A Data-Driven Analysis of NFL Quarterbacks

Jayson Rhea
School of Computing
Southern Adventist University
Collegedale, TN 37315

***Abstract*—Fantasy football has become a focal point of interest for both fans and teams alike. The NFL employs a ranking system that aids teams and managers in assessing the quality of their chosen quarterbacks (QB) and predicting their performance. This experiment seeks to enhance the accuracy and predictability of player recommendations (Highly Recommended - HR, Recommended - R, Not Recommended - NR) for fantasy football teams by utilizing NFL QB player data. While existing analyses exist for player classification, this study augments these analyses with additional data to provide more precise player representations based on specific statistics.**

The primary objective of this study is to provide valuable insights into the worth of NFL players, benefiting both team owners and fantasy football enthusiasts. Of particular interest is the assessment and maximization of the potential value of quarterbacks, who are widely regarded as one of the most critical assets to a football team. What sets this research apart is its incorporation of efficiency quantifiers, including completion percentage and yards per completion, to provide a more comprehensive evaluation of player performance.

I. Introduction

Fantasy football has transformed the way fans engage with the sport, turning casual spectators into strategic team managers. In this dynamic landscape, the nfl's ranking system has emerged as a critical tool, assisting team owners and managers in evaluating the quality of quarterbacks (QBs) and projecting their on-field performance. However, the constant pursuit of improving predictions and player recommendations continues, and this study embarks on that journey by leveraging NFL QB player data. Our goal is to refine the accuracy of recommendations for fantasy football teams, classifying QBs as Highly Recommended (HR), Recommended (R), or Not Recommended (NR) based on their statistical profiles. While established analyses have laid the groundwork for player classification, this research introduces additional variables to create a more precise and nuanced representation of each player's potential.

II. Exploring the Data

A. Data Collection and Preparation

The data set was retrieved from an online source[1] and consisted of 83 quarterbacks' data for the 2021 season. Data cleaning involved removing unnecessary variables and rows with missing values.

B. Methodology

- Descriptive statistics were used to summarize the data.
- Quarterbacks were classified into three categories (HR, R, DNR) based on rankings.
- Efficiency metrics such as completion percentage and yards per completion were calculated to enhance precision.
- The data was scaled to reduce dimensions and allow for faster convergence during machine learning.

C. Machine Learning

The following classifiers were used in this project:

- k-Nearest Neighbors
- Decision Tree Classifier
- Logistics Regression
- Support Vector machine
- Artificial Neural Network

The data was augmented so that the class distribution could be balanced in the classification report.

III. Machine Learning Statistics

A. k-Nearest Neighbor

	Precision	Recall	F1-Score	Support
DNR	0.62	0.93	0.74	14
HR	1.00	0.64	0.78	14
R	0.5	0.43	0.46	14
Accuracy			0.67	42
Macro Avg	0.71		0.67	0.66	42
Weighted Avg	0.71	0.67	0.66	42

TABLE I
k-NN

The KNN model (Table 1) excels in correctly identifying "Highly Recommended QBs" (HR), achieving a perfect precision of 1.00 and a high F1-score of 0.78 for this class. However, it encounters challenges when identifying "recommended QBs" (R), with a precision of 0.50 and a relatively low F1-score of 0.46. The report overall suggests that while the KNN model demonstrates strength in certain areas, further improvements may be needed to enhance its accuracy, particularly in correctly classifying "recommended QBs."

	Precision	Recall	F1-Score	Support
DNR	1.00	1.00	1.00	14
HR	1.00	1.00	1.00	14
R	1.00	1.00	1.00	14
Accuracy			1.00	42
Macro Avg	1.00	1.00	1.00	42
Weighted Avg	1.00	1.00	1.00	42

TABLE II
DECISION TREE MODEL

	Precision	Recall	F1-Score	Support
DNR	0.67	1.00	0.80	14
HR	1.00	1.00	1.00	14
R	1.00	0.50	0.67	14
Accuracy			0.83	42
Macro Avg	0.89	0.83	0.82	42
Weighted Avg	0.89	0.83	0.82	42

TABLE IV
SUPPORT VECTOR MODEL

B. Decision Tree

Remarkably, the Decision Tree Model (Table 2) achieves perfect scores of 1.00 across all metrics for each of the three classes: "DNR," "HR," and "R," as well as for accuracy, macro-average, and weighted-average scores. While these high scores might initially suggest excellent model performance, the report rightly points out that such perfect results across a relatively small data set, represented by a support value of 14 for each class, are indicative of over-fitting. Over-fitting occurs when a model has learned the training data so well that it struggles to generalize to unseen data. Therefore, it is essential to consider the data set's size and diversity, as these perfect scores may not reflect the model's true performance in larger, more varied data sets.

C. Logistics Regression

	Precision	Recall	F1-Score	Support
DNR	0.72	0.93	0.81	14
HR	1.00	1.00	1.00	14
R	0.9	0.64	0.75	14
Accuracy			0.86	42
Macro Avg	0.87	0.86	0.85	42
Weighted Avg	0.87	0.86	0.85	42

TABLE III
LOGISTICS REGRESSION MODEL

The data from table 3 demonstrates strong predictive capabilities, particularly in identifying "Highly Recommended QBs" (HR), achieving a perfect precision, recall, F1-score, and support of 1.00 for this class. Additionally, it exhibits respectable performance in classifying "Do Not Recommend QBs" (DNR) with a precision of 0.72 and an F1-score of 0.81. However, it faces challenges in correctly classifying "Recommended QBs" (R), with a precision of 0.90 but a lower recall of 0.64, leading to an F1-score of 0.75. Overall, the model achieves a high accuracy of 0.86, indicating its overall effectiveness. The report underscores the Logistic Regression model's near-perfect predictability for HR while acknowledging that correctly predicting R remains a more intricate task.

D. Support Vector Model

The support vector model (Table 4) demonstrates a robust ability to identify "Do Not Recommend QBs" (DNR), achieving a precision of 0.67 and an F1-score of 0.80, and it excels in classifying "Highly Recommended QBs" (HR) with perfect precision, recall, and F1-score, reflecting flawless performance for this class. However, the model faces challenges in distinguishing "Recommended QBs" (R), as indicated by a perfect precision but a lower recall, resulting in an F1-score of 0.67 for this category. Overall, the model achieves a commendable accuracy of 0.83, highlighting its strong predictive capability. The macro and weighted averages, which consider overall performance across all classes, reinforce the model's reliability, with macro-average values of 0.89 for precision, 0.83 for recall, and 0.82 for the F1-score, while weighted-average values align closely, each at 0.89, 0.83, and 0.82, respectively. These metrics collectively suggest that the Support Vector Model performs well across the dataset, with particular strength in identifying HR, though it exhibits some limitations in correctly classifying R.

E. Artificial Neural Networks

	Precision	Recall	F1-Score	Support
DNR	0.67	1.00	0.80	14
HR	1.00	1.00	1.00	14
R	1.00	0.50	0.67	14
Accuracy			0.83	42
Macro Avg	0.89	0.83	0.82	42
Weighted Avg	0.89	0.83	0.82	42

TABLE V
ARTIFICIAL NEURAL NETWORKS

Lastly, the artificial neural networks model (Table 5) exhibits strengths and challenges in its classification abilities. Notably, it achieves high precision and recall for "Do Not Recommend QBs" (DNR), with a precision of 0.82 and recall of 1.00, resulting in an impressive F1-score of 0.90 for this category. However, for "Highly Recommended QBs" (HR), while maintaining perfect recall, the model struggles with precision, yielding an F1-score of 0.74. It faces significant difficulties in classifying "Recommended QBs" (R), with perfect precision but an extremely low recall, resulting in a low F1-score of 0.13. Overall, the model achieves an accuracy of 0.69, indicating its moderate performance across all classes. The macro and weighted averages, which provide an overview of overall model performance, show moderate scores, highlighting the trade-offs between precision and recall. While excelling in identifying DNR, the ANN model encounters challenges in classifying HR and particularly in recognizing R, suggesting potential areas for improvement and further tuning.

IV. CONCLUSION

In the world of fantasy football and team management, the value of informed decision-making cannot be overstated. This study has sought to bridge the gap between data and actionable insights, enhancing the ability of team owners and fantasy football enthusiasts to make strategic choices. By delving into NFL QB player data and enriching the classification process with efficiency metrics such as completion percentage and yards per completion, we have contributed to a more robust understanding of player worth. The quarterback, often considered the linchpin of a football team, holds immense significance, and our study has provided a refined lens through which to assess their contributions.

In this comprehensive data-driven analysis of NFL quarterbacks, the study has addressed the pressing issue of improving the accuracy and predictability of player recommendations for fantasy football teams. By leveraging NFL QB player data, this research contributes valuable insights to both team owners and fantasy football enthusiasts. What sets this study apart is its incorporation of efficiency quantifiers such as completion percentage and yards per completion, which provide a more comprehensive evaluation of player performance.

The exploration of the data involved meticulous data collection, cleaning, and preparation, followed by an extensive methodology involving descriptive statistics, player classification, and the calculation of efficiency metrics. Machine learning models, including k-Nearest Neighbors, Decision Tree Classifier, Logistics Regression, Support Vector Machine, and Artificial Neural Network, were employed to enhance classification accuracy.

The analysis of the machine learning models revealed important findings and trade-offs. For instance, the k-Nearest Neighbors model demonstrated strength in identifying ”Highly Recommended QBs” but faced challenges in classifying ”Recommended QBs.” The Decision Tree Model, while achieving perfect scores, indicated potential overfitting due to the relatively small dataset. The Logistics Regression Model displayed strong predictive capabilities, particularly for ”Highly Recommended QBs,” but encountered difficulties in classifying ”Recommended QBs.” The Support Vector Model demonstrated robustness in identifying ”Do Not Recommend QBs” and ”Highly Recommended QBs,” with challenges in classifying ”Recommended QBs.” The Artificial Neural Networks model showed strengths in classifying ”Do Not Recommend QBs” but struggled with ”Recommended QBs,” indicating potential areas for improvement.

In conclusion, this study emphasizes the value of data-driven decision-making in the context of fantasy football and team management. By enhancing player recommendations and offering a refined understanding of player worth, this research serves as a valuable resource for navigating the complexities of fantasy football. Furthermore, it underscores the broader potential of data analytics in sports management and fan engagement. As fantasy football continues to evolve, data-driven insights will play an increasingly pivotal role in maximizing success in the virtual gridiron. Through ongoing exploration and refinement, the future promises even more precise player recommendations, benefiting teams and enthusiasts alike.

V. Acknowledgements

The author(s) would like to acknowledge the use of ChatGPT, a language model developed by OpenAI, as a tool for generating text and providing assistance during the writing process.

References

[1] minhnguyen14797, M. N. (2022, May 6). MINHNGUYEN14797/NFL-2021-data-extraction https://github.com/minhnguyen14797/nfl-2021-data-extraction/tree/main

[2] Sessler, M. (2023, February 28). NFL QB Index: Ranking all 68 starting quarterbacks from the 2022 NFL season. NFL.com. https://www.nfl.com/news/nfl-qb-index-ranking-all-68-starting-quarterbacks-from-the-2022-nfl-season