# Final Report

## Decision Trees VS Neural Networks

In this study two data mining techniques were used to compare and contrast the results obtained by both, this also led to the determination of which of the techniques is more suitable for this specific scenario. The same models were fed to the different techniques and the results pinpoint some differences. For this study it is clear that Neural Networks are better suited. Results in Figure 1. illustrate the average percentage of correctly predicted matches by both techniques. While Decision Trees fatigued to get results above 50%, Neural Networks managed to obtain high accuracy rates. The results presented above build on the positive results obtained in the study by *(S. Mohammad Arabzad et al. 2014). (S. Mohammad Arabzad et al. 2014),* using an artificial neural network managed to correctly guess both matches and final league positions. As discussed, results obtained throughout this study are complimentary to those obtained by *(S. Mohammad Arabzad et al. 2014)* and confirm the suitability of Neural Networks when making predictions related to Football matches.

*Fig 1.*

|  |  |  |
| --- | --- | --- |
|  | Neural Networks | Decision Trees |
| BPL | 87.30% | 47.44% |
| Serie A | 87.10% | 46.59% |
| Bundesliga | 88.95% | 43.33% |
| Ligue 1 | 85.55% | 42.43% |
| La Liga | 86.74% | 45.94% |

## Performance of odds has been consistent throughout the years

*Fig 2.*

The above figure (Fig 2.) illustrates the results obtained by the Neural Network, each season, for each of the domestic league included within the study. For the first few years results obtained inconsistent, and there were discrepancies between leagues. In the first season the highest percentage was of 64.5% correctly predicted matches in the English Premier League (BPL in Fig 2.), while the worst percentage (38.62%) was obtained on the Spanish La Liga, creating a gap of nearly 26% between the best and worst percentage. However, from the season of 2004-2005 these discrepancies are minimal. As can be seen from the figure the percentage of correctly predicted matches from the 2004 – 2005 season rarely go below the 90% for any league included. This illustrates that the predictive qualities of odds provided by betting companies has been consistent throughout the last 14 years. Such result is consistent with the statement made by *(Constantinou and Fenton, 2013),* who state that ability of odds has seen no improvement over the years.

## Bookmakers are not equally effective nor are they equally effective in different Leagues

Fig 3.

In the figure above, the average performance of the odds provided by each of the companies on each of the leagues involves are illustrated throughout the years. From this figure it is notable that bookmakers are not equally effective. Since there was a lack of data on the odds provided by comp 8 (Data was only available for a single season), for this scenario comp 8 was removed from Fig 4. Although odds provided by each of the companies in Fig 4., it is visible that comp2, 7 and 10 produce more accurate bets while companies 3 – 6 and 9 produce slightly inferior (less accurate) odds on average. While companies 2,7 and 10 produce odds that on average are 90%+ accurate, companies 3 – 6 do not achieve this percentage in any of the league while company 9 achieves 90.44% only on the Bundesliga. In spite of the fact that Company 1 goes slightly over 90% on three of the leagues; percentile is still inferior to that of companies 3,7 and 10. This suggests that not all companies produce equally effective odds. This finding is agreement with the statement made by *(Štrumbelj and Šikonja, 2010)* where they highlighted how the results they obtained revealed that odds in some leagues act as better forecasts than odds on other leagues.

Discrepancies can be seen on the percentage obtained by a single company on the different leagues. For example, company 6 had an average percentage of 86.82% correctly predicted matches on odds provided on the Bundesliga but an average of 82.21% on odds on Ligue 1. These discrepancies can between the percentages can be seen in the results obtained by each company. This result suggests that odds of a singular company are not equally effective amongst the different leagues.

Fig 4.

## Odds on the French Ligue 1 perform the worst.

Fig 5.

Fig 5. Is a graph of the results obtained from Fig 1. Fig 5 shows the average percentages of correctly predicted results in each of the leagues. As discussed above, odds perform differently in different leagues. The quality of the teams within the league can determine the performance of the odds. If a league has teams of equal quality than results will be harder to predict, whilst if a league has teams of unequal qualities these can make the end result more predictable. From Fig 5. It can be seen that the French League (Ligue 1) has the lowest percentage, both with Neural Networks and with Decision Trees. From this result it can be said that odds on Ligue 1 perform the worst when compaed to other leagues. This is in agreement with the findings presented by by *(Štrumbelj and Šikonja, 2010)* in their results stated that odds offered on Ligue 1 performed the worst. They also highlighted how, since the teams in Ligue 1 are more evenly matched, match outcomes are harder to predict.

# Draw Odds are relevant

Football matches tend to be won by either participating sides, and a draw result is infrequent. Throughout this study odds on the three possible results were observed. When looking at the results obtained in both of the data mining techniques used, draw matches were also being correctly predicted. In almost all the cases odds on the draw result are the highest in price, thus meaning that the result in less probable. Throughout the results outputted by the data mining techniques used, results that ended in the draw were correctly predicted as well. These results differ from the results provided by *(S. Mohammad Arabzad et al. 2014),* as in this study the Neural network did not manage to predict any draws. It also differs from the statement made by (Hvattum, 2012), who noted that odds on the draw outcome have no noteworthy predictive properties. Figure 6 and 7 are an example of the output provided by both data mining techniques. Fig 6 represents output of a decision tree on Ligue 1 of season 2010-2011, while figure 7 represent the output of the same league and season by a neural network. Green boxes indicate the correctly predicted outcomes

Fig 6.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Predicted  A | D | H |
| Actual |  |  |  |
| A | 0.7% | 12.4% | 11.7% |
| D | 6.6% | 21.9% | 8% |
| H | 2.9% | 21.9% | 13.9% |

Fig 7.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Predicted  A | D | H |
| Actual |  |  |  |
| A | 13.7% | 1% |  |
| D | 2% | 29.3% | 1.3% |
| H | 7.3% | 3.4% | 42% |

Bibliography

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]

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]

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>

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].