# Results

## Models

As previously discussed in the previous section of methodology, models were produced by using MSSQL and multiple queries to select the required data. A model for each season available was produced. Models that do not contain the ELO ratings of football teams were constructed as follows; the first two columns had the names of the teams involved in the match. The following column held the result at the end of the match. The rest of the columns after contained the odds given by the betting companies on each of the three ways the result could go (Home, Draw & Away). For models that contained ELO Ratings, the first two columns after FTR, contained the ELO ratings of the home and away team prior to the match. ELO ratings fluctuate over time, they increase or decrease based on the performance of the club. Figure 1 demonstrates the complete structure for Models that do not contain ELO ratings while Figure 2 demonstrates the structure for Models that included ELO ratings.

Figure 1

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| FTR | Home | Away | BetComp1H | BetComp1D | BetComp1A | …. | BetComp10H | BetComp10D | BetComp10A |
| H | Team1 | Team2 | 1.1 | 3.7 | 2.5 |  | 1.15 | 3.62 | 2.35 |
|  |  |  |  |  |  |  |  |  |  |

Figure 2

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| FTR | Home | Away | ELOH | ELOA | BetComp1H | BetComp1D | BetComp1A | …. | BetComp10H | BetComp10D | BetComp10A |
| H | Team1 | Team2 | 1500 | 1300 | 1.1 | 3.7 | 2.5 |  | 1.15 | 3.62 | 2.35 |
|  |  |  |  |  |  |  |  |  |  |  |  |

# 2. Results Obtained from Decision Trees

## 2.1 Decision Trees without ELO ratings

In this section results obtained from the decision trees alone will be discussed. As previously stated in the methodology, a decision tree for all the data in the model was produced as well as a decision tree for each company that was present within the model. From the results obtained by using the models without ELO ratings hinted certain indications. Figure 3 represents the percentage of games correctly classified by the decision trees when using all the data within the given model. It can be clearly seen that throughout the nineteen years of football odds included in this study, no notable progress in the percentage of correctly classified matches can be highlighted. This finding is compliant with the findings presented by *(Constantinou and Fenton, 2013)* were it was highlighted that over the course of the fourteen years included in their study no apparent improvement was noticed. However, the same finding contrasts (*Štrumbelj and Šikonja, 2010)* as in their study they find that improvements where visible over the years when using odds for forecasts. The second notable observation is that all the obtained results except that of the 2004/2005 season all relied between 40 and 50% or slightly over. Such finding is loosely complimentary to *(Constantinou and Fenton,2013)’s* finding where they state that accuracy of odds between bookmakers is extremely consistent. This argument is even further confirmed when looking at the results obtained by the decision trees when given odds of just one company (Figure 4). While performing slightly better than the decision trees with all the data, most of the results obtained still fell within the same range (40-50%). Such observation is in an agreement with the finding made by (Marek, 2018) where it is suggested that such close results suggest that the model’s bookmakers use to produce the probabilities are very similar to one another.

Figure 2.1 A

When looking closely at the results obtained by the use of a cross table in R studio (Figure 5 and 6) one can clearly see that the most correctly classified results are those matches who finished in wins by the home team. To compliment this argument are the findings provided by (*S. Mohammad Arabzad et al. 2014*), who highlight that in most cases the home team is predicted to outscore the away team, thus winning the match. This finding could also suggest that the home team is more likely to win a match. One factor that may cause such findings is the travelling involved. Away teams must travel to the home teams’ ground. Such argument is sustained by the finding presented by *(Goddard and Asimakopoulos, 2003)* where it is statedthat the home advantage may increase in cases where the teams are from opposite ends of a country (Thus having to travel further). *(Goddard and Asimakopoulos, 2003)* also note that such advantage loses some of its affect when the match is a derby between two local clubs. Another factor that may lead to such advantage for the home team could be the support during the matches played in the home stadium, which can boost home team morale while intimidating (in sports terms) the away team.

Figure 2.1 B

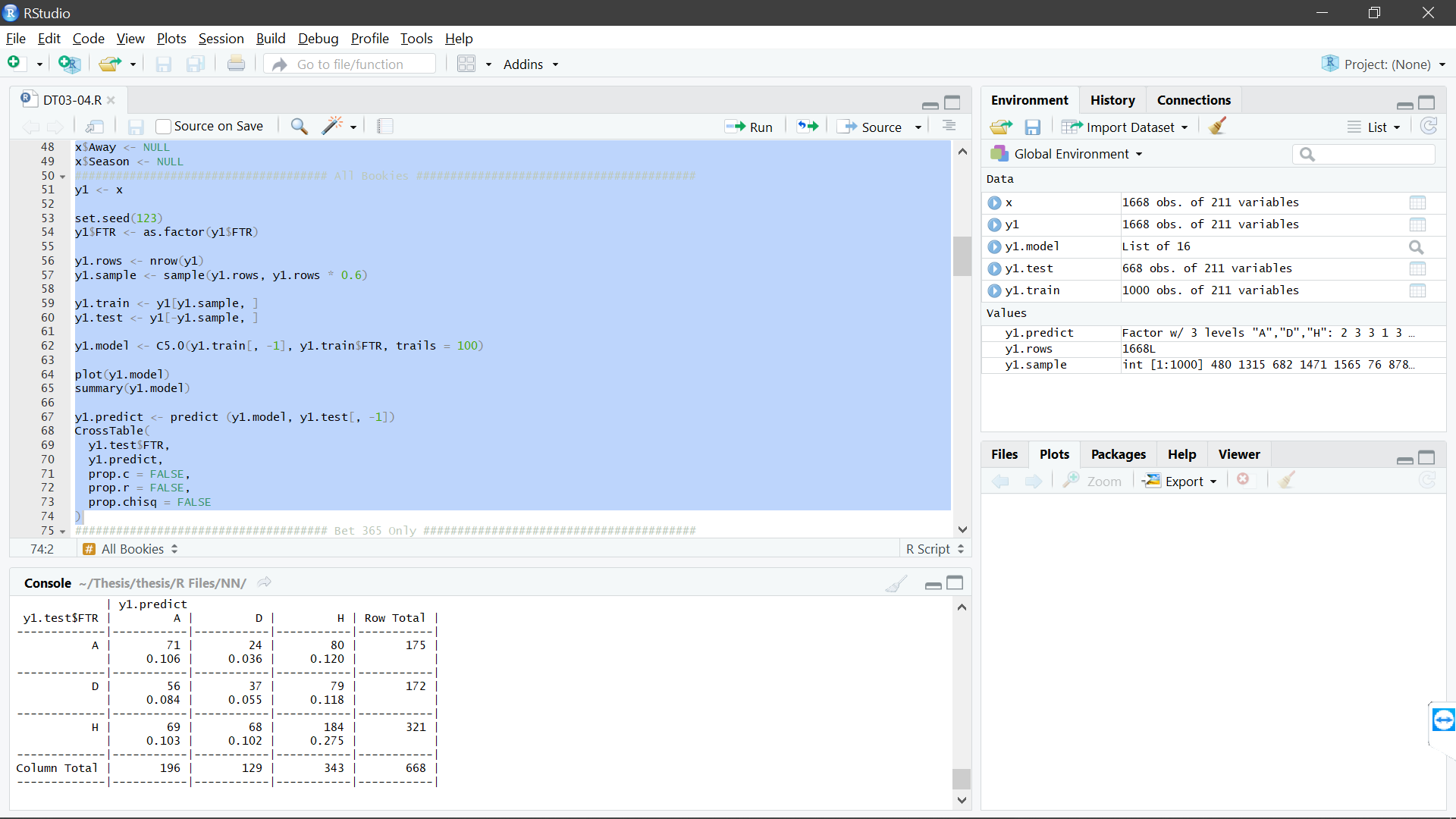
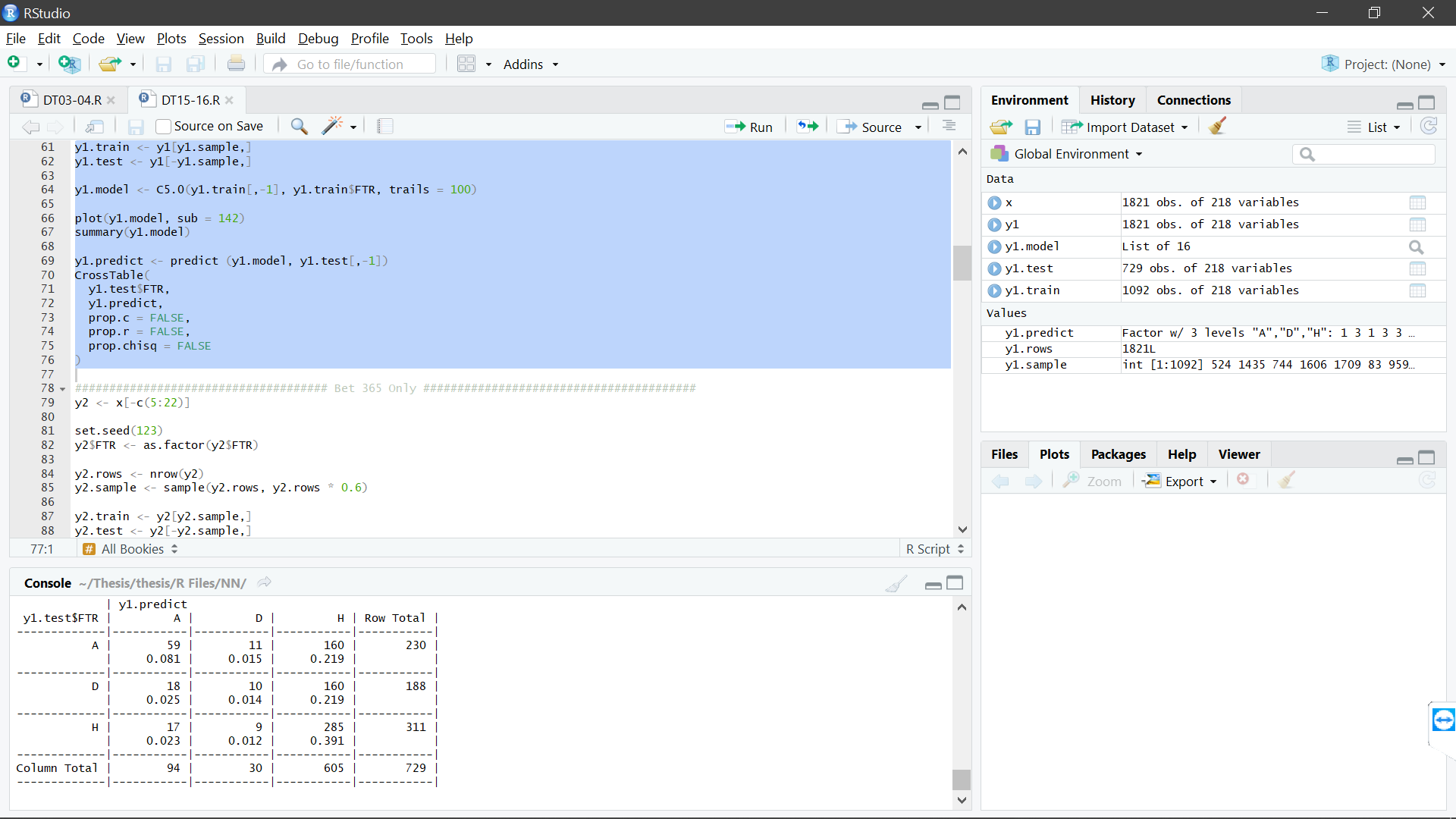


Figure 2.1 C



More often than not, while using the cross table to examine the results, the least correctly classified matches were those who finished as a draw. Obtaining such results could indicate that the draw odd has little predictive value, thus complimenting the statement by (Hvattum, 2013). (Hvattum, 2013) stated that odds on the draw outcome have no noteworthy predictive properties. Another indication could be that correctly predicting a draw is difficult since most of the matches in football, end up in a win for one of the sides involved.

## 2.2 ELO rating inclusive decision trees

To further enhance the model ELO ratings were added into the model. These were added to indicate the quality of the teams involved in the match, thus aiding the model to produce better classification results. As can be seen from the figure below, in most cases, decision trees with ELO rating performed better than those who didn’t. This means that if one were to add more data rather than having just the odds, better results can be obtained. However, it is worth noting that not in all cases did the decision tree perform better with the ELO than without them, especially when looking at individual betting companies. When looking at the results obtained by these decision trees it is clearly visible that the accuracy rate between the years involved, tends to fluctuate much more than that of decision trees without the provided ELO rating.

Figure 2.2 A

# 2.3 Leagues

Further investigation was carried out by investigating the behaviour of odds within a given league over the same period of the other models. From Figure 8 one can notice that although there are certain abnormal fluctuations, the percentage of accurately classified matches of all leagues, is similar to that of previous models. However, when going further into detail and looking at each of the leagues individually some factors become more apparent.

## 2.3.1 English Premier League

When looking at the English Premier league one can notice a slight dip in accuracy between the seasons of 2009/10 - 2012/13 and again between 2014/15 - 2015/16. These dips in percentage can be explained when looking at the league tables of those respective seasons. In 2009/10 a single point separated champions Chelsea F.C from second placed Man Utd. In 2011/12 even though the title was won with 9 points of a difference, both second and third place teams had the same amount of points (71) while the fourth had a mere three points less. This means that the race to finish in a champions league spot was very tight. The infamous 2012/13 season was won literally in the last minute as both teams from Manchester had the same amount of points, with Man City edging out as winners only on goal difference. Such findings are supportive of the argument presented by (Štrumbelj and Šikonja, 2010), who state that such close and hard predict results cause the odds to be more inaccurate than on other competitions. Causing the lower accuracy percentage in 2015/16 is Leicester city winning the premier league when odds on the same team becoming champions were 5000/1.

Figure 2.3.1 A

## 2.3.2 Serie A

When looking at the results obtained on the Italian Serie A, the results are contrasting to those in the premier league. Instead of a decrease, one can see an increase in accuracy in the last three years. The increase is of 21.8%, going from 37.4% in the season of 2014/15 up to 59.2% in the last season. This increase could be due to the gap in the quality between the participating teams of the Serie A in recent years. In the last 7 seasons included in this study (from 2011/12 onwards) only one team has been crowned champion of Italian football, with that team being Juventus. This finding is also compliant with the statement made by (Štrumbelj and Šikonja, 2010) mentioned prior. Since the results are quite predictable, the accuracy of odds has increased drastically.

Figure 2.3.2 A

## 2.3.3 German Bundesliga Leagues

When looking at the German Bundesliga, results show that in the first years of this study there were some fluctuations, but accuracy has stabilised around the 35%-40% percentile in the last 5 years, suggesting that there is a common level of quality between the German teams.

## 2.3.4 La Liga

As one can clearly notice at first glance, other than the two fluctuations in the 2009/10 and the following season, La Liga has a rather consistent percentage of accuracy. The fluctuations could have been caused by the fact that in the 2009/10 season the title was two-team race between Barcelona and Real Madrid. Barcelona won he league with 99 points while Real placed second with 96. The following season followed along the same lines, with Barcelona and Real finishing in the same positions and points identical to the season prior. Although the season of 2011/12 was no different for the two top teams, the difference between the third and sixth position, positions which allow qualification for European cup tournaments was only of 6 points

## 2.3.5 Ligue 1

Compared to all the results obtained above, Ligue 1 is the only league that does not have a single season with an accuracy of 50% or higher. This result is similar to the one obtained by (Štrumbelj and Šikonja, 2010). In their study (Štrumbelj and Šikonja, 2010) comment on how odds involving the French top division, just like in this study, performed the worst. Another observation on the results obtained is that the lowest percentage of accuracy obtained is also within this league. The accuracy for the 2001/02 season was that of 22.3%. The fact that the league winner was decided in the final match and high average of goals scored in each match (2.3) could have led to the poor predictive capability of odds that season. In the 2012/13 season, race for European qualification and the fight to avoid relegation was also fierce. This could have caused the drop in accuracy for that particular season.

# Bibliography

For Past League Standings:

<https://www.skysports.com/serie-a-table>

<https://www.premierleague.com/tables>

<https://www.ligue1.com/ligue1/classement?sai=70&journee1=1&journee2=34&cat=Gen>

<https://www.transfermarkt.com/primera-division/tabelle/wettbewerb/ES1/saison_id/2018>

References

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]

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].

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]