1.0 Problem Statement

Fantasy football is a game where participants create and manage a team using real NFL players, competing weekly against others in a league. Players score Fantasy Points (FPS) based on real-life performance, with the highest scorer winning the matchup. Participants set their starting lineup weekly by evaluating projected points and past performances, but projections don't always predict actual outcomes. Factors like location, weather, or team advantage can affect a player's performance, yet there's no easy way to account for them quantitatively. This **project proposes developing a Machine Learning (ML) model to improve FPS projections and support better player selection.**

2.0 Proposed Method

During preliminary research, three published research papers[1][2][3] were reviewed. These papers focused on using machine learning methods to predict a complete fantasy football lineup and algorithms focused on predicting the FPS of a single position (QB). While these studies did not produce a high degree of accuracy, with one paper reporting accuracy in the 31st percentile[3], there are opportunities to continue the research and improve the algorithms used.

This includes:

- Each position has different factors that can affect performance. Therefore, this project will build a **model focused on a single position (QB)**.
- One research paper suggested reducing features to prevent overfitting[3], while another recommended adding more hidden layers to a neural network for accuracy[1]. For this project, **finding the correct number of features that best correlate with player performance will be critical.**
- Each paper used different ML algorithms, including linear regression[2], support vector regression (SVR)[1], and neural networks[1] [3]. Based on the findings, **this project will focus on building a neural network, with plans to explore multiple hidden layers to boost accuracy,** as suggested by one paper (REF).

3.0 Stakeholders

The following stakeholders can benefit from this project proposal:

- **Fantasy football participants** can leverage this to make better-informed decisions on which player to start.
- **ML researchers** have written several research papers on this topic. As this project aims to extend previous research, future research may also benefit.

4.0 Project Risks / Obstacles

The following project risks have been identified:

- Training a neural network requires a high degree of computational resources, which can be expensive. I have a 16-core laptop, but there is a risk that more computational power is needed to train a model in a reasonable amount of time. To mitigate this risk, I also have a desktop with an NVIDIA GPU that could be used to improve training time.
- Most papers so far have yet to publish promising results. However, the overall goal of this project is to improve the algorithm rather than look for perfect predictions.

5.0 Novelty

While the whole project is not considered novel, focusing on adding multiple hidden layers to the neural network and a stronger emphasis on identifying the correct features may lead to a novel solution. I would also like to include weather forecasts as a feature, which I have yet to see in a previous research paper.

6.0 Project Plan

The following sections describe the high-level project plan. See Appendix A for a detailed work back schedule

6.1 Data Sets

Existing data sets exist for NFL statistics; however, many of them are behind paywalls or from sources, and I am unclear on where they collected their data. Based on this, I plan to build my own data sets using ESPN's API to collect statistics. I have performed a proof of concept on extracting this data using ESPN REST APIs. During this proof of concept, I created a platform that will make it easier to extend and collect data from various ESPN REST API endpoints asynchronously. Finally, weather data will be collected from weather.gov's API. Please see https://github.com/amcheste/fball_data_collection for details on this data collection platform.

6.2 Data Organization

Once the data has been collected, preprocessing will be performed. As part of this, I plan to organize my data based on the features I will use as input into the model. This will include removing outliers (e.g., QB who has not started in x games). Data must also be normalized to ensure they are all in the same format. Since this data comes from a single source, ESPN, this is less of a concern; however, if a field needs a value, I must determine how to address it. I plan on doing some independent research and learning the optimal way to handle missing data from the class.

Preliminary features:

- Previous game FPS
- Previous game stats that factor into FPS collection
- Team offensive rank
- The opponent team's defensive rank
- Home or away
- Over/Under
- Weather conditions (e.g., rain, wind strength, etc.)

6.2 Data Analysis

One of my goals with this project is to help refine the features used to improve accuracy. To do that, I expect to analyze the data to understand which features impact or correlate with the projected FPS. I plan on using data visualizations to identify if there is a correlation or trend with various features.

6.2 Model Type

As described in the proposed method section, I plan on building a neural network with a second hidden layer to extend previous research projects. One area I will still need to research with the model is the type of activation function. Preliminary research shows sigmoid as a popular activation function; however, I am unclear if this will be optimal for this use case. To reduce the effort required to implement a neural network, I plan on leveraging open-source technologies such as PyTorch or TensorFlow.

6.2 Model Validation

The collected data set will be split into training and validation data to measure the model's performance. I plan on using a 90%/10% split of training and validation data, respectively. Finally, since we are in an active football season, I will test the model with this active season to get real-life validation.

Appendix A: Project Schedule

Name	Description	ETA
Project Proposal	Create a project proposal document	10/3
Finalize data collection	Finalize data collection project and collect required data in a .csv format.	10/10
Data Preprocessing / Feature selection	Load the data in pandas, remove outliers, normalize data where required and identify features.	10/17
Create model	Create code that represents the neural network model. This will include selection of the activation function.	10/24
Train model	Train the model with the complete data set.	10/31
Buffer	Scheduling buffer to help mitigate risks.	11/7
Final report	Create final report document.	11/14
Final Presentation	Demo prep.	11/21

Appendix B: References

[1] - Lutz, R. (2015). Fantasy Football Prediction. *ArXiv*. /abs/1505.06918

[2] - Dolan, P., Karaouni, H., & Powell, A. (2015). Machine learning for daily fantasy football quarterback selection. *CS229 final report*.

[3] - Mahoney, J. M., & Paniak, T. B. (2023). Method and Validation for Optimal Lineup Creation for Daily Fantasy Football Using Machine Learning and Linear Programming. *arXiv preprint arXiv:2309.15253*.