Lorenzo Dube

ME397M Term Project Proposal

3.29.2023

Best XI of the MLS

For my project I would like to examine the MLS and create a model to predict the Best XI (all star team). For my project I plan to use the dataset at the following link on Kaggle: <https://www.kaggle.com/datasets/josephvm/major-league-soccer-dataset>. This dataset includes player season stats from the 1996 season through the 2020 season and table results for all seasons from 1996-2020. I would also use fbref.com to create .csv files for the data from 2021 and 2022.

I will use SQL tables to hold the data and allow for querying across the data as necessary. I will clean the data using pandas, analyze the data with relevant ML libraries, and visualize through matplotlib or plotly.

For the specific analysis, I will use this data to perform logistic regression to find the likelihood of a player being a Best XI caliber player based on full regular season performance. From there, I will show the qualities most relevant to being named a Best XI caliber player by position using correlation. Additionally, I would hope to find any players who might have been deserving at being named to the Best XI, but were not. As a deliverable, I would like to have an interactive dashboard that would show the true and expected Best XI from each season. Additionally, I will output some graphics showcasing the most important factors from each position.