Fantasy Football Draft Assistant

The General Managers - Group 7

**Data Science Capstone Project
Launch Report**

Date:

April/17/2025

Team Members:

Name: Caleb Miller - cm3962@drexel.edu

Name: Hashim Afzal - ha695@drexel.edu

Name: Thomas Kiefer - tmk326@drexel.edu

Name: David Blankenship - dwb65@drexel.edu

The System/Product

System/Product Name:

Fantasy Football Draft Assistant

Introduction:

Fantasy football is a multi-billion-dollar industry with nearly 30 million participants in the United States alone. As competition intensifies and the demand for analytical insights grows, amateur fantasy football managers are increasingly seeking data-driven tools to gain a competitive edge. Our project, the Fantasy Football Draft Assistant, aims to meet this need by providing real-time, personalized draft recommendations using advanced machine learning models and historical NFL data.

In DSCI 591, we built a robust dataset by scraping and merging several years of NFL player performance data from "golden standard" professional football data sources like Pro Football Reference and NFL.com. We engineered meaningful features such as average draft position and player fantasy point outputs for both Point Per Reception (PPR) and Standard scoring. We then investigated multicollinearity and the distribution of our features to better understand our data.

Now, in DSCI 592, we are focused on developing and evaluating fantasy point prediction models (e.g., XGBoost, Neural Network, etc.), constructing a recommendation engine based on these predictions, and delivering our results through an interactive dashboard UI. Our recommender system will also incorporate user-defined draft strategies and positional needs, guiding managers to make smarter draft choices in real-time. Ultimately, the main deliverables for this phase include the finalized fantasy point prediction models (Standard and PPR formats), the customizable draft recommender system, and the user-friendly dashboard for real-time player recommendations.

Highlighted Features:

- Predictive Fantasy Point Modeling
    - Develop Linear, XGBoost, Deep Neural Network, and MLP Regression models to predict player fantasy performance for the upcoming season.
- Player Recommender System
    - Factors in predicted player performance and live draft player availability and positional needs.
- Custom Draft Logic
    - Examples include ignoring Quarterbacks in early rounds or prioritizing "flex" position players.
- Real-Time Draft Assistant Dashboard
    - Interactive UI with support for multiple fantasy scoring formats.

Sponsor or Proxy User:

We do not have a specific external sponsor for this project. Instead, our target users are the millions of amateur fantasy football managers across the U.S. Our team members, who are all passionate football fans and experienced fantasy football players, act as proxy users for design and evaluation purposes. We aim to build a system that reflects the real-world needs and preferences of our target users. Through iterative testing and domain expertise, we will ensure the tool delivers relevant and actionable draft insights.

Issues:

One issue that we will be tackling early is the need to switch from a wide format to a long format for our data. The long format will include player performance for a single season with the target being the next year's fantasy points scored. We had initially grouped several years of data for individual players in one row, but after a discussion on the true purpose of the models we have elected to rework the data into a long format. The goal of this rework is to allow us to predict the top performers for the next year based on past performance profiles. Additionally, in order to increase the size of our dataset we decided to expand the years that are included. While this will involve reworking our data acquisition, we expect no major issues. This is due to our decision to develop reusable tools in DSCI 591 for data collection/web scraping which we can repurpose to our new format.

This new format will allow us to develop a train/test/predict regime for our model. This involves a train-test split in which we will use 2018-2022 data as our training data and 2023 data as our test data. We will also incorporate what we're calling a predict split where we will use 2024 performance metrics to predict the 2025 Fantasy Football performance. This, along with a rules-based recommendation system relating to Fantasy Football gameplay, will allow us to recommend the best players.

In regards to expertise our team has experience with Football and Fantasy Football and are familiar with the relevant statistics.

The Team

Team Name:

The General Managers

Team Members and their specialties:

We are continuing on to this quarter with the same team members from the DSCI 591. Our structure is a flat hierarchy with tasking defined and updated continuously as we move into new phases of the project. Our weekly meetings will serve to orient and direct our efforts. Below are our team members and our various backgrounds and strengths.

- Caleb Miller - Caleb is currently working as a Data Engineer for NeuroFlow in Philadelphia, PA. In this role Caleb displays his Python, SQL, and Snowflake skills to build and maintain NeuroFlow's database. He also uses his machine learning skills to provide actionable insights in the mental health realm. In the past, he has taken a course on creating interactive dashboards (using Python's Dash library), and he has incorporated this into his project portfolio. Additionally, Caleb has taken on various projects with web-scraping based data acquisition.
- Hashim Afzal - Hashim is currently a Machine Learning and Environmental Management Researcher at the Los Alamos National Laboratory in Los Alamos, NM. In this role Hashim leverages his knowledge of machine learning and programming languages to model potential environmental risks. Apart from his research work, he is currently taking a course on Recommender Systems here at Drexel and plans to pass on his knowledge to his fellow team members.
- Thomas Kiefer - Thomas currently works as an Informatics Data Analyst Intern for IBX, where he utilizes his experience with preprocessing, exploratory data analysis, and feature engineering via Python and SQL skills to help handle model development and operationalization. Similarly, this position, as well as his successful completion of several applied machine learning courses at Drexel, has given him the opportunity to sharpen his knowledge on and comfortability working with machine learning algorithms, such as XGBoost and Neural Networks. In addition to his technical background, Thomas's academic and professional experience in exercise physiology allows him to provide a unique perspective that combines data science and the real-world application of player performance metrics.
- David Blankenship - David has a solid foundation in the Python programming language having focused primarily on data science tools and libraries. From his time in the Navy, he has strong experience with project management in highly technical fields. As a faculty research assistant at a UMD-College Park's Astronomy department he developed technical documents for comet archival and is familiar with scientific writing having done so during his co-op working with Drexels Integrated Circuits and Electronics (ICE) Design and Analysis Lab. While working with ICE he developed Neural Network-based models to predict circuit design metrics. His interest in this project is primarily in developing his skills in building a recommendation engine. In addition to the above skills he will be bringing his experience in preprocessing, EDA, feature engineering, and model pipeline development to the team.

Team Communication:

Two of the team members are on-campus students and two are remote students. We have worked together in the previous quarter and will be continuing the collaborative format that we found success with previously. First, we will primarily be communicating remotely with in-person reserved as necessary for those on-campus. We have regularly scheduled Zoom meetings, a text message group, and each other's email addresses. Our planned meeting schedule is as follows:

- Regular Meetings at 7:30 PM on Monday with the full team for discussion, tasking, and updating team members on results of previous weeks work.
- Ad hoc meetings with full team or individual members as needed dependent on our tasking.

We are using two primary sources for document sharing: Github and Google Drive. Our Github, linked here https://github.com/caleb10miller/FantasyFootballHelper, will be reserved for our public, user-facing code base related to the end product. Google Drive will be for private, team-facing documents and uploads to ensure we can share documents as we continue to develop our project. This will ensure our final product on the Github page is a well maintained repository. Both Github and Google Drive were created in the previous quarter and will continue to be used.

Team Issues:

The largest potential issue we've identified for this project is the dashboard development. It is the area with which we have the least experience. However we have been developing our knowledge over the last quarter and will continue to do so as we approach the dashboard development later in this quarter. This issue is further mitigated by the fact that at least one team member, Caleb, has familiarity with dashboard development and will be able to assist and lead the other team members. Further we have allocated a sizable portion of time for our project to this line of effort to allow for time to learn as we develop the dashboard.

Currently, we expect no issues with scheduling as all members have agreed on our meeting times and we have already begun our regular meetings. If an issue arises the team member will let the rest of the team know and we will flex and find a new time we can all use as needs require.

The Project Plan

Project Statement: Our goal for this project is to develop a recommender system tailored towards Fantasy Football drafts and create a dashboard for users to develop their Fantasy team.

Our key deliverable is a recommender system capable of providing Fantasy Football managers data-based draft suggestions that gives them a competitive edge leading to improved fantasy team performance for the user. We also would like to provide a dashboard for the user to easily interact with our recommender system without a barrier of technical knowledge.

Our weekly milestones for this quarter are listed in the Weekly Milestone Table in the appendix. We highlight the major milestones for this quarter below that will occur over the course of the Spring 2025 quarter.

- Week 2 - Launch Report
- Week 3 - Outline & Steps Presentation
- Week 4 - Pitch Presentation
- Week 7 - Machine Learning Presentation
- Week 9 - Predictive Modeling Report
- Week 10 - Final Project Presentation

We are currently working on developing models to predict player performance. We will be looking at linear regression, XGBoost, a multi-layer perceptron, and a deep neural network. We may also expand the models that we look at as necessary to continue boosting the performance of the model. We will also test different preprocessing techniques and conduct hyperparameter tuning to optimize our model. Our goal is to have this finished by the end of week 5. We will also be developing rules that allow us to intelligently recommend players that consider Fantasy rules; for example. limits on the number of players in a specific position. Once we are satisfied with our model and rule development we will begin to develop a dashboard that will allow users to interact with our model. Our goal is to develop a dashboard such that a user will be able to build out a fantasy team using the output of our models in an intuitive way. We are tentatively looking at developing the dashboard using Plotly and Dash. Our goal is to finish this by the end of the quarter; well before the beginning of the 2025 NFL season in September.

Weekly Milestones
This table will be updated weekly with lines of effort and goals to reflect our team's progress.

Weekly Milestones Table			
Week	Deliverables	Lines of Effort	Goals
1. 31MAR - 06APR		Rework Data Format Launch Report	
2. 07APR - 13APR		Format Rework Model Development	Finish data format rework Finish Launch Report
3. 14APR - 20APR	Launch Report Outline & Steps Presentation	Tune Models Explore additional features Build and Record Presentation	Submit Launch Report Submit Presentation
4. 21APR - 27APR		Initial Modeling Results Explore more complex models	Have Features Finalized
5. 28APR - 04MAY	Pitch Presentation	Evaluate Model Begin Dashboard Development Build and Record Presentation	Finished Models Submit Presentation
6. 05MAY - 11MAY		Dashboard Development Build and Record Presentation	Working Prototype of Dashboard
7. 12MAY - 18MAY	Machine Learning Presentation	Dashboard Development Build and Record Presentation	Submit Presentation Rough Draft of Dashboard

Weekly Milestones Table

8. 19MAY - 25MAY		Dashboard Development Predictive Modeling Report	Finish dashboard development
9. 26MAY - 01JUN	Predictive Modeling Report	Predictive Modeling Report Develop Final Presentation	Submit Predictive Modeling Report
10. 02JUN - 08JUN	Final Project Presentation	Record Final Presentation	Submit Final Presentation

Table of Contributions

The table below identifies contributors to various sections of this document.

	Section	Writing	Editing
1	Project	Caleb, David, Hashim, Thomas	Caleb, David, Hashim, Thomas
2	Team	Caleb, David, Hashim, Thomas	Caleb, David, Hashim, Thomas
3	Plan	Caleb, David, Hashim, Thomas	Caleb, David, Hashim, Thomas

Grading

The grade is given on the basis of quality, clarity, presentation, completeness, and writing of each section in the report. This is the grade of the group. Individual grades will be assigned at the end of the term when peer reviews are collected.