Report on important features for predicting football match outcomes.

Introduction

The aim of the report is to discuss the most important features, within the data which have been provided, to consider when predicting the possible outcome of a football match between two teams. Some features within the data have been considered to have some form of relevance towards achieving the outcome whilst other pieces of data have less relevance and will likely not play a role in determining the characteristics in the final predictor for football matches.

Method

The most important piece of data seemed to be the results data gathered from all the years of each football league. The data was stored inside a csv file accessible via the ‘pandas’ python library. The column of most interest came from the ‘Result’ which states the score of the home-team against the away-team.

It was important to consider how much each team won against their opponent as this may play a major role in the predictor. It was also another important detail to look at the broader scope of the teams, not just 1-on-1 matches, but also how the team performed overall over the entire year. A team which would win more matches over an entire year compared to another would likely have a higher chance of winning future matches and a football match predictor should reflect this greater chance.

Another feature considered was the number of matches played by a team through the entirety of the year. A team with a larger number of matches played would provide a more reliable way to predict their future matches, as future outcomes will be less likely to deviate than if the teams had a smaller population of data that could be worked with. With many more results from larger number of matches played, the potential impact of ‘upset games’ where teams, that are not as good, beat better teams due to factors such as luck or the better team having an off day. With a larger dataset, these anomalous results will have a smaller impact on the final predictor and is unlikely to skew the predictor into favouring the team which is not as good.

When looking at a future match-up against two football teams, another factor which could play a role into a team having a greater advantage is considering their performance in home or away games. It is expected that teams who play at home have an advantage in terms of the home crowd. Therefore, it is worth analysing the data to observe whether teams do have a home-field advantage or not and whether their performance increases when playing in front of their home crowd.