# **AI Model for Fantasy Soccer League Draft and Roster Management**

## **Project Overview**

Data is the lifeblood of the fantasy sports industry. Unfortunately, when everyone is looking at the same data in the same way, there is little competitive advantage to be gained. In this project, we seek to create a AI model that gathers the relevant publicly-available statistics on players in the Major Leage Soccer league to create a program that assists team “owners” with the creation and season-ling management of their teams.

## **Objectives**

1. Identify the relevant data sources, connect to their API, and access the relevant statistical data.
2. Analyze this data and identify 1-3 ways in which we can combine these statistics to enable novel and proprietary analysis in order to provide a competitive advantage
3. Using these novel analytics, alongside traditional measurements, rank available players by position for draft preparation
4. On an ongoing basis, the model will take in new player data, updated with the most recent match’s stats. Owners will be able to submit their roster, and the model will provide an analytics-driven lineup for that week’s competition
5. Provide owners with graph-based data visualizations to support the recommendations

## **Key Questions**

1. How have owners traditionally used player stats to build their teams? What are the statistics that typically drive these decisions?
2. How do owners manage their teams during their season? What are the key analytics that drive lineup decisions, trades, and other in-season roster moves?
3. What stats can help us to understand if a player is “overvalued” or “undervalued?” And how can we leverage “conventional wisdom” in order to provide a novel approach to statistical analysis of players?
4. Can we provide a differentiated product? Is the mousetrap we seek to build actually better?

## **Data Sources**

* **TBD**

## **Hypotheses**

1. There non-obvious correlations between player statistics that could lead to novel predictive insights.
2. We believe that combining multiple machine learning models have the potential to improve prediction accuracy.
3. Because player valuations in fantasy sports are often driven by individual biases (e.g., recency bias, confirmation bias). By understanding and adjusting for these biases, the AI model can identify undervalued or overvalued players, providing a competitive advantage.

## **Project Timeline**

|  |  |  |
| --- | --- | --- |
| Wed | 8/28/2024 | Data source exploration |
|  |  | Finalize data sources |
|  |  | Assign data-related tasks |
|  |  |  |
| Thu | 8/29/2024 | Individual Data Source Exploration |
|  |  | Develop Model Architectural Framework |
|  |  | Begin EDA |
|  |  |  |
| Fri-Tu | 8/30/2024 - 9/3/24 | Continue EDA, Begin Data Cleaning & Preprocessing |
|  |  | Initiate coding |
|  |  |  |
| Wed | 9/4/2024 | Coding and collaboration |
| Thu | 9/5/2024 | Coding and collaboration |
|  |  |  |
| Fri-Sat | 9/6/2024 - 9/7/2024 | Everyone wrap tasks for Sunday collaboration |
|  |  |  |
| Sun | 9/8/2024 | Collaborate and begin fine-tuning |
|  |  |  |
| Monday | 9/9/2024 | Model completed - conduct final testing in class |
|  |  |  |
| Tuesday | 9/10/2024 | Complete the PowerPoint, Zoom to discuss Presentation |
|  |  |  |
| Wed | 9/11/2024 | Project Presentation |

## **Dependencies**

This project requires Python and the following Python libraries installed:

* import pandas
* import sklearn libraries
* import time
* import requests
* import os
* from dotenv import os
* from dotenv import load\_dotenv
* from pathlib import Path
* import matplotlib.pyplot
* import hvplot.pandas
* import numpy
* import json
* from datetime import datetime, timedelta
* import requests

## **Directory Structure (abbreviated), Installation, and Usage**

**TBD**

## **Project Structure**

## **Installation and Usage**

## **Results and Analysis**

## **Contributors**

* Matt Cannon
* David Kaplan
* Dan Quinn
* Roberto Reis

**License**

GNU GENERAL PUBLIC LICENSE, Version 3, 29 June 2007