Fantasy Football Draft Assistant

The General Managers - Group 7

Data Science Capstone Project
Launch Report

Date:

01/13/2025

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 massive, lucrative game that has exploded in the United States over the last quarter-century. As of 2022, nearly 30 million people in the United States play fantasy football annually, resulting in an industry that consistently generates over $10 billion per National Football League season[1]. While impressive on its own, fantasy football makes up just a small portion of the growing machine of sports gambling in the United States. Sports gambling as we know it is legal in 39 U.S. states, and Americans spent over $120 billion on sports gambling in 2024 alone[2]. This combination of increased popularity and lucrativeness has led to a fervent community of amateur fantasy football managers who are always looking for new ways to gain a competitive edge over their weekly opponents. The Fantasy Football Draft Assistant recommender system is an attempt to provide these managers with a way to improve their decision making in real time when drafting their teams . By harnessing advanced machine learning techniques and historical NFL and fantasy football data, we plan to deliver a recommender system UI that managers can utilize to assess which available players will increase the value and performance of their teams the most.

Highlighted Features:

- Advanced player recommendation system incorporated into an easy to use UI.
    - Utilizes historical NFL and fantasy data, including, but not limited to, offensive and defensive performance statistics as well as average draft positioning for different players.
    - Plan to use ML based regression to predict fantasy point scored. Tentatively plan to use MLP Regression (neural network), XGBoost Regression, and Linear Regression
- Real-time player availability and team composition tracker to be updated by the user.
- Customizable filtering for position-based recommendations and “sleeper” picks.

Sponsor or Proxy User:

We do not have a specific sponsor externally, as the system itself is meant to be used by amateur fantasy football managers. As mentioned in the introduction, we aim to provide insight and a competitive edge for as many of the nearly 30 million fantasy football managers that play in the U.S. as possible.

Issues:

While we have not encountered any issues accessing the historical NFL and fantasy draft data needed for this system to be trained and operated, one challenge that may arise is incorporating all of our data seamlessly, as it is being webscrapped across 10 different sources. In addition, player data on injuries and suspensions is too difficult to reliably scrape from the internet, leading to its exclusion from the system. Furthermore, the impact of new coaches and team schemes on player performance is highly subjective and cannot be accurately quantified, so this factor will not be incorporated into the model. Lastly, even in the ideal scenario that our Fantasy Football Draft Assistant successfully provides great recommendations based on accurate fantasy point predictions, they are still just predictions. The NFL games themselves contain a plethora of unpredictable variables, making it impossible to perfectly capture

exactly what will happen before the game is even played. As far as interpreting the insights produced by the system, we are confident that our team's extensive knowledge of fantasy football and data science methodologies will enable us to effectively analyze the output and translate it into clear, actionable strategies for draft decision-making.

The Team

Team Name:

The General Managers

Team Members and their specialties:

Our planned 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 serves as a Data Science Intern for Drexel's Clinical Exercise Physiology Lab, where he utilizes his experience with preprocessing, exploratory data analysis, and feature engineering to handle initial model pipeline development. 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:

Three of the team members are on-campus students, while one is a remote student. We have several considerations outlined here to ensure we are able to work collaboratively. 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:

- 7:30 PM on Monday with Professor for Project Progress Update
- 7:30 PM on Wednesday 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. Github 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 are already created and being used.

Team Issues:

The two largest potential issues we've identified for this project are two skill gaps. The primary skill gap for our team is with recommender systems. As the implementation of this system will be done next quarter, we will be using this quarter to prepare by educating ourselves on recommender systems. We will utilize a variety of resources to do so. We are also considering implementing a dashboard dependent on time. As with the recommender system we will be looking at developing our skills this quarter in order to prepare for us implementing this next quarter. We include a list of reference material in the appendix below for both recommender systems and dashboards.

Both of these issues are further mitigated by the fact that at least one team member has familiarity with the subject and will assist the other team members with our study of these over this quarter. Hashim is familiar with recommender systems and Caleb is familiar with dashboards.

Currently, we expect no issues with scheduling as all members have agreed on our meeting times and we have already been regularly attending 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.

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.

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. Our stretch goal deliverable, dependent upon time in the next semester, is 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. We will also update these milestones in our next quarter as we continue the project.

- Launch Report
- Acquisition & Preprocessing Report
- Pitch Presentation
- EDA Report
- End of Quarter Status Presentation

The above milestones will occur over the course of this quarter (Winter 2025).

We have already begun the process of web scraping the requisite data immediately after team formation. In week 3 we will submit the launch report and continue the data acquisition & preprocessing work. Over the course of the next 3 weeks we will continue finalizing the dataset. This will include verifying missing data and developing our imputation strategy where alternative sources can't be identified. Towards the end of this 3 week section we will develop our Pitch presentation and Acquisition & Preprocessing Report.

In week 6 we will deliver the Pitch presentation and Acquisition & Preprocessing Report and begin the EDA and Feature Engineering portion of our work over the next 3 week period. This will include univariate analysis, missing value and outlier analysis, feature engineering with the goal of completion of each one week after the next. We will conduct other necessary data analysis to ensure a strong baseline understanding of our data set and return to acquisition and preprocessing as the necessity shown in the analysis drives us. In week 9 we will deliver the EDA report and in week 10 we will deliver our end of quarter Status Presentation. In the next quarter we will begin our model building, evaluation, and model deployment work.

Appendix

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. 06JAN - 12JAN	Team formation - 06JAN	Team formation	Form team and finalize meeting schedule
2. 13JAN - 19JAN		Begin web scraping	Have initial data scraping finished by the end of this week. Identify additional variables to incorporate
3. 20JAN - 26JAN	Launch Report - 20JAN	Continue data scraping and begin preprocessing - Verify missing data - Develop imputation strategy - Identify alternative sources as needed	Scrape additional variables – begin dataset finalization – minor data cleaning Submit Launch Report.
4. 27JAN - 02FEB		Continue preprocessing	
5. 03FEB - 09FEB		Continue preprocessing	Have preprocessing finished. Have the report and pitch finished by 10FEB.
6. 10FEB - 16FEB	Pitch Presentation - 10FEB Acquisition & Preprocessing Report - 10FEB	Begin EDA and feature engineering	Submit Pitch Presentation and A&P Report Try to have univariate analysis complete
7. 17FEB - 23FEB		Continue EDA and feature engineering	Try to have missing value and outlier analysis complete

Weekly Milestones Table			
8. 24FEB - 02MAR		Continue EDA and feature engineering Begin creating Status Presentation	Finish feature engineering Finish the EDA report
9. 03MAR - 09MAR	EDA Report - 03MAR	Continue developing Status Presentation	Submit EDA Report
10. 10MAR - 17MAR	Status Presentation		Submit Status Presentation

Sources Cited

1. Oehy, Alessandro. "Fantasy Football - NFL's Fan Engagement Engine." Medium, The AO, 30 Aug. 2024, medium.com/the-ao/fantasy-football-nfls-fan-engagement-engine-c713699859e6#:~:text=As%20of%202022%2C%2029.2%20million,industry%20in%20the%20U.S.%20alone.

2. Hardy, Kevin. "Growth of Sports Betting May Be Linked to Financial Woes, New Studies Find." Stateline, 5 Dec. 2024, stateline.org/2024/12/05/growth-of-sports-betting-may-be-linked-to-financial-woes-new-studies-find/.

Recommender Systems References

- https://github.com/recommenders-team/recommenders
- https://cseweb.ucsd.edu/~jmcauley/pml/
- Charu C. Aggarwal, Recommender Systems: The Textbook (RST), Springer, 2016. ISBN:9783319296579
- Jannach, D., Zanker, M., Felfernig, A., & Friedrich, G, Recommender Systems: An Introduction (RSI), Cambridge University Press, 2010. ISBN: 0521493366
- Ricci, F., Rokach, L., Shapira, B., & Kantor, P. B. Recommender Systems Handbook. 2011, ISBN: 9780387858203
- Kim Falk, Practical Recommender Systems. 2019, ISBN: 9781617292705

Dashboard References

- https://plotly.com/examples/dashboards/
- Interactive Dashboards and Data Apps with Plotly and Dash by Elias Dabbas.
    - https://github.com/PacktPublishing/Interactive-Dashboards-and-Data-Apps-with-Plotly-and-Dash
- A variety of additional textbooks are available online through Drexel Library

Table of Contributions

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

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