journal homepage: http://iacss.org/index.php?id=30

sciendo

DOI: 10.2478/ijcss-2025-0008




# Players' Performance Prediction for Fantasy Premier League, Using Transformer-based Sentiment Analysis on News and Statistical Data

*Tamimi, M., Tran, T.*

*University of Ottawa, School of Electrical Engineering and Computer Science, Canada*

## Abstract

Fantasy sports have become increasingly popular, with millions of players engaging in strategic team management and competition. In the realm of Fantasy Premier League (FPL), effective player analysis and performance prediction are crucial for success in each game. This paper presents an innovative approach to enhance FPL analysis and performance prediction by integrating news sentiment and players' injury with statistical data sources. A dataset of weekly news articles was enriched through pretrained transformer-based sentiment analysis toolkit and combined with different boosting and neural network algorithms for prediction tasks. Our findings demonstrate that integrating these features enhances model performance, with the CNN architecture achieving a reduction in MSE from 6.27 to 5.63 outperforming the state of the art model. These results highlight the potential of leveraging diverse data sources for more accurate predictions and informed decision-making in FPL.

KEYWORDS: FANTASY PREMIER LEAGUE (FPL), MACHINE LEARNING, NEURAL NETWORKS, SENTIMENT ANALYSIS, TRANSFORMERS

.

## Introduction

Sports analysis widely employs data science techniques to derive insights across various sports disciplines (Chmait and Westerbeek, 2021), given the vast amount of data generated during games and training sessions, including both individual and team-based disciplines such as tennis (Giles et al., 2020) and football (Ati et al., 2023). One critical insight derived from sports data is performance analysis, which plays an important role across all sports, particularly in team sports. It serves as an essential tool for managers and coaches (Wright et al., 2014), enabling informed decisions regarding player recruitment, dismissals, tactical adjustments, and game strategies. The ability to forecast player performances and assemble the most effective team for upcoming events is valuable to sports managers (Szymanski and Smith, 1997). This strategic selection process can help a team to win in a competition.

Fantasy Premier League (FPL) is a game platform where people can select their own team from all players in EPL (English Premier League—an English professional league for men's association football clubs—) and compete with each other in a leader board. FPL stands as the premier fantasy football game worldwide, boasting a record of over 11 million managers participating during the 2023/2024 EPL season (AllAboutFPL 2022; fantasy.premierleague.com). Accurately predicting a high-performing team can significantly benefit FPL players in winning the game. This capability could also be valuable for betting platforms. Additionally, since the prediction process is based on real data from EPL, it can assist team managers and coaches in making informed decisions about player acquisitions, substitutions, and team lineups. Crafting an optimal team within constraints—such as player positions and maximum roster size—transforms the team selection process into a performance prediction and optimization problem. This challenge has been explored in various studies in FPL (Gupta, 2019; Rajesh et al., 2022) as well as other fantasy sports like hockey and basketball (Beal et al., 2020; Hermann and Ntoso, 2015). Our aim in this paper was to provide a novel technique to predict players' performance for an upcoming event. This method holds potential utility in aiding team selection for FPL competitions.

The EPL dataset is a real-world publicly accessible resource that undergoes weekly updates. It encompasses fundamental statistics for each player, game-week-specific data, and the historical performance of players throughout the season. The majority of methodologies employed in this field leverage this dataset, employing various traditional Machine learning-based (ML-based—"ML-based" is used throughout the paper to refer to all ML methods that do not involve deep learning techniques—) and Deep Learning-based (DL-based) algorithms such as Linear Regression (LR), Decision Trees (DT), Random Forests (RF), Gradient Boosting (GB) Regression, Long Short Term Memory (LSTM), and Convolutional Neural Networks (CNN) (Bangdiwala et al., 2022; Frees et al., 2024; Gupta, 2019; Rajesh et al., 2022) to forecast players' performance.

Incorporating external sources beyond statistical data for sports analysis has been explored in several studies. Inputs such as match previews (Beal et al., 2021), news articles, and text generated around games have been utilized to enhance the understanding of sporting events and improve the accuracy of performance predictions. For example, Schumaker et al. (2016) demonstrated that analyzing sentiment in tweets could enhance spread and win predictions in the EPL. In FPL domain, research has similarly expanded beyond the confines of statistical datasets to include additional data sources. Bonello et al., (2019), for instance, integrated diverse data streams including betting odds, tweets, and sentiment scores from English blog posts. While they made these datasets available through different APIs (Application Programming Interfaces), most of which were not free, accessing their dataset proved challenging, rendering their methods non-reproducible. In contrast, Baughman et al., (2021) explored a wider array of

textual and non-textual data sources, employing techniques such as Document2Vector to incorporate these sources into the prediction process.

In the Premier League, football matches occur weekly, with player data provided in the EPL dataset for each game week. In this paper, we aimed to enhance predictive capabilities by incorporating textual and players' injury data alongside the EPL dataset. To achieve this, we collected a new dataset comprising news articles about each player and players' injury for every game week. We then conducted sentiment analysis on the textual data to generate additional features for predicting player performance in the subsequent week and determining their likelihood of participation in the next match. The news dataset was collected from the Google News, The Guardian (open-platform.theguardian.com), and GDELT (blog.gdeltproject.org/gdelt-doc-2-0-api-debuts) APIs, and we employed a pretrained transformer-based sentiment analysis toolkit for labeling sentiment in each article.

We have decided to integrate sentiment of news articles mentioning each player's name, gathered in the week leading up to each match, to enhance our predictive model. While statistical data provide essential player features for each game, incorporating sentiment of news enriches our dataset by offering insights into the positivity or negativity surrounding each player. Sentiment analysis represents a crucial aspect of Natural Language Processing (NLP), with various tools available that are trained on diverse text corpora for labeling new text as positive, negative, or neutral. In pursuit of this, we have chosen to employ pretrained transformer-based toolkits, for thier advanced capabilities. Transformer structures, renowned for their state-of-the-art design incorporating attention mechanisms (Vaswani et al., 2017), have demonstrated exceptional performance across a spectrum of NLP tasks.

In this paper, we present two methods. In the first method, our predictive tasks involved both regression, for predicting player performance, and classification, to determine if a player will play in the next game week. To address these challenges, we employed various boosting methods, including the GB (Algorithm Friedman, 2001), CatBoost (Dorogush et al., 2018), and XGBoost (Chen and Guestrin, 2016). Alongside news' sentiment, we extracted approximately 70 features from the EPL dataset.

In the second method, we expanded our dataset to include two sequential seasons using data from The Guardian and GDELT APIs. Additionally, we collected injury data for each player, creating new features for each week before each game. We applied various algorithms, including Ridge Regression (Hoerl and Kennard, 1970) , LightGBM (Ke et al., 2017), 1-dimensional CNN (O'shea and Nash, 2015), and LSTM (Hochreiter and Schmidhuber, 1997) architectures. We also compared our model with the state-of-the-art model proposed by Frees et al., (2024). To the best of our knowledge, our approach represents the first instance of employing a pretrained transformer-based sentiment analysis method to extract features from news articles in this domain.

In the upcoming sections of this paper, we will delve into related researches in Section 2. Following this, we will elaborate on our proposed methodologies, encompassing data collection and algorithmic approaches employed for these tasks in Section 3. Subsequently, we will look over the results and evaluations, then in the section 5 we will highlight the value of our work, and discusses its challenges and limitations. Finally, we will draw conclusions, and outline the future research direction in the conclusion section.

## Related Works

Numerous recent studies have focused on performance prediction for FPL, utilizing datasets provided by the FPL API. Additionally, some models have incorporated supplementary data

sources, such as tweets or news articles, and metadata on player injuries. These prediction tasks have been explored using both DL-based and ML-based models. A summary of some proposed methods in the literature provided in the Table 1.

Table 1. Comparison of proposed methods in Related Works.

| Refrence | EPL Data | Textual Data | Other Metadata | ML-based | DL-based |
|---|---|---|---|---|---|
| Bangdiwala et al., 2022; Hermann and Ntoso, 2015; Rajesh et al., 2022; Shah et al., 2023 | ✓ | ✗ | ✗ | ✓ | ✗ |
| Gupta, 2019; Lombu et al., 2024; Ramdas, 2022 | ✓ | ✗ | ✗ | ✓ | ✓ |
| Bonello et al., 2019 | ✓ | ✓ | ✓ | ✓ | ✗ |
| Baughman et al., 2021; Frees et al., 2024, **Our Model** | ✓ | ✓ | ✓ | ✓ | ✓ |

Bangdiwala et al., 2022 leveraged three different ML-based models including: LR, DT, and RF. They have considered features such as fixture difficulty, form of the two teams, creativity, and threat of the footballer. Another work provided in 2023 (Shah et al., 2023), considered different rules for team selection and used Multi Criteria Decision Analysis (MCDA) specifically TOPSIS (Technique for Order of Preference by Similarity to Ideal Solution) which consider both the positive and negative aspects of the alternatives in initial team selection. Subsequently, XGBoost Regressor is used to predict the points the team will obtain. Rajesh et al., (2022) employed RF and GB Regression to forecast players' performance across various positions in a team. In another domain, fantasy basketball team selection tackled by Hermann and Ntoso, (2015), they have used LR model based on box score statistics and modified multinomial naive Bayes classifier to this aim.

Several DL-based models have also been developed in this domain using the EPL dataset, Gupta, (2019) employed a combination of Auto-regressive Integrated Moving Average (ARIMA) with Recurrent Neural Networks (RNNs) and LSTM to predict player points as a team in time series data. Linear Programming was used to optimize the total points by selecting the most promising team based on the calculated predictions. In a research thesis, Ramdas, (2022) introduced a method that combines CNN and LSTM networks to predict players' performance. Another work in 2024, (Lombu et al., 2024) proposed a method and compared the performance of CNN and LSTM to predicting performances in FPL.

Incorporating other dataset and features rather than the ones provided by FPL API, Bonello et al., (2019) introduced a methodology for team selection in EPL, combining statistical analysis, historical data, and human sentiment. They employed GB Machine to determine player selection, utilizing inputs such as player historical data, FPL API data, betting odds, and sentiment extracted from Twitter and web articles. They accessed diverse datasets from various domains via APIs, although many of these APIs were not accessible for free at the time of this paper's writing, limiting their utilization. The datasets included betting odds obtained through the football-api (Rapid API), tweets extracted using the Tweepy library, and sentiment scores for English blog posts retrieved for the top 100 search results per query via the Aylien API. These data sources were utilized to analyze player performance in each previous match throughout the current football season. Another novel machine learning system designed to enhance ESPN Fantasy Football team management by analyzing vast amounts of natural language text and multimedia data proposed by researchers from IBM and Disney ESPN (Baughman et al., 2021). The system employs trained statistical entity detectors and DL models to understand and classify player performance trends. The system's training data includes web archives, ESPN statistics, and injury reports, with 2017 fantasy football data used for testing. Overall, the system aims to address the challenge of information overload in fantasy sports management and improve decision-making processes. This method, utilizes a ML pipeline involving statistical entity

detectors and document2vector models to comprehend natural language from over 50,000 sources. DL feed-forward neural networks are then employed to classify players.

The most recent approach in this domain was proposed by Frees et al., (2024). They introduced a novel method based on a 1-dimensional CNN and a transfer learning model that utilized news data about each player collected from The Guardian. Their results showed that the transfer learning model did not perform well, while the CNN architecture outperformed both the transfer learning method and other machine learning-based approaches.

Considering the innovative approaches in the domain of performance prediction for FPL, which include techniques leveraging only the EPL API dataset as well as those incorporating additional data sources, we introduce a novel method. Our approach integrates additional textual data collected for each game week, followed by sentiment analysis using transformer- based techniques.

## Methods

In this section, we provide an overview of the data utilized in our two methods. Next, we define the classification and regression models employed in this study. Finally, we discuss the evaluation methods applied to assess model performance.

### *Data*

This study utilized three distinct data sources: (1) the EPL dataset, containing historical match data; (2) a news dataset, gathered from various news APIs; and (3) an injury dataset, obtained using the WorldfootballR package (github.com/JaseZiv/worldfootballR).

#### *EPL Dataset*

EPL dataset is a publicly accessible open-source data (github.com/vaastav/Fantasy-Premier-League). This dataset comprises historical data collected from real football events in the EPL from the 2016-17 season onwards. It encompasses various statistics including overview statistics for a particular season, game-week-specific statistics for each season, player-specific statistics for each game week, and historical performance statistics for individual players.

The data for each player in every game week spanning the 2022/23 and 2023/24 seasons was downloaded and merged together. For the first method we preprocessed the EPL data to enrich the dataset, several new features were introduced. These additions were some statistics such as points, bonus points, goals, saves, and more, accumulated by each player across previous seasons. Additionally, a feature indicating the position of the player's team and their opponents on the final EPL table of the preceding season was incorporated. Another enhancement was the inclusion of a column representing the percentage of a player's value—derived from the FPL dataset, which reflects the player's cost to managers in the game and adjusts based on performance and transfers (March, 2024)—relative to their team's total value. This was calculated by dividing each player's value by the sum of all players' values on their team for each game week. Additionally, another column was added to indicate the number of players in the same position within their respective teams deemed more valuable. Unused columns, such as those containing "xP," "opponent-team," "expected-assists," and others, were omitted due to either excessive missing values. Further enhancements involved the creation of historical feature sets, where lists were utilized to store the last three values of specific player statistics across the preceding three gameweeks. These included metrics like assists, goals, and saves. Subsequently, the mean and standard deviation (stdev) of these historical features were calculated, providing deeper insights into player performance trends.

Our aim in the second method was twofold: first, to utilize a larger dataset, and second, to compare our model with the one proposed by Frees et al., (2024). To ensure a fair comparison, we replicated the preprocessing steps and used the same features of EPL dataset as those in Frees et al.'s model (github.com/danielfrees/mlpremier). For this method, we employed approximately 20 features provided in the EPL dataset and used the same data for training all models, collected from the 2022-23 and 2023-24 seasons.

### *News Dataset*

The data streams emerging from social media and news platforms offer a valuable window into societal perceptions and discussions surrounding various contexts or topics. Recognizing the potential insights embedded within these data streams, we decided to integrate them into our predictive model. Our reason behind this integration come from the understanding that while the majority of features within the EPL dataset are derived from match-related statistics, a plenty of external factors can influence player performance. These factors may include injuries sustained during or after matches, personal circumstances, or the overall sentiment circulating about a player within society.

To leverage this wealth of external information, we incorporated news articles generated about each player in the week leading up to their respective matches. Subsequently, we conducted sentiment labeling on these articles to measure the general sentiment surrounding each player. The sentiment related to players served as the foundation for generating a new feature, thereby enriching our model with additional insights derived from societal perceptions.

In the first method to collect news pertaining to each player, we first filtered the EPL dataset to include only data from the last month, focusing on players involved in games held in March 2024. Subsequently, we went through the EPL dataset, extracting the top 10 news articles in each previous week concerning each player in each game using the Google News API. As a result, for every player in every game, we obtained a set of 10 news articles from the preceding week. Collecting news from the Google News API had several limitations, including the inability to retrieve data beyond the past month. Therefore, for the second method's data collection, we had to use alternative sources. We employed The Guardian and GDELT APIs to obtain a set of news articles for each player for every game from the preceding week.

Once the data was compiled, we subjected each of the news articles pertaining to each player to extract their sentiment. This process involved assessing whether the sentiment expressed in each article was negative, positive, or neutral. Subsequently, we conducted a voting process among these sentiments for each player, aggregating the sentiment results to create a new feature. Therefore, we added one new feature from the Google News API for the first method, and two new features from The Guardian and GDELT APIs for the second method.

For extracting sentiment from news articles in the first method, we employed a pretrained transformer-based sentiment analysis toolkit (Pysentimiento (Pérez et al., 2021)) trained specifically on tweets. Due to the large scale of data in the second method, we decided to use the pretrained Sentiment Analysis Pipeline provided by Hugging Face (DeBERTa-v3-small-ft-news-sentiment-analisys). We utilized pretrained transformer-based models in both methods to label the sentiment of our news dataset. These models were selected to leverage the robust capabilities of transformers and attention mechanisms in NLP tasks, which are well-suited for sentiment analysis. Furthermore, given the absence of sentiment labels for the collected news data, these pretrained models provided an effective solution for accurately labeling sentiment. Moreover, collecting data for two seasons for the second method from both APIs and performing sentiment analysis on it took three weeks.

### *Injury Dataset*

For our second method, we created an additional feature named "injury." Using the WorldfootballR package (github.com/JaseZiv/worldfootballR), which offers functionality for gathering data from a variety of football data sources, including Transfermarkt (transfermarkt.com), Understat (understat.com), and FBref (fbref.com/en/). We retrieved player injury histories from Transfermarkt. Leventer et al. (2016) evaluated the data's validity and reliability. This data spans the 2022–23 and 2023–24 English Premier EPL seasons. Similar to the news data, we checked if each player was injured in the week preceding each game and created a boolean feature indicating the injury status for our second method.

## *Classification and Regression Models*

In this paper, we investigate two different approaches to data collection, preprocessing, and comparison. In the first method, we present a method capable of predicting the performance of each football player for upcoming events, as well as forecasting whether each player will participate in the forthcoming match. To achieve this, we preprocessed a small portion of the EPL dataset (one month), generating 70 features. Additionally, we collected news articles from the Google News API for the week preceding each match and extracted sentiments of the news using a transformer-based toolkit. These sentiment scores were then integrated into the EPL dataset, resulting in the creation of a new, enriched dataset. By applying various ML-based algorithms to this dataset, our model could predict player performances and determine player presence for upcoming games.

In the first method, we aimed to predict two key aspects: firstly, forecasting the points each player would arise in the next game, thereby framing the task as a regression problem. Secondly, we endeavored to predict whether each player would participate in the upcoming event or not, transforming this challenge into a classification problem. For the regression problem, we utilized the players' points as labels, while for the classification task, we assigned labels based on whether a player had participated in a match, with those playing over 10 minutes labeled as 1 and those playing less than 10 minutes labeled as 0. With these labels established for both regression and classification tasks, we proceeded to employ various boosting methods to tackle both challenges in the first method. We employed boosting algorithms for both tasks due to their strong performance in classification and regression and their feature-handling capabilities. Additionally, given the limited data available for each player's position, boosting methods were particularly advantageous as they can enhance predictive performance while mitigating bias in models. Notable algorithms utilized included the GB Algorithm (Friedman, 2001), Catboost (Dorogush et al., 2018), and XGBoost (Chen and Guestrin, 2016). By harnessing the capabilities of these boosting techniques, we aimed to develop robust models capable of accurately predicting player performance in terms of points gained and their likelihood of participation in upcoming matches.

The first method highlighted the importance of incorporating additional data to enhance prediction accuracy. Consequently, we built the second method. Recognizing the impact of adding external data beyond the EPL dataset, we used a larger EPL dataset spanning two seasons and augmented it with additional data from various sources. Specifically, we utilized The Guardian and GDELT APIs to collect news about each player for the week preceding each game, followed by sentiment labeling to evaluate overall feedback about each player. We also gathered weekly player injury data and integrated these features into our data. By incorporating these features, we analyzed their influence on predicting the points gained by each player each week. We employed different ML-based algorithms, CNN and LSTM architectures, to predict player points for each game. The processed dataset, when used with the CNN architecture, yielded best

performance.

For the second method, we focused on the regression task, aiming to predict the points each player would gain in each game. We employed various regression methods to determine if including textual and injury features could enhance model performance. Specifically, we used Ridge regression (Hoerl and Kennard, 1970), LightGBM (Ke et al., 2017), 1-dimensional CNN (O'shea and Nash, 2015), and LSTM (Hochreiter and Schmidhuber, 1997) architectures. The implementation of Ridge regression, LightGBM, and CNN was consistent with the work proposed by Frees et al., (2024) (github.com/danielfrees/mlpremier). However, our implementation introduced three new features generated from additional sources, resulting in improved performance. Following Frees et al.'s (2024) methodology, we also considered Ridge Regression and LightGBM as baseline models. For advanced approaches, we employed CNN and LSTM architectures. Since our primary objective was to compare our model's performance with Frees et al.'s work, we adopted similar methodologies and model selections in the second method.

To investigate the capability of the sequential nature of the data, we decided to use an LSTM architecture. We used the same features as those used for other algorithms in the LSTM architecture. During preprocessing, we created data sequences of varying lengths for each player, assigning the next week's point as labels to the previous week's data to maintain sequentiality. We tested our model with different sequence lengths, number of layers, learning rate, hidden dimension, epochs, and dropout rates. Additionally, we assessed the impact of incorporating textual and injury datasets on the model's performance and compared the results with other algorithms.

### *Evaluation Method*

For evaluation purposes, we aim to assess the impact of integrating news sentiments on the performance of both classification and regression tasks. To achieve this goal, we conducted experiments with two versions of the EPL dataset: one containing sentiment and injury features and another without. We employed the same classification and regression algorithms mentioned earlier on both datasets. The objective was to compare the predictive capabilities of models trained on datasets with and without additional features.

In the first method, to ensure a robust evaluation, we adopted a train-test splitting approach consistent with some previous researches, such as studies by Rajesh et al. (2022) and Baughman et al. (2021). Following prior works, we sorted the data by date and selected the last 10 percent of the dataset as the testing set, with the remaining data allocated to the training set. By analyzing the results obtained from both versions of the dataset, we aimed to gain insights into the efficacy of incorporating news sentiments in enhancing the predictive performance of classification and regression models in the context of FPL analysis.

In the second method, the dataset was split into training, validation, and test sets in proportions of 60%, 25%, and 15%, respectively. To ensure robust results, we applied 3-fold cross-validation to all algorithms, including the baselines, LSTM, and CNN models.

For the CNN models, we began with a grid search to tune hyperparameters on two datasets—one incorporating the new features and the other without. A random state of 42 was applied. The grid search focused on the following hyperparameters: *Window sizes*: 3, 6, 9; *Kernel sizes*: 1, 2, 3, 4; "*Drop low playtime*" metric: Binary values (True/False); *Stratification*: Based on player skill and standard deviation (stdev) scores; *Number of filters*: 64; *Dense layer size*: 64; and *Activation function*: ReLU. These parameters were selected following the methodology outlined by Frees et al. (2024). Due to computational constraints on Google Colab, we streamlined the grid search process in subsequent iterations, reducing the search space to focus on a smaller

subset of hyperparameters. After the initial tuning in the firsr step, we refined the hyperparameters based on the validation set for the next step, performing both grid search and cross-validation. The final subset of optimal hyperparameters for grid search included: *Window sizes*: 3 and 9; *Kernel sizes*: 1 and 2; *Drop low playtime*: True; and *Stratification*: stdev. Separate CNN models were developed for each player position (GK, DEF, MID, FWD), with hyperparameter tuning. For example, in one iteration of cross-validation on the dataset with additional features, the selected hyperparameters were: *Window size*: 3 for GK and MID; 9 for DEF and FWD; and *Kernel size*: 1 for GK and FWD; 2 for DEF and MID.

Cross-validation were performed for baseline*s,* LSTM *and* CNN models, ensuring consistency across train-test splits. The CNN implementation and preprocessing adhered to Frees et al. (2024), with adaptations to address resource limitations by narrowing the grid search space. For CNN models performance metrics were recorded for each iteration of cross-validation, with test results reported for the best hyperparameters derived from the validation set.

### *Evaluation Metrics*

This paper aims to assess the performance of both regression and classification tasks in first method, and regression task in second method, with and without the inclusion of additional features, across various algorithms using a comprehensive set of evaluation metrics.

For the classification task, we will report two key metrics: accuracy and F1 score. Accuracy measures the overall correctness of predictions, while the F1 score balances precision and recall to provide a assessment of classification performance. Precision evaluates the accuracy of positive predictions, while recall assesses the model's ability to identify all instances of positive predictions. These metrics, when analyzed together, offer valuable insights into the classification task's effectiveness.

In the regression task, we will report Mean Squared Error (MSE) and Root Mean Squared Error (RMSE). MSE quantifies the average squared difference between actual (observed) values and predicted values, providing a measure of the predictive model's performance. RMSE, the square root of MSE, offers a more intuitive understanding of the model's predictive accuracy.

## Results

In this section, we examine the impact of incorporating sentiment analysis on news articles and adding players' injury data as additional features alongside the EPL dataset. We will report the evaluation metrics for all mentioned algorithms utilized in both classification and regression tasks, focusing on how these additional features affect performance.

Table 2. Performance of Algorithms Trained on Datasets with and witout News Sentiments, Classification Task.

| Algorithm/First Method | Accuraccy% | F1 score% |
|---|---|---|
| Baseline | **86.22** | **82.35** |
| Catboost | 84.45 | 76.59 |
| Catboost+News Sentiment | **85.15** | **77.89** |
| XGBoost | 83.74 | 75.78 |
| XGBoost+News Sentiment | 84.09 | 75.93 |
| GB | 84.09 | 76.43 |
| GB+News Sentiment | 83.39 | 74.86 |

### *First Method, Influence of News Sentiments on Classification Task*

In the classification task, our objective was to predict whether a player will participate in the

next match or not. Table 2 presents the Accuracy and F1 score for different algorithms trained on datasets with and without news sentiments. We also included a baseline method, assuming that same players will play again. This simplistic baseline assumes that a player who participated in the previous match will also participate in the next match.

As shown in Table 2, the baseline achieved the highest F1 score and accuracy among all methods. This performance is driven by the baseline's recall, which is perfect (1.0), indicating that it predicts all actual participants correctly. However, its precision is lower (0.7000), as it overpredicts participation. For example, the best-performing boosting model, CatBoost+News Sentiment, achieved a recall of 0.8131 and a precision of 0.7474. Boosting algorithms can considered in this problem because they move beyond the simplistic assumptions of the baseline and can identify more complex patterns in the data. While the baseline captures all players who participate, it lacks precision and includes many false positives. Boosting models strike a better balance by reducing false positives while maintaining recall, resulting in more reliable predictions. Additionally, these models leverage multiple features to uncover trends that the baseline model cannot detect. The other explanation for the performance gap between the baseline and the boosting models lies in the imbalanced nature of the dataset. In most games, only a subset of players participate, meaning that the majority class consists of players who do not play. This imbalance may have caused the boosting models to struggle in correctly identifying all participants. Addressing this imbalance could further improve the performance of these models.

Considering boosting algorithms, the Catboost algorithm outperformed other algorithms for both datasets. This performance can be attributed to CatBoost's optimization for training on large datasets and its efficient handling of high-dimensional feature spaces. Conversely, XGBoost and GB algorithms may encounter performance challenges or require additional tuning when dealing with datasets containing a large number of features.

Importantly, the inclusion of news sentiment features in the dataset yielded improvements in performance metrics for Catboost and among all algorithms. Specifically, the accuracy increased from 84.45% to 85.15%, and the F1 score improved from 76.59% to 77.89%. This outcome underscores the significance of incorporating news sentiments features into the EPL dataset. Despite the dataset's high dimensionality, integrating news sentiment data facilitated enhanced algorithm performance. These findings underscore the utility of incorporating additional features, particularly those derived from external sources like news sentiment, in improving predictive modeling outcomes.

Table 3. Players' Presence Prediction Using Catboost for Different Postions, with and witout News Sentiment.

| Players Position | Accuraccy% | F1 score% |
|---|---|---|
| GK | 87.87 | 74.99 |
| GK+News Sentiment | **95.23** | **87.71** |
| DEF | 77.52 | 73.68 |
| DEF+News Sentiment | **80.00** | **73.84** |
| MID | 86.66 | 80.85 |
| MID+News Sentimen | **92.66** | **88.57** |
| FWD | 76.92 | 62.50 |
| FWD+News Sentiment | **92.50** | **82.35** |

Another comparison was conducted to assess the performance of the Catboost algorithm in classifying whether each player would participate in the next game or not. This evaluation was based on the validation set, which was derived from selecting 10 percent of the training data. The comparison of Catboost performance for different player positions is presented in Table 3. Table 3 illustrate that including news sentiment for prediction significantly impacted the

accuracy and F1 score of the model. Table 3 demonstrates that incorporating news sentiment resulted in improved prediction accuracy across all player positions. This finding underscores the importance of integrating news sentiment data into the predictive modeling process, as it enhances the model's ability to accurately predict player participation in upcoming matches across various positions.

### *First Method, Influence of News Sentiments on Regression Task*

As previously mentioned, the regression task in this paper involves predicting the points each player will accumulate in the upcoming match. To ensure meaningful predictions, we filtered the dataset to include only players who had participated in each match for more than 10 minutes. The results of training the three mentioned algorithms are presented in Table 4, illustrating the Mean Squared Error (MSE) and Root Mean Squared Error (RMSE). Table 4 highlights that Catboost outperformed the other algorithms in the regression task, similar to its performance in the classification task.

RMSE, which represents the difference between the actual and predicted points collected by each player, offers a more intuitive understanding of prediction accuracy compared to MSE. Analyzing the RMSE metric in Table 4 reveals minimal differences across all algorithms with and without the inclusion of news sentiment. While Catboost exhibited better performance in terms of RMSE without incorporating news sentiment, this outcome may be attributed to the dataset's reduced size after filtering only played players. Consequently, the impact of news sentiment on the regression task may have been diminished.

Table 4. Performance of Algorithms Trained on Datasets with and without News Sentiments, Regression Task.

| Algorithm/First Method | MSE | RMSE |
|---|---|---|
| Catboost | **7.27** | **2.69** |
| Catboost+News Sentiment | 7.51 | 2.74 |
| XGBoost | 7.46 | 2.73 |
| XGBoost+News Sentiment | 7.40 | 2.72 |
| GB | 7.37 | 2.71 |
| GB+News Sentiment | 7.74 | 2.78 |

### *Second Method, Influence of News Sentiments and Players' injury on Regression Task*

As previously mentioned, we added three new features to the EPL dataset and utilized various algorithms, including ML-based and DL-based approaches, to predict each player's performance in upcoming events. For all the algorithms, we analyzed the impact of incorporating these new features on prediction performance and compared the results with the performance of models trained on the original dataset.

Table 5 illustrates the MSE metric for all the algorithms, including Ridge Regression, LightGBM, CNN, and LSTM, which were trained on datasets with and without news sentiment and player injury features. Our objective was to compare the influence of our newly engineered features on the prediction performance of player points in each match. The first comparison in Table 5 could be between each algorithm with and without the added features. As shown, in most cases, the addition of the engineered features performed equally to the models without the new features. However, when comparing CNN and CNN+F—+F indicates training algorithms with the additional news and injury features—, a significant difference can be observed in the Goalkeeper (GK) prediction. This indicates that the new engineered features improved the model's performance. For other positions, CNN+F generally performed better than CNN without

the added features. On average, CNN+F showed strong performance with an MSE of 5.63 and an RMSE of 2.37, compared to its main competitor, CNN, which had an MSE of 6.27 and an RMSE of 2.50. It is noteworthy that we report the best value for each model, with stratification based on skill for CNN and based on stdev for CNN+F. Additionally, examining the table reveals that using a CNN structure for this task generally yields better results. Most algorithms perform similarly for the Defender (DEF) position. Moreover, while LSTM performed well for the GK position in both versions, it did not perform as well for other positions compared to other algorithms. This suggests that further investigation into feature engineering for the LSTM model may lead to improved results.

Table 5. MSE for Algorithms Trained on Datasets with and without Additional Features, Second Method. +F indicates training algorithms with the additional news and injury features.

| Players Position | Ridge | Ridge+F | LightGBM | LightGBM+F | CNN (Frees et al., 2024) | CNN+F | LSTM | LSTM+F |
|---|---|---|---|---|---|---|---|---|
| **GK** | 7.61 | 7.61 | 7.54 | 7.47 | 7.34 | 4.89 | **3.61** | **3.61** |
| **DEF** | **5.51** | **5.51** | **5.51** | **5.51** | 5.71 | **5.51** | 7.03 | 6.97 |
| **MID** | 6.05 | 6.06 | 5.95 | 5.94 | **5.46** | 5.92 | 7.02 | 7.08 |
| **FWD** | 9.09 | 9.07 | 9.21 | 9.24 | 6.56 | **6.20** | 10.16 | 11.20 |
| **Average** | 7.06 | 7.06 | 7.05 | 7.04 | 6.27 | **5.63** | 6.95 | 7.21 |

## Discussion

This study presented a solution for predicting player presence and performance in FPL by augmenting the EPL dataset with external features. The integration of news sentiment and player injury data highlights the importance of leveraging external sources to complement the statistical dataset. Similar studies have shown that incorporating external data can enrich sports prediction models by providing valuable insights into player and team dynamics (Bonello et al., 2019; Baughman et al., 2021, Beal et al., 2021). Our findings demonstrate the potential of incorporating these features to enhance predictive models. Moreover, our approach illustrates the value of leveraging textual data through advanced NLP techniques, such as pretrained transformer-based models. These methods allow for the extraction of insights from unstructured text, which statistical features alone cannot provide. While our approach performed well within the specific context of FPL, its generalizability to other fantasy sports platforms or sports prediction tasks remains untested.

The inclusion of news sentiment improved model performance in the classification task in the first method. Specifically, CatBoost achieved a high performance. These improvements underscore the utility of integrating contextual external data, consistent with findings from Schumaker et al. (2016), who highlighted the importance of sentiment analysis in sports prediction. However, despite these gains, the simplistic baseline model achieved the highest accuracy and F1 score due to its perfect recall. While the baseline captured all participants, it suffered from low precision, indicating a high rate of false positives.

The dataset's inherent imbalance—where the majority of players do not participate in a given match—may have contributed to the performance gap observed between the baseline and advanced models in the classification task. Incorporating balancing techniques, could improve model reliability. Moreover, the high performance of the baseline raises questions about the added value of more complex algorithms. While boosting models demonstrated better precision by reducing false positives, the baseline's simplicity and recall advantage highlight the need to evaluate the trade-offs between model complexity and accuracy. A broader range of baselines to contextualize the performance of advanced models could be considered.

In the first method's regression task, we observed minimal performance differences between models trained with and without news sentiment. One potential explanation is the filtering process, which reduced the dataset size. This reduction may have limited the model's ability to utilize the added features effectively. Nonetheless, the results suggest that while sentiment features provide benefits for classification tasks, their impact on regression tasks vary depending on the dataset size and the nature of the target variable. To further explore this effect, we implemented the second method using a larger dataset to better assess the influence of the additional features on model performance.

In the second method, the addition of news sentiment and player injury features had mixed effects across algorithms and player positions. For instance, CNN+F demonstrated better performance compared to CNN alone (Frees et al., 2024). Furthermore, the addition of these features enabled CNN+F to achieve the best average performance across all player positions, highlighting the significance of these features in enhancing model performance and surpassing the state-of-the-art model proposed by Frees et al. (2024).

This study faced several challenges, primarily related to data access and dataset size. Data collection posed significant challenges due to API restrictions and access limitations. For instance, the inability to access Twitter data after February 2023 forced us to rely on alternative sources such as Google News, The Guardian, and GDELT APIs. These APIs imposed their own constraints, including request limits and restricted retention periods, complicating the process of gathering consistent and high-quality data especially in the first method using Google News API. These limitations influenced the volume and variety of data available for analysis, potentially constraining the generalizability of our findings.

The limited size of our dataset, particularly in the first method, posed challenges for training both boosting models in regression task and DL models like LSTM. Neural networks are known to require extensive datasets to effectively capture sequential patterns and temporal dependencies (Hochreiter & Schmidhuber, 1997). However, our dataset, covering only one month, was insufficient to train the LSTM models to their full potential, resulting in suboptimal performance. This limitation underscores the importance of obtaining larger and more diverse datasets for future studies.

## Conclousion and Future Works

In conclusion, this study underscores the critical role of incorporating external data sources, such as news sentiment and players' injury, in enhancing predictive modeling tasks in the domain of FPL. By augmenting the EPL dataset with new additional features, we observed improvements in model performance for both classification and regression tasks. Our findings highlight the importance of diversifying data sources to create more comprehensive datasets, which ultimately contribute to more accurate predictions. In the future, research efforts could be directed towards exploring additional sources of external data and refining predictive models to enhance performance in FPL analysis. These additional resources may include tweets from Twitter, weather data, injury reports, and social media posts authored by each player. As a potential future direction, acquiring a Twitter API account could allow for the collection of more accurate and diverse data from Twitter over several months or years, thereby expanding the dataset. Another research direction could involve further investigation into feature engineering, specifically for LSTM, to better utilize its capabilities in time series problems. Moreover, given the performance of CNN and LSTM, developing a method to leverage a combination of these two models could be beneficial.

## References

AllAboutFPL. (2022, June). Number of people who played FPL each season: How many play FPL? https://allaboutfpl.com/2022/06/number-of-people-who-played-fpl-each-season-how-many-play-fpl/

Ati, A., Bouchet, P., & Ben Jeddou, R. (2024). Using multi-criteria decision-making and machine learning for football player selection and performance prediction: a systematic review. *Data Science and Management*, *7*(2), 79–88. https://doi.org/10.1016/j.dsm.2023.11.001

Bangdiwala, M., Choudhari, R., Hegde, A., & Salunke, A. (2022). Using ML Models to Predict Points in Fantasy Premier League. *2022 2nd Asian Conference on Innovation in Technology (ASIANCON)*, 1–6. https://doi.org/10.1109/ASIANCON55314.2022.9909447

Baughman, A., Forester, M., Powell, J., Morales, E., McPartlin, S., & Bohm, D. (2021). *Deep Artificial Intelligence for Fantasy Football Language Understanding*. http://arxiv.org/abs/2111.02874

Beal, R., Middleton, S. E., Norman, T. J., & Ramchurn, S. D. (2021). Combining Machine Learning and Human Experts to Predict Match Outcomes in Football: A Baseline Model. *Proceedings of the AAAI Conference on Artificial Intelligence*, *35*(17), 15447–15451. https://doi.org/10.1609/aaai.v35i17.17815

Beal, R., Norman, T. J., & Ramchurn, S. D. (2020). Optimising Daily Fantasy Sports Teams with Artificial Intelligence. *International Journal of Computer Science in Sport*, *19*(2), 21–35. https://doi.org/10.2478/ijcss-2020-0008

Bonello, N., Beel, J., Lawless, S., & Debattista, J. (2019). *Multi-stream Data Analytics for Enhanced Performance Prediction in Fantasy Football*. https://doi.org/10.48550/arXiv.1912.07441

Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. *Proceedings of the 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining*, 785–794. https://doi.org/10.1145/2939672.2939785

Chmait, N., & Westerbeek, H. (2021). Artificial Intelligence and Machine Learning in Sport Research: An Introduction for Non-data Scientists. *Frontiers in Sports and Active Living*, *3*, 363. https://doi.org/10.3389/fspor.2021.682287

Dorogush, A. V., Ershov, V., & Gulin, A. (2018). *CatBoost: gradient boosting with categorical features support*. https://doi.org/10.48550/arXiv.1810.11363

Frees, D., Ravella, P., & Zhang, C. (2024). *Deep learning and transfer learning architectures for english premier league player performance forecasting*. https://doi.org/10.48550/arXiv.2405.02412

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, *29*(5), 1189–1232. https://doi.org/10.1214/aos/1013203451

Giles, B., Kovalchik, S., & Reid, M. (2020). A machine learning approach for automatic detection and classification of changes of direction from player tracking data in professional tennis. *Journal of Sports Sciences*, *38*(1), 106–113. https://doi.org/10.1080/02640414.2019.1684132

Gupta, A. (2019). *Time Series Modeling for Dream Team in Fantasy Premier League*. https://doi.org/10.48550/arXiv.1909.12938

Hermann, E., & Ntoso, A. (2015). *Machine learning applications in fantasy basketball*. https://api.semanticscholar.org/CorpusID:15791576

Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, *9*(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, *12*(1), 55–67. https://www.tandfonline.com/doi/abs/10.1080/00401706.1970.10488634

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T.-Y. (2017). Lightgbm: A highly efficient gradient boosting decision tree. *Advances in Neural Information Processing Systems*, *30*. https://proceedings.neurips.cc/paper_files/paper/2017/file/6449f44a102fde848669bdd9eb6b76fa-Paper.pdf

Leventer, L., Eek, F., Hofstetter, S., & Lames, M. (2016). Injury Patterns among Elite Football Players: A Media-based Analysis over 6 Seasons with Emphasis on Playing Position. *International Journal of Sports Medicine*, *37*(11), 898–908. https://doi.org/10.1055/s-0042-108201

Lombu, A. S., Paputungan, I. V, & Dewa, C. K. (2024). Predicting fantasy premier league points using convolutional neural network and long short term memory. *Jurnal Teknik Informatika (Jutif)*, *5*(1), 263–272. https://jutif.if.unsoed.ac.id/index.php/jurnal/article/view/1792

March, S. (2024). Ex-winner Simon March: What does value really mean in FPL? *Fantasy Football Scout*. https://www.fantasyfootballscout.co.uk/2024/09/11/ex-winner-simon-march-what-does-value-really-mean-in-fpl

O'Shea, K., & Nash, R. (2015). *An Introduction to Convolutional Neural Networks*. http://arxiv.org/abs/1511.08458

Pérez, J. M., Rajngewerc, M., Giudici, J. C., Furman, D. A., Luque, F., Alemany, L. A., & Martínez, M. V. (2021). *pysentimiento: A Python Toolkit for Opinion Mining and Social NLP tasks*. http://arxiv.org/abs/2106.09462

Rajesh, V., Arjun, P., Jagtap, K. R., M, S. C., & Prakash, J. (2022). Player Recommendation System for Fantasy Premier League using Machine Learning. *2022 19th International Joint Conference on Computer Science and Software Engineering (JCSSE)*, 1–6. https://doi.org/10.1109/JCSSE54890.2022.9836260

Ramdas, D. (2022). *Using convolution neural networks to predict the performance of footballers in the fantasy premier league*. https://doi.org/10.13140/RG.2.2.10010.72645/2

Schumaker, R. P., Jarmoszko, A. T., & Labedz, C. S. (2016). Predicting wins and spread in the Premier League using a sentiment analysis of twitter. *Decision Support Systems*, *88*, 76–84. https://doi.org/10.1016/j.dss.2016.05.010

Shah, D., Kapadia, P., Kakwani, H., & Dabre, K. (2023). Multi Criteria Decision Making in Fantasy Sports. *2023 IEEE Pune Section International Conference (PuneCon)*, 1–6. https://doi.org/10.1109/PuneCon58714.2023.10450054

Szymanski, S., & Smith, R. (1997). The English Football Industry: profit, performance and industrial structure. *International Review of Applied Economics*, *11*(1), 135–153. https://doi.org/10.1080/02692179700000008

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention Is All You Need. *Advances in Neural Information Processing Systems*, *30*. https://doi.org/10.48550/arXiv.1706.03762

Wright, C., Carling, C., & Collins, D. (2014). The wider context of performance analysis and it application in the football coaching process. *International Journal of Performance Analysis in Sport*, *14*(3), 709–733. https://doi.org/10.1080/24748668.2014.11868753