Problem Statement

The goal of this project is to develop multiple machine learning models that accurately predict fantasy football points for the 2023 season. The models will use player performance statistics from the previous seasons (2019–2022), sourced from Football Reference. The key predictive features include metrics commonly used in fantasy football scoring systems, such as yards from scrimmage, touchdowns, receptions, rushing attempts, passing yards, and other relevant player statistics. The objective is to create a reliable tool that helps fantasy football enthusiasts make more informed drafts, lineup, and trade decisions throughout the season.

By using machine learning algorithms, this project aims to uncover underlying patterns in player performance that can forecast future outputs. These models will not only enhance decision-making but also provide valuable insights into how different player metrics interact to contribute to a player's fantasy value. Ultimately, the goal is to empower fantasy players to make data-driven choices that maximize their chances of success.

Articulation of Value

Fantasy football has transformed from a casual pastime into a billion-dollar industry, involving millions of participants, from casual players to high-stakes competitors. In this context, the ability to predict player performance with greater accuracy has immense value for players and the broader ecosystem that supports fantasy football, including betting markets, analytics providers, and media platforms. This project, which aims to predict fantasy football values by developing multiple machine learning models using five years of NFL data, presents significant value on both an individual and market level. For fantasy football players, the essence of the game lies in making accurate predictions about the future performance of players. This process involves drafting players, making trades, and setting weekly lineups. Each decision directly affects the outcomes of league standings and, in many cases, monetary winnings. However, predicting player performance is a challenging task, as it requires understanding various statistics, injury reports, game scripts, and other variables that can influence outcomes.

By using machine learning models to predict fantasy football values, participants can benefit from a more objective and data-driven approach to decision-making. Machine learning excels at identifying complex patterns in large datasets, offering predictive insights that may not be immediately obvious to a human player. For example, while a player might focus on a few key statistics like yards from scrimmage and touchdowns, a well-tuned model can consider many more variables, including situational factors (such as matchups against weak defenses), trends over time, and interactions between different types of performance metrics.

The development of accurate predictive models provides participants with a substantial competitive advantage. For instance, during drafts, players can rely on model-generated rankings to select undervalued players, thereby maximizing their potential point output for a given round. Similarly, in weekly matchups, models can assist in identifying which players are likely to perform

well in a particular game, even if that performance is not immediately obvious based on traditional metrics. This leads to better lineup decisions, improved chances of winning, and ultimately, a more enjoyable and successful fantasy football experience.

The fantasy football ecosystem is supported by a wide array of content providers, ranging from independent analysts to major sports media companies. These entities produce rankings, articles, podcasts, and videos aimed at helping fantasy players make informed decisions. However, many of these insights are based on subjective opinions, historical performance, or simplistic models. A machine learning approach to predicting fantasy football values offers content providers a robust, data-driven method for player analysis. By integrating machine learning predictions into their offerings, they can provide more accurate and actionable insights to their audiences. For example, a predictive model can highlight breakout candidates based on trends in yards from scrimmage or offer dynamic rankings that adjust as new data from the ongoing season comes in. This enhances the quality and credibility of the content being offered to users.

Additionally, analytics providers can use machine learning models to offer advanced tools and dashboards to their audiences. These tools can include predictive models that allow users to simulate future outcomes, evaluate trade proposals, or optimize their starting lineups based on projected points. This provides significant value to players, who can make more informed decisions while saving time and effort compared to manually crunching numbers or relying on less sophisticated analysis.

The fantasy football market has grown exponentially, with millions of participants and billions of dollars at stake. According to the Fantasy Sports & Gaming Association (FSGA), approximately 60 million people participate in fantasy sports in the U.S. alone, with football being the dominant sport. Fantasy football is not only a recreational activity but also a substantial driver of revenue through league fees, subscriptions to expert services, advertisements, and sports betting. By introducing accurate predictive models, this project can help increase player engagement, particularly for more casual players. Fantasy football can be intimidating for newcomers due to the vast amount of data and analysis involved. Machine learning models, which provide easy-to-understand and actionable predictions, can make the game more accessible by reducing the amount of time and effort needed to make decisions. As a result, the project has the potential to expand the market by attracting more players and keeping current players more engaged throughout the season.

Furthermore, predictive models have implications for the broader fantasy sports and sports betting industries. Sports betting has become increasingly intertwined with fantasy football, as

many platforms offer daily fantasy sports (DFS) contests where participants bet on player performances each week. Machine learning models that accurately predict player points can be leveraged by these platforms to offer better odds, set more accurate betting lines, and even create new types of contests based on predictive data. For example, participants could place bets on whether a player will exceed a model's predicted output, or platforms could offer "over/under" contests based on predicted fantasy scores. These models could also be used in the backend of fantasy platforms themselves to improve user experience. For instance, predictive algorithms can be integrated into fantasy platforms to offer lineup optimization features, waiver wire suggestions, or trade evaluations. This type of functionality could become a premium feature for users, providing additional monetization opportunities for the platforms while delivering greater value to the players.

To further articulate the value of predictive models in fantasy football, consider a rough calculation of the potential economic impact. Fantasy football is a $7.22 billion industry, with millions of players contributing through league fees, subscriptions, and DFS contests. If a machine learning model improves a player's decision-making ability by just 1%, this could have a sizable economic impact. If 20% of fantasy football participants win some form of prize in their leagues, and the average entry fee is $100, with 10% allocated to the prize pool, each participant contributes roughly $10 to the prize pool. If a model increases the likelihood of winning by 1%, it could add an average of $0.10 per participant. Multiplied by 12 million prize-winning participants, this represents an additional $1.2 million in economic value derived directly from improved decision-making due to the model. This calculation only considers the direct financial impact on players. The indirect benefits, such as increased subscriptions to fantasy content providers, enhanced user engagement, and more strategic betting in DFS contests, could further increase the overall economic value of predictive models in the fantasy football space.

The articulation of value for this project lies in its potential to enhance decision-making for fantasy football participants, improve the quality of content offered by analytics providers, and generate additional revenue streams in the broader fantasy football and sports betting markets. By leveraging machine learning to predict fantasy football values, this project can drive better outcomes for individual players, content creators, and the industry. The integration of advanced predictive models into the fantasy football ecosystem represents a significant opportunity to create a competitive advantage, improve user engagement, and unlock new sources of economic value.

13 Week Plan

Week 1: Problem definition and scope finalization.

Week 2: Data collection from Football Reference

Week 3: Data cleaning and preprocessing (handling missing data, normalizing stats).

Week 4: Exploratory data analysis (EDA) to understand key features.

Week 5: Feature engineering (creating new features like yards per carry, target share).

Week 6: Train-test split of the dataset and model selection.

Week 7: Baseline model development (linear regression, decision trees).

Week 8: Model refinement (hyperparameter tuning, feature importance analysis).

Week 9: Evaluation of model performance using cross-validation.

Week 10: Model optimization and final tuning.

Week 11: Deployment preparation (exporting model, creating API).

Week 12: Report writing and final documentation.

Week 13: Project presentation and submission on GitHub.

Data

For this project, I will use a variety of datasets sourced from Football Reference, a comprehensive and reputable resource for historical NFL data. These datasets cover player statistics from the 2019 to 2022 NFL seasons. By compiling and analyzing this data, I aim to create machine learning models capable of predicting fantasy football points for the 2023 season. Football Reference provides detailed player statistics for each NFL season, broken down by individual games, cumulative season totals, and advanced metrics. I will be extracting data relevant to fantasy football scoring, focusing on key statistics such as the following: yards from scrimmage, touchdowns, receptions (for PPR leagues), passing yards and passing touchdowns, rushing attempts, targets, fumbles, and interceptions, these datasets will be combined into spreadsheets that model the inputs used to calculate fantasy football points in standard league formats. Football Reference ensures data integrity and historical accuracy, making it an ideal source for modeling purposes.

The compiled data will consist of several key tables, each focusing on different player statistics that contribute to fantasy football scores. Each row will represent a player's individual season or game, while the columns will represent different features such as rushing yards, receiving yards, touchdowns, and receptions. The main features included in the dataset are player name, position, team, games played, yards from scrimmage, receptions and touchdowns, passing stats, rushing, and receiving stats. This data will serve as the foundation for building predictive models, as it encapsulates the primary drivers of fantasy football scoring.

The problem is to accurately predict the fantasy football points of players for the 2023 NFL season. Fantasy football points are typically calculated using various statistics, such as yards gained, touchdowns, and receptions. By focusing on these statistics over the last five seasons (2019-2022), the dataset will allow the machine learning models to learn patterns and relationships between these key performance indicators and their effect on fantasy football points.

This approach solves the problem statement in several ways. Player statistics from previous seasons provide a strong basis for predicting future performance. By analyzing how players have performed in similar situations (in terms of teams, game plans, and opponents), the model can infer which players are likely to succeed in the 2023 season. One of the most significant metrics in fantasy football is yards from scrimmage, which combines both rushing and receiving yards. Players with higher yardage totals often contribute more fantasy points. By including this metric, along with touchdowns and receptions, the model can predict fantasy scores based on real-world performance data. The dataset will account for the different types of statistics relevant to each position. For example, quarterbacks rely heavily on passing yards and touchdowns, while wide receivers and running backs depend on receiving and rushing yards, respectively. Creating a dataset that separates and organizes these position-specific metrics allows for more tailored and accurate predictions for each position. The dataset includes multiple years of data, allowing the models to account for variability in player performance due to factors such as injuries, trades, or changes in team dynamics. This historical data provides a comprehensive view of how players have evolved over time, helping to predict their potential future outputs. With this dataset, the machine learning models will not only predict total fantasy points for the season but can also offer insights into weekly projections. This will assist fantasy football participants in making more informed decisions about drafting, lineup changes, and trades throughout the season.

To improve model performance, the dataset will undergo preprocessing, including data cleaning, feature engineering, and normalization. Cleaning the data will include removing any outliers, missing values, or irrelevant columns. I will be removing the 2020 season as the NFL did not play a full season of games, which will lead to outliers. Creating new features, such as player consistency scores, red-zone usage, and opponent strength, to enhance the model's ability to predict fantasy points. Normalizing the data ensures features like yards and touchdowns are on similar scales. These preprocessing steps will ensure the dataset is optimized for machine learning models and enhance prediction accuracy. The data from Football Reference provides the raw player performance statistics needed to address the problem statement. By organizing this data into spreadsheets and focusing on key metrics like yards from scrimmage, touchdowns, and receptions, the model can deliver accurate and actionable predictions for fantasy football players in 2023.

Modeling

For this fantasy football prediction project, the most suitable modeling approach is supervised learning, specifically a regression problem. The goal is to predict the number of fantasy football points a player will accumulate based on historical data, such as yards from scrimmage, touchdowns, receptions, passing yards, and other relevant statistics. Since fantasy points are a

continuous numeric value rather than a discrete category, the problem lends itself to regression rather than classification.

Supervised learning is appropriate because the model is being trained on historical data where both the input features (player statistics) and the target variable (fantasy points) are known. In supervised learning, the algorithm learns from these labeled examples to make predictions on new, unseen data. The goal is to build a model that can generalize from past performance to predict future fantasy football points. The supervised learning approach fits well with this task for a couple of reasons. The problem is inherently predictive, as the goal is to forecast future performance based on prior knowledge. Football Reference provides historical data with both input features and corresponding fantasy point totals, which can be used to train the model.

Why Regression?

Since fantasy football points are numerical values, this problem is best approached using regression algorithms, which are designed to predict continuous outcomes. In contrast to classification, where the goal is to assign instances to a particular category, regression aims to estimate a value. Fantasy football points can range from single digits to much larger totals, depending on player performance, so a model needs to predict these values as accurately as possible.

Some potential regression models that could be used in this project include Linear Regression, Random Forest, and Gradient Boosting Machines. Linear regression is a simple but powerful method for modeling relationships between input variables and a continuous target variable. It works well when there is a linear relationship between the input features (e.g., yards from scrimmage, receptions) and the output (fantasy points). Random forest is an ensemble learning method that uses multiple decision trees to improve accuracy. It's useful for handling complex relationships and interactions between features, such as how a player's total yards and touchdowns jointly contribute to their fantasy score. GBMs are a more advanced ensemble technique that builds models sequentially, improving upon previous models' errors. It can capture more nuanced relationships in the data and is often used in predictive tasks like this one.

While classification is a common type of supervised learning, it's not appropriate for this project because we are not categorizing players into discrete classes (e.g., "good" or "bad" players). Instead, we're predicting a specific numerical value (fantasy points), making regression the correct modeling approach.

In summary, the fantasy football prediction problem calls for a supervised regression model. Supervised learning is ideal because the model is trained on labeled data where both input features and outcomes are known, and regression is the right choice because the target variable (fantasy points) is continuous. By applying regression algorithms such as linear regression, random forests, or gradient boosting, the model will be able to provide accurate predictions that can help fantasy football participants make more informed decisions in their drafts and weekly lineups.

GitHub Link: https://github.com/PQuinn00/ADAN-8888-Project