## 2.1 Studies on the sport of football.

In a recent study on the popularity of sports more than 40% of the people surveyed replied that they were either interested or very interested in Football (Nielsen Sport, 2018). With such popularity, it comes as no surprise that there is a considerable amount of research on topics related to the numerous areas related to the sport. A popular area amongst researchers is the area of prediction. This includes match results, total amount of goals scored in a match, title winners of both national and international competitions and many more. Studies tend to focus more on predicting the outcome of a single match instead of total goals scored in a match and this could be due to match prediction being less complicated, as it requires less parameters and the calculations are considered simpler (Goddard and Asimakopoulos, 2003). Another popular area amongst researchers is market efficiency which focuses on the possibility of having betting strategies that generate constant profits. Football is believed to be an unpredictable sport and it has provided huge surprises, a clear example of this is Leicester city FC surprising everyone and becoming champions of England in the season of 2015/16 when odds at the start of the season were valued at 5000/1, meaning it was virtually impossible for them to win the English Premier league. Although being highly unpredictable, research has shown that matches have predictable elements, such as comparison, evaluation of team’s performance in the prior season and experts’ forecasts and are not completely determined by chance alone (Goddard and Asimakopoulos, 2003) with odds prevailing as the best base for predictions.

## 2.2 Football betting in the Betting market.

The presence of football betting in the economic market and the how odds presented to the bettor have changed over the years will be examined in the following. In 2015 sports betting held the major part of the online gambling market with 48% (Marek, 2018). Football has increased its popularity amongst bettors. In 1998 the British football betting had a turn over close to 2£ billion whereas in 2008 a single betting company registered a 31.4% increase over the previous year and a turnover of 2.92£ billion alone (Constantinou and Fenton, 2013). Bookmakers have their own models to calculate the probabilities of match results. The odds presented to those who are betting are not entirely fair ((Marek, 2018) and (Štrumbelj, 2014)). This means that when the values of the probabilities presented by the bookmaker are added to each other the answer always exceeds 100%. These additional margins are added to cover the company expenses and make profit (Štrumbelj and Šikonja, 2010).

Although the posted odds do not entirely reflect the true probabilities, they can still be regarded as forecasts. The margins added to the true predictive value has decreased over time and this could be due to the ever-increasing number of betting companies thus an increase in selling competition ((Marek, 2018) and (Štrumbelj, 2014)) and (Sheridan, 2012) notes how the use of online betting has further enhanced the competition. Considering the premier league, margins have decreased notably whereas for the 2005/06 premier league they lied between 7.33% and 11.1%, they have recently fallen to 2.71% and 5.79% (Marek, 2018). Odds are a very valuable tool when it comes to forecasting sports events ((Wunderlich and Memmert, 2018) and (Štrumbelj and Šikonja, 2010)). It is noted that odds on the draw outcome have no noteworthy predictive properties (Hvattum, 2012) and that the result of a match and the amount of goals scored in the previous match are not enough to predict the following match results (Wunderlich and Memmert, 2018). This shows how thanks to the success and ever-growing competition regards to betting, companies are now providing the bettor with odds that are closer to the real probable value of that particular result.

## 2.3 The football betting market, Arbitrage & Biases.

Efficiency means that no bettor or bookmaker can achieve greater winnings than the rest. It is said that the market is highly inefficient at the end of the football season (Constantinou and Fenton, 2013). The strongest evidence of efficiency are the multiple studies in which the models presented in said studies failed to outperform the odds presented by the bookmakers (Constantinou and Fenton, 2013). Market efficiency also ensures that there are few arbitrage opportunities to be exploited by those betting (Constantinou and Fenton, 2013).

## 2.3.1 Arbitrage.

Arbitrage is when a bettor takes advantage of the differences in the odds presented by multiple bookmakers on the same event, such that in any circumstance the margin is in favour of the bettor (Vlastakis, Dotsis and Markellos, 2009). Arbitrage opportunities are highly profitable, with winning returns starting from 12% going up to 200% of the money placed (Vlastakis, Dotsis and Markellos, 2009), albeit there is limited evidence, lower divisions have higher profits when one finds such arbitrage opportunities (Constantinou and Fenton, 2013). Although highly profitable these opportunities are very rare. While they are limited to 1 in 200 matches in overall betting, they are reduced to 1 in every 1000 matches when only online bookmakers are considered. Out of a total of 10,374 matches studied there were only 10 cases of arbitrage opportunities, meaning a probability of 0.96%, further emphasizing the rarity of such opportunities (Vlastakis, Dotsis and Markellos, 2009).

When closely looking at the opportunities of arbitrage found, most of them lied in international competitions (E.g. Champions League, Europa League and the FIFA World Cup), where between the teams involved there may be a big gap in the overall quality (Vlastakis, Dotsis and Markellos, 2009). 80% of the time, in these opportunities, the home team are the favourites for the win (Vlastakis, Dotsis and Markellos, 2009). These opportunities have become more common as competition in the market increased and their return rates have remained high (Constantinou and Fenton, 2013). These paragraphs highlight how although the market is highly efficient, strategies such arbitrage exists that allow a punter to make considerable profits and how these can also indicate a match with teams of uneven qualities.

## 2.3.2 Biases.

The home advantage mentioned before wears off even on betting, producing the bias known as the *home-away bias*. In the majority of cases, betting on the home team produces the highest total return (Constantinou and Fenton, 2013). Betting on the home team even when it is the least probable result, is very profitable in England and Germany with profit rates of 6.51% and 5.12 % respectively, it is also noted that this bias does not perform well in Italy and Spain where profit rates are those of -17.61% and -14.08% (Constantinou and Fenton, 2013). The *favourite-longshot bias* is another well-known bias.The bias is described as the over evaluation of the longshot result and under evaluation of the favourite result by those placing the bet. The *favourite-longshot bias* produces higher returns when bets are placed on the favourites (Andrikogiannopoulou & Papakonstantinou, 2011), (Goddard and Asimakopoulos, 2003) and (Vlastakis, Dotsis and Markellos, 2009), and is present in both the result and in the total number of goals scored (Cain, Law and Peel, 2000).

## 2.4 Odds and How they are used as means of predictions.

One study suggests that the predictive power of odds has improved over time (Wunderlich and Memmert, 2016). Another paper however negates this, highlighting how over the course of 7 seasons (Starting from season 05/06 all the way to 11/12) the accuracy has remained the same (Constantinou and Fenton, 2013). It is noted that odds are more precise in top divisions (Constantinou and Fenton, 2013). When comparing the odds bookmakers present, there will always be some small difference in the pricing. This difference could represent superior knowledge one bookmaker has or indicate a match whose outcome is tough to predict. (Deschamps & Gergaud, 2007) state that these differences can be exploited to earn an uncharacteristic return. Odds prevailed as the best source of information to predict results on. The smaller the odd value presented by the bookmaker on a certain team, the more likely that the outcome will be in favour of that team. With the same reasoning the higher the value of the odd, the less probable that result is (Vlastakis, Dotsis and Markellos, 2009).

## 2.4.1 Reliability of Odds a base of prediction

The reliability of odds as a base for predictions differs between the different leagues and are more accurate in certain leagues. Using a mean Brier score and a ranked probability score, it is noted that the most accurate odds were those on the Scottish Premier league, while the worst were those on Frances’ top division, Ligue 1. A probable cause for these results is the lack of quality that lies between the top two teams of Scotland (Celtic F.C and Rangers F.C) and the rest of the participating teams. By the same scale it can be said that the French league has contestants of equal value, making the result much harder to predict thus causing the odds to be more inaccurate than on other competitions (Štrumbelj and Šikonja, 2010). It is noted that the mean accuracy differs slightly between companies. The highest for the season 2014/15 was found in the England’s third tier, the *English Football League two,* where the most accurate company had an accuracy of 53.3% on all of the matches and the lowest had an accuracy rate of 49.3% (Marek, 2018). Such close results suggest that the model’s bookmakers use to produce the probabilities have minimal differences (Marek, 2018).

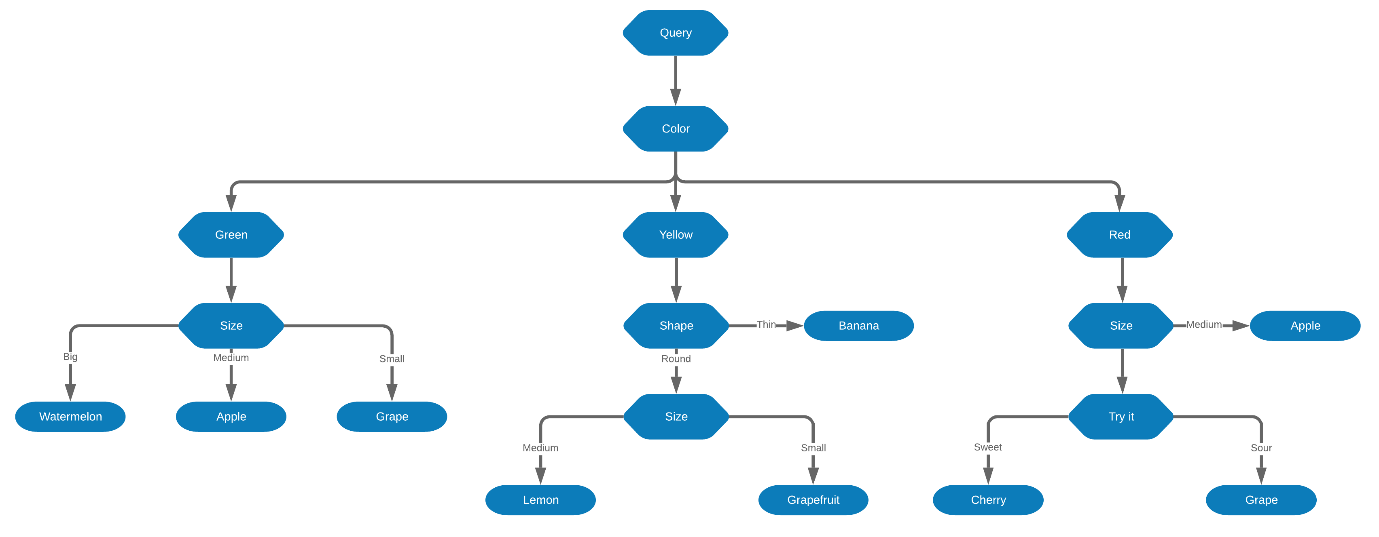
2.5 Machine Learning

Machine learning consists of data-based algorithms that can mimic and understand tasks of information processing. According to (Barber, 2012) these algorithms are often used to enhance human processing, such as fast retrieval of information and prediction within the stock market. While machine learning is associated with Artificial Intelligence, machine learning primarily focuses on driving and adapting the model based on the data given. Decision Trees, Neural Networks and random Forests are all types of machine learning algorithms.

2.5.1 Decision Trees

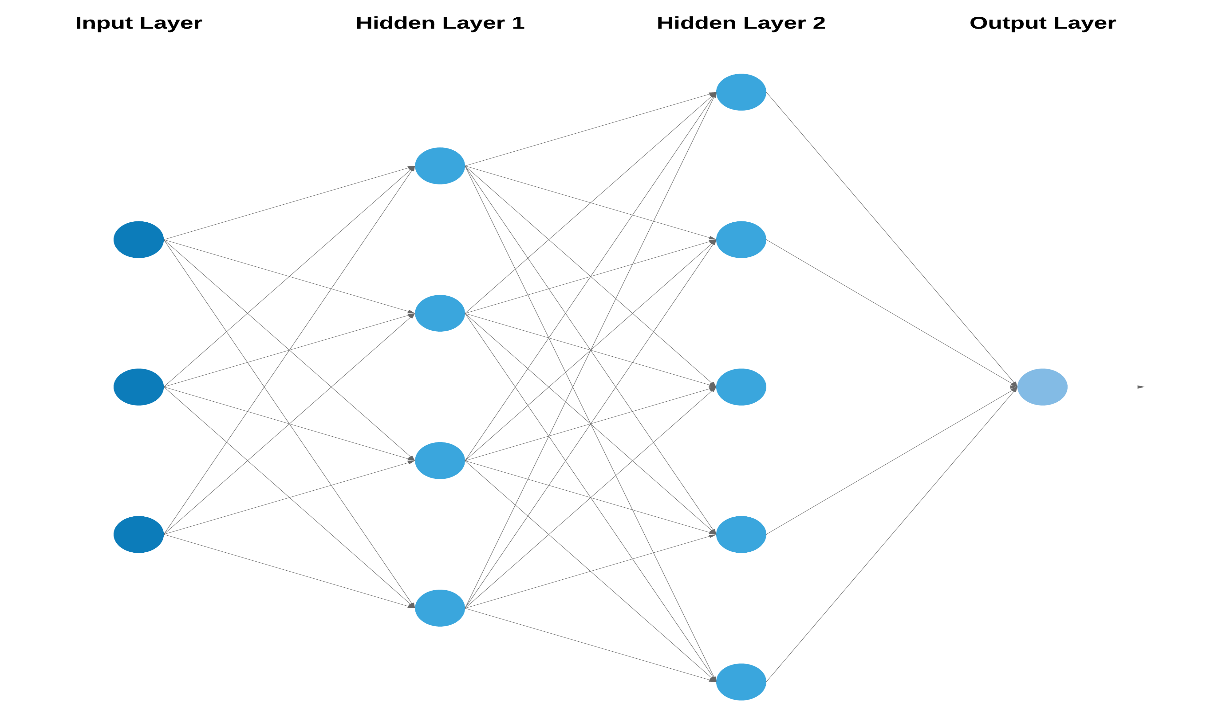
Decision trees is data mining technique that is used to solve classification and prediction problems. These trees are made up of nodes and leaves. Each node represents a feature that requires testing. Leaf nodes give a classification to the instances that reach that leaf, this is done by sorting the instances starting from the root node downwards until a leaf is reached.

Figure 2.5.1.a

**

2.5.2 Neural Networks

A neural network is a set of linked neurons which are organised by layers. The first layer is the input layer where the initial data is inputted. Layers that are present between the input and output layer are called Hidden layers, here through an activation function, the artificial neurons take weighted inputs and produce an output. In the output layer the outputs produced

**Figure 2.5.2.a

## 2.6 Factors to produce a good prediction model

There are numerous factors that need to be considered and that can aid in producing more accurate predictions on the result of a match. The significance of the match, a teams’ cup competitions status and the geographical distance between the teams participating, all have a significant influence on the result ((Goddard and Asimakopoulos, 2004) and (Goddard and Asimakopoulos, 2003)). If the match is of high importance to one team (E.g. A match that determines whether a team wins the title or is relegated to an inferior division.) it is likely to reflect on the result, in particular cases even results that persist season after season can be of significance to the end result (Goddard and Asimakopoulos, 2003). Studies have shown that the home advantage has a significant influence on the results, and that the advantage may increase in cases where the teams are from opposite ends of a country (Goddard and Asimakopoulos, 2003). It is also noted that the greater the competitiveness amongst the teams in a local derby, the home advantage may lose some of its effect (Goddard and Asimakopoulos, 2003).

Other significant factors include the size of the club, where the greater club between the two participants of a match usually wins (Goddard and Asimakopoulos, 2003). The size of the club is not only useful for those predicting, but also for the bookmakers issuing the odds as it taken into consideration when they are pricing the bet (Štrumbelj and Šikonja, 2010). While it is noted that experts (E.g. Former players and punters) or professional bettors are not generally better than naïve individuals at making predictions (Wunderlich and Memmert, 2016), it is advised that expert opinion should be given attention and evaluated as it can have highly valuable information . Contrary to popular belief elimination from cup competitions appears to have a negative effect on the teams’ league result (Goddard and Asimakopoulos, 2003). The factors discussed show how the can affect the accuracy of a prediction, it also noted that former players and punters are not any better than those who are less knowledgeable in football.

## 2.7 Performance of multiple prediction models

Against multiple types of prediction methods and algorithms, odds prevailed as the best source of prediction. An MLP Artificial neural network (a Multilayer Perceptron is a neural network that consists of at least three layers of nodes) with 10 inputs and 1 output was used in a study conducted on the Iranian league and it managed to correctly predict 5 out of 8 matches and in which competition (AFC Champions league or Relegation to the Iranian second division) 5 out of 6 teams where going to participate in (S. Mohammad Arabzad et al. 2014). (Wunderlich and Memmert, 2018) eliminated the margins added to the odds, making sure that the probability added up to 100%. These were then compared to 3 different ELO rating types and all of them were outperformed by odds.

When compared to results produced by statistical models, bookmakers’ odds outperformed these statistical models (Štrumbelj and Šikonja, 2010). This argument is further enhanced by multiple studies. When predicting cup winners’, predictions based on odds produced better results than predictions based on ratings ((Hvattum, 2012) and (Wunderlich and Memmert, 2016)). While make use of Shin’s model, probabilities based on betting odds were also proven to be more accurate than the probabilities that were based on a basic normalisation or on regression models (Štrumbelj, 2014). Although odds have been highlighted for their high relevance in predictions, relevance of time in conjunction with an odd itself has not been discussed. Using machine learning techniques, this study will highlight and discuss the relevance of time when placing bets.

Bibliography

Andrikogiannopoulou, A., & Papakonstantinou, F. (2011). Market efficiency and behavioral biases in the sports betting market. HEC Geneva*,*working paper*.*

Barber, D. (2012). *Bayesian Reasoning and Machine Learning*. 1st ed. Cambridge: University Cambridge Press.

Cain, M., Law, D. and Peel, D. (2000). The Favourite-Longshot Bias and Market Efficiency in UK Football betting. *Scottish Journal of Political Economy*, [online] 47(1), pp.25-36. Available at: <https://onlinelibrary.wiley.com/doi/abs/10.1111/1467-9485.00151>.

Constantinou, A. C. and Fenton, Norman E. (2013). Profiting from arbitrage and odds biases of the European football gambling market*. The Journal of Gambling Business and Economics,* [online]7(2)*: p.41.* Available at: <http://constantinou.info/downloads/papers/evidenceofinefficiency.pdf> [Accessed 22 Sept. 2018]

Deschamps, B. and Gergaud, O. (2007). EFFICIENCY IN BETTING MARKETS: EVIDENCE FROM ENGLISH FOOTBALL. *The Journal of Prediction Markets*, [online] 1, pp.61–73. Available at: <http://ubplj.org/index.php/jpm/article/view/420> [Accessed 20 Oct. 2018].

Goddard, J. and Asimakopoulos, I. (2003) Modelling football match results and the efficiency of fixed-odds betting, Working Paper, Department of Economics, Swansea University.

Goddard, J. and Asimakopoulos, I. (2004). Forecasting football results and the efficiency of fixed-odds betting. *Journal of Forecasting*, [online] 23(1), pp.51-66. Available at: <http://previsaosimples.pbworks.com/w/file/fetch/65223638/soccerForecasting.pdf> [Accessed 12 Sept. 2018]

Hvattum, L. (2012). Analyzing information efficiency in the betting market for association football league winners. Available at: http://www.enexto.com/FootballLeagueSimulator/Hvattum\_2012.pdf [Accessed 10 Oct. 2018]

Marek, P. (2018) *Bookmakers' Efficiency in English Football Leagues*. University of West Bohemia.

Mesarić, J. and Šebalj, D. (2016). Decision trees for predicting the academic success of students. *Croatian Operational Research Review*, [online] 7(2), pp.367-388. Available at: https://hrcak.srce.hr/ojs/index.php/crorr/article/view/4547 [Accessed 15 Oct. 2018].

Nielsen Sport (2018). *World Football Report 2018*. [online] Available at: https://nielsensports.com/wp-content/uploads/2014/12/Nielsen\_World-Football-2018-6.11.18.pdf [Accessed 9 Jan. 2019].

S. Mohammad Arabzad, M.E. Tayebi Araghi, S. Sadi-Nezhad, and Nooshin Ghofrani. (2014). Football Match Results Prediction Using Artificial Neural Networks; The Case of Iran Pro League. Journal of Applied Research on Industrial Engineering 1, 3 (2014), 159-179. <http://www.journal-aprie.com/article_43050.html>

Sharma, H. and Kumar, S. (2016). A Survey on Decision Tree Algorithms of Classification in Data Mining. *International Journal of Science and Research (IJSR)*, [online] 5(4), pp.2094-2097. Available at: https://www.ijsr.net/archive/v5i4/v5i4\_02.php [Accessed 15 Oct. 2018].

Sheridan, D. (2012)*. Modelling football match results and testing the efficiency of the betting market. Master of Literature in Economics.* National University of Ireland.

Štrumbelj, E. (2014). On determining probability forecasts from betting odds. *International Journal of Forecasting*, [online] 30(4), pp.934-943. Available at: <http://lkm.fri.uni-lj.si/uploads/eriks/strumbelj_workingpaper2013.pdf> [Accessed 26 Sep. 2018].

Štrumbelj, E. and Šikonja, M. (2010). Online bookmakers’ odds as forecasts: The case of European soccer leagues. *International Journal of Forecasting*, [online] 26(3), pp.482-488. Available at: <https://www.researchgate.net/publication/46497620_Online_bookmakers'_odds_as_forecasts_The_case_of_European_soccer_leagues> [Accessed 16 Oct. 2018].

Vlastakis, N., Dotsis, G. and Markellos, R. (2009). How efficient is the European football betting market? Evidence from arbitrage and trading strategies. *Journal of Forecasting*, [online] 28(5), pp.426-444. Available at: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/for.1085> [Accessed 7 Oct. 2018].

Wunderlich, F. and Memmert, D. (2016). Analysis of the predictive qualities of betting odds and FIFA World Ranking: evidence from the 2006, 2010 and 2014 Football World Cups. *Journal of Sports Sciences*, [online] 34(24), pp.2176-2184. Available at: <https://www.dshs-koeln.de/fileadmin/redaktion/Institute/Kognitions-_und_Sportspielforschung/Publikationen/Paper/Wunderlich_2016_Analysis-of-the-predictive-qualities-of-betting-odds-and-FIFA-World-Ranking.pdf> [Accessed 14 Oct. 2018].

Wunderlich, F. and Memmert, D. (2018). The Betting Odds Rating System: Using soccer forecasts to forecast soccer. *PLOS ONE*, [online] 13(6), p.e0198668. Available at: https://journals.plos.org/plosone/article/file?id=10.1371/journal.pone.0198668&type=printable [Accessed 11 Oct. 2018].