Introduction

The sport of football is undoubtedly one of the major sports around the world. The sport has not only risen in its popularity in terms of spectators or fans who follow the sport, but also in the area of gambling. Betting on football matches has become a major part of the gambling market around the world. The ability to correctly predict results of matches is an enigma which one rarely manages to solve. With regards to football correct predictions rely on multiple factors such as, quality of opponent, whether match is played in home ground or away, teams current form and many more. The number of factors that can be decisive make it much more challenging to correctly predict the result of a match.

Focus of this study lies on the odds provided by betting companies and their predictive qualities. This study aims to discover if the odds provided by several different betting companies can be used as means of prediction together with multiple machine learning techniques. This study asses the odds provided on the leagues that are considered to be the top five leagues in Europe over the course of nineteen seasons. Such leagues include the English Premier League, French Ligue 1, Italian Serie A, German Bundesliga, and the Spanish La Liga. The second goal of this study is to assess to predictive quality over the years, assessing whether or not odds have become more precise over the course of the years involved in this study. The third goal of this study is to uncover in which league have odds been the most reliable to use as means of predictions.

Machine learning techniques today have garnered a lot of attention and are being adapted to our everyday lives. Technologies such as self-driving cars, Facial recognition which is nowadays a common feature in our smartphones are all forms of machine learning. The use of machine learning algorithms to be able predict the matches of a sport which is deemed unpredictable and full of surprises was the main motivation of this study. Such study is relevant as it will reveal whether the odds that are provided to betting enthusiasts, give them a fair chance of correctly predicting (thus winning) matches or if these odds are biased in favour of the company earning more profits. This study is also relevant as it can identify possibilities of using technologies such as machine learning algorithms to determine the fairness of odds and the applicability of such technologies with regards to predictions on football.

To achieve the goals of this study multiple models will have to be produced. The first model will include data of leagues for a particular season, while the second will have the data for a particular league and particular season. To assess the predictive qualities of odds these models will be used together with two different machine learning techniques. First these models will be given to a decision tree which will output the predicted results by classifying them. To further asses the odds the same models will be processed through a Neural network using back propagation. Using back-propagation allows the Neural network to adjust the weights of the inputs to obtain the least margin of error.

The rest of this paper is structed in the following manner; The second section garners information from past works and papers related to efficiency of the betting market, how odds and machine learning techniques can be used to correctly classify results in both domestic and international competitions and the reliability of odds as means of prediction. The third section documents the gathering of the dataset. This section also documents the structure of the models fed to the machine learning techniques as well as the structure of said techniques. The fourth section will describe the results obtained from both machine learning techniques, while in the fifth section these results will be discussed.