Capstone Project: Using Machine Learning to Predict Optimal Soccer Positions

Suraj Daru
CSCI 1420

May 15, 2024

1 Abstract

Soccer is the most popular sport in the world, and scouting/team management are huge components at the professional level. Figuring out how to most optimally create a lineup, what type of substitutions to make, and how to determine the optimal position for players are just a few aspects of analysis that are integral to a successful team. In this paper I chose to focus on the later, using neural networks, K-nearest neighbors (KNN), and Principal Component Analysis to determine the most optimal positions based on Premier League dataset from Kaggle. I was able to create a neural network model that accurately classifies outfield positions at $\sim 85.8\%$ within the testing set after using cross-validation-loss to determine appropriate hyperparameters. Interestingly, the linear Model performed just as well as the nonlinear Model, and the most optimal KNN model did ever so slightly better with a testing accuracy of 86.1% – PCA resulted in information lost that reduced the testing accuracies. This provides a good point to improve upon and expand on classifying more specific player positions with an expanded dataset.