Premier League Match Predictor

With this data analysis project, I’ve pulled data from the web on team results from the last premier league season. With this data, I utilized Pandas and NumPy to predict the match results, and by extension the league table for the upcoming premier league season. Through this project, I was able to form solid predictions for the upcoming season, and analyze the effectiveness of my prediction algorithm.

Chart, line chart

Description automatically generatedThis line plot represents my predicted point totals for each team in the upcoming season vs. the point totals for teams from last season (DISCLAIMER: For the promoted teams, I predicted results based on the performances of relegated teams from last season). From the chart, it is clear that my formula favored teams that placed higher in the league last season and went against teams that struggled in the previous season. The teams with high point totals last season are predicted to have even more points with my algorithm, while teams at the bottom are predicted to do even more poorly. To see if my predictions are part of a larger pattern, I pulled the point totals for teams in each of the previous 5 seasons and plotted the results for each season in the plot below.

Chart, line chart

Description automatically generatedAs you can see, there is not a clear pattern of increasing disparity in Premier League point distributions. With my algorithm, the teams at the bottom of the table are predicted to have significantly lower point totals than in previous seasons, while the top teams are slightly closer to past results. I believe this flaw in my algorithm comes from the favor it gives to teams that won by large amounts last season. Within my match algorithms, it gives a team additional goals in any game where they won by multiple goals last season. Since usually the teams at the bottom lose games by the most, and the teams at the top win by a lot, this increased the disparity in results between these sets of teams. As far as goals go, my algorithm seemed to be much more similar to the results of the previous season. The chart shows that some teams are predicted to score more goals, while an equal amount are predicted to score less. The only glaring disparity in this table is that the bottom team is predicted to score significantly less, which could easily be the case in the upcoming season and doesn’t seem to be an unreasonable predicted outcome. On the other end, with predicting goals conceded, only the top 2 teams are predicted to concede far less goals than the previous season. This shows the reason for the point disparities from last season for the top and bottom teams.

Chart, bar chart

Description automatically generatedChart, bar chart

Description automatically generated

Overall, I was successfully able to use data from previous seasons to predict outcomes for the upcoming year. While there are certainly flaws in the way my algorithm handles games between mismatched opponents as it doesn’t provide many upsets, it is certainly a reasonable way to analyze the data. This is the case for most prediction algorithms as big upsets are outliers that occur too rarely to account for in a simple machine learning analysis.