26th International Congress on Modelling and Simulation, Adelaide, SA, Australia, 30 Nov–4 Dec 2025
mssanz.org.au/modsim2025

Descriptive and Predictive Modelling for Professional Darts

Adam Aly [a] **and Hussein A Abbass** [b]

[a]*Australian National University, Canberra, Australia.*
[b]*School of Systems and Computing, University of New South Wales, ACT 2600, Australia*
Email: adamhusseinaly@gmail.com

Abstract: Darts is a two-player game. A match consists of 6 to 15 legs. Each player starts each leg with 501 points; we call it a budget. Players take turns. In each turn, each player throws three consecutive darts at the dartboard. After each throw, the points corresponding to the area the dart landed on the dartboard are subtracted from the budget. The winner of a leg is the first player who takes their budget down to exactly zero. Despite its popularity, only a few papers have been written on darts exploring gender differences and the influence of the first throw on the second throw. At the start of professional games, the TV displays certain statistics about each player. To validate how useful these statistics are, the aim of this paper is two-fold. First, we aim to develop descriptive models of the game with the aim of understanding the factors within a game that influence winning. In this regard, we use decision trees to model statistical features within a game. Second, we aim to identify early signs of winning in a game by developing predictive models of winning from statistical features that get collected only from the first leg in a game. Decision trees suggest that the accuracy of this predictive modelling approach is low. We conclude that the statistics displayed on the TV before a match starts, while useful, lack predictive powers. Therefore, these statistics could be misleading when attempting to predict the outcome of the game. Significant opportunities exist to advance game analytics for Darts, and more advanced analytics could provide fruitful insights into the game and player's strategies.

Keywords: *Darts, Decision Trees, Sport Analytics, Machine Learning*

1 INTRODUCTION

In the professional sports arena, there is an ongoing argument that the games, as they are being played today, are less enjoyable than the "golden era", pre in-depth sports analytics. The development of sports statistics and data analytics ultimately drive the way professionals play and practise sport today and will continue to do so in the future.

In pubs and bars around the world, visitors find the game of darts. While it is seen commonly as a so called pub game, in the past two decades the professional darts landscape has expanded significantly, and is now worldwide. In professional darts, the premier governing body today is the Professional Darts Cooperation (PDC). Historically, other organisations have hosted professional darts; however, today the PDC is seen as the premier organisation, hosting the best players in the world with significantly higher prize money, and an unparalleled worldwide reach. In recent decades, the sport of professional darts has spiked in viewership and global reach, inspiring a new generation of dart players.

A dartboard, see Figure 1, consists of 20 numbered slices (numbered 1 - 20) with a double ring around the outside of the board, where scores are worth double the number of the slice, and a triple ring between the double ring and the centre of the board, where scores are worth triple the number of the slice. The double bullseye is the small red section in the middle of the board; it is worth 50 points. It is surrounded in green by the single bullseye that is worth 25 points. If a dart hits the black section outside of the board, or hits a wire on the dartboard and bounces out, it is worth 0 points. The dart that takes a players budget to exactly 0, must be a double. This means it must land either in the double ring or in the double bullseye. Once a player reaches 0, or "checks out", they win what is known as a "leg". In the PDC, the majority of tournaments will play a best of x number of legs, where x ranges anywhere between 11 and 29. While some tournaments, most notably the darts world championship, play a best of x sets (similar to tennis), where a set is first to 2 or 3 winning legs, this paper will not deal with matches that engage in multiple sets.

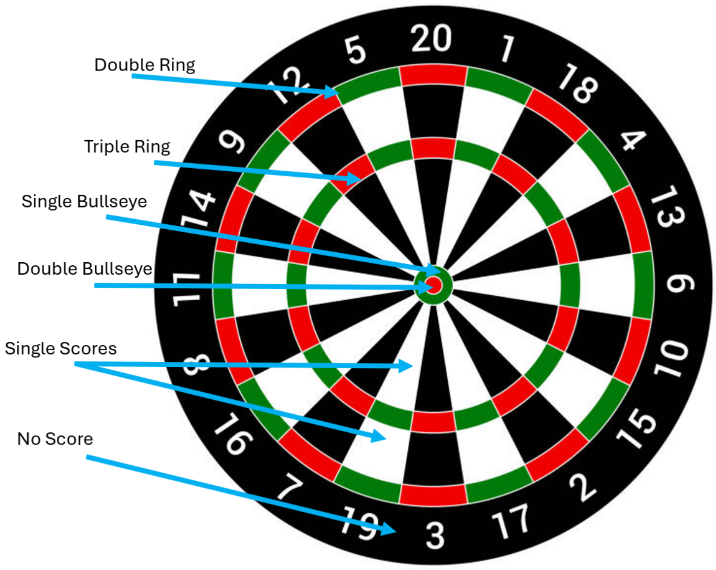

Figure 1. Labelled Dartboard

On TV, at the beginning of dart matches, statistics typically get shown before the match starts to compare the two players who are about to play. Statistics such as three and nine dart averages, as well as checkout percentages get displayed. Gambling odds are also typically visible on the screen, raising a question of the scientific grounds of these winning probabilities.

This paper has two aims. The first one is to validate if indeed the statistical feature of three and nine dart averages, and checkout percentages within a game contain sufficient predictive power of a game's outcome. This aim here is not to design a predictive model but rather a descriptive model that connects these statistics within a game and the game outcome. In our second aim, we design predictive models. The first uses statistics collected from the first leg of a game to predict the game outcome. The second uses statistics from historical games (last five) to predict a game outcome. The former model aims to predict the outcome of a game after watching one leg, while the latter aims to predict the outcome of a game before a game even starts.

2 LITERATURE REVIEW

There is not a particularly large field of research on the sport of darts, however, some notable studies stand out. Tibshirani et al. (2011) tackled the problem of maximising score based on a level of skill, developing a mathematical model that finds the expected score given a 1mm target range on the board, which is dependent on a players level of skill, measured as the mean distance away from the bullseye after throwing at it 50 times. Using this model, they developed a method to create personalised heat maps based on a 1mm grid of the dartboard. Further, in noticing the challenge in identifying a players skill, Haugh and Wang (2024) built a skill model using the Dispersed Measure (DM) distribution[1] for players aiming at a particular point, that could be used as input for Tibshirani *et.al*'s maximisation model.

Aside from this, there has been a significant research effort on the part of Linda Duffy, amongst other scholars, see for example Duffy (2002); Duffy et al. (2004, 2007), on the role of gender in darts, and whether variables such as recruitment rates, physical size or arm lengths can explain the disparity in darts performance between men and women. Ötting et al. (2020) study the effect of pressure on professional dart players, finding that they neither excel nor choke under pressure. In separate cognitive analysis, Kiliçarslan and Özant (2024) found that plays irrespective of age, gender, and education level enjoy physical activities highly. They also found that the more experience a player had, the more ability they had to reach a flow state when playing darts.

The above studies are all significant within the darts academic landscape for different reasons, but a particularly notable study that serves as the baseline for this paper is Liebscher and Kirschstein (2017)'s attempt to predict the outcome of dart tournaments. Within this, Liebscher and Kirschstein (2017) attempted to find indicators that can predict the result of single set dart matches. They estimate a binomial regression[2] model by applying a prohibit link function to estimate the probability of a player A winning.

$$P(Y = 1|X) = \Phi(X' \cdot \beta)$$

where X is the set of n covariates, and Φ is the cumulative distribution function of the standard normal distribution. They test four hypotheses:
H1: There is an advantage for the person who throws first (player A)
H2: The higher the advantage of player A in terms of legs, the higher player A's winning probability.
H3: The higher checkout rate of A compared to player B, the higher player A's winning probability.
H4: The higher the average points per turn (3 dart average) player A has compared to player B, the higher player A's winning probability.
They test these hypotheses using the glm function in the software package R by applying iterated least squares regression over four different models to find the variables that result in the most accurate generalised linear regression model. Their favouritism for checkout percentage per turn, rather than checkout percentage is a result of the performance of their model named IIa. This model using just leg advantage and checkout percentage resulted in a Pseudo-R^{2}[3] of 60%, but when replacing checkout percentage with checkout percentage per turn, the R^2 increased to 67%, which confirmed their suggestion that checkout percentage per turn is a better indicator than checkout percentage. Then, in their most successful model, they predicted player A's win probability by player A's leg advantage, the difference between the checkout per turn of player A and B, and the difference between the 3 dart averages of player A and B. This model had an Pseudo-R^2 of 79%. They go on to predict multi-set matches, and then tournaments, but these will not be discussed in this particular paper.

3 DATA ANALYSIS AND RESULTS

Liebscher and Kirschstein (2017) collected data from the website (https://live.dartsdata.com/); the quantity and quality of data collected is unparalleled in available resources today. However, this website and others like it

[1] A DM distribution uses measures of dispersion such as range, variance and standard deviation to describe the spread or grouping of data points in space

[2] A Binomial Regression model is used when the dependent variable follows a binomial distribution; for example, when its domain is binary. A transformation function is used to project the linear combination of the dependent variables as a probability, replacing the binary values with continuous variables representing their probabilities.

[3] When the independent variable is binary, a classic coefficient of determination, R^2, is replaced with a pseudo-R^2 on the transformed variable, for example, on the logistic regression.

do not exist currently. We obtained a copy of the data from Prof. Dr. Thomas Kirschstein's website at `https://prodlog.wiwi.uni-halle.de/team/ehemalige_mitarbeiter/kirschstein/`.

To ensure a fair reference to the work of Liebscher and Kirschstein (2017), we used the same dataset of 859 matches, removing multi-set matches, those after November 27th 2016, and matches resulting in a tie. Each row in the dataset represents an individual throw, resulting in over 220,000 rows (instances or records). We then wrote python code to convert the files so that each match represented by one row and four key statistics were generated during the conversion. These statistical attributes for each player and each match are three dart average, nine dart average, checkout percentage, and checkout percentage per turn.

"Three dart average" is arguably the most common statistic displayed when watching professional dart matches. It is calculated as the sum of the score of every throw in the match, divided by the number of throws in match, and then multiplied by three to approximate the average over three throws (one turn).

"Nine dart average" is similar, but it is the average score of the first nine throws in every leg, divided by three as to represent one turn. This statistic is typically used to represent a players' scoring ability, their ability to decrease their score to a place where they can check out from. The reason for using the first nine is because nine darts is the minimum number of darts a player needs to check out. Players must throw nine perfect darts including 8 triples and one double to checkout in the minimum number of throws. Therefore, the maximum nine dart average is 167.

"Checkout percentage" is another common statistic; it is the number of successful checkouts divided by the number of opportunities to check out. A check out opportunity is any time a player has an even score between 2 and 40 inclusive, or 50 and attempts to hit a double to effectively "checkout".

"Checkout Percentage per Throw" is a less common statistical attribute proposed by Liebscher and Kirschstein (2017), which they found was a better measure than simply checkout percentage. It is calculated by the number of successful checkouts divided by the number of turns where a player had an opportunity to checkout. This means that if a player had a score of 100 before their 13^{th} throw, they hit a triple-20, miss above the double-20 twice, and in their next turn they hit the double-20 on their second throw, they would have a checkout percentage of 33% and checkout percentage per turn of 50%.

These statistical attributes were generated for both players A and B in the match, where player A is the player who threw first in the first leg. At this point, each match was split into two rows, one for player A and a second one for player B. The following attributes were then inserted for each player; name, three dart average, opponent's three dart average, nine dart average, opponent's nine dart average, checkout percentage, opponent's checkout percentage, checkout percentage per turn, opponents checkout percentage per turn, difference in three dart average, difference in nine dart average, difference in checkout percentage, difference in checkout percentage per turn, and whether this player won (1) or not (0). We used WEKA 3.8.6 and the J48 algorithm which is an extension of the C4.5 algorithm designed by Quinlan (2014). We built eight different descriptive models and four different predictive models; key statistics about each model are shown in Table 1.

3.1 Descriptive Models

In the descriptive modelling phase, our aim is to establish the relationship between four attributes and game outcome. The four attributes are the three dart average, nine dart average, the checkout, and the checkout turn percentage. These four attributes were calculated for a player and the opponent. Our hypothesis is that if these statistics are useful, one would expect that there is a model that could relate the outcome of a game to these statistic. This exercise should not be seen as aiming at designing predictive models because the statistics are calculated within a full match; that is, it does not make sense to use them to predict the outcome of the game since this outcome would be known once the match has been played in the same way that these statistics become available once a match has been completed.

As shown in Table 1, the eight different descriptive models vary in the features that have been used. The "All Variables" used all four statistics for both players and the difference between players; that is in total, twelve features used. In the "Difference" model, only the difference between players was used, resulting in four features only, 35% reduction in the size of the decision tree, and an increase in accuracy to near perfection. This suggests that by simply subtracting Player B statistics from Player A for each of the four attributes, these differences are perfectly descriptive of game outcome. When contrasting these results to attributes for both players without including the difference and the attributes for the winning player alone, accuracy drops to 0.874 and 0.817, respectively, and the size of the decision tree increases by 75%. The "Play A's Statistics Only" suggests a reasonably high level of accuracy based on simply player A's statistics. However, given the

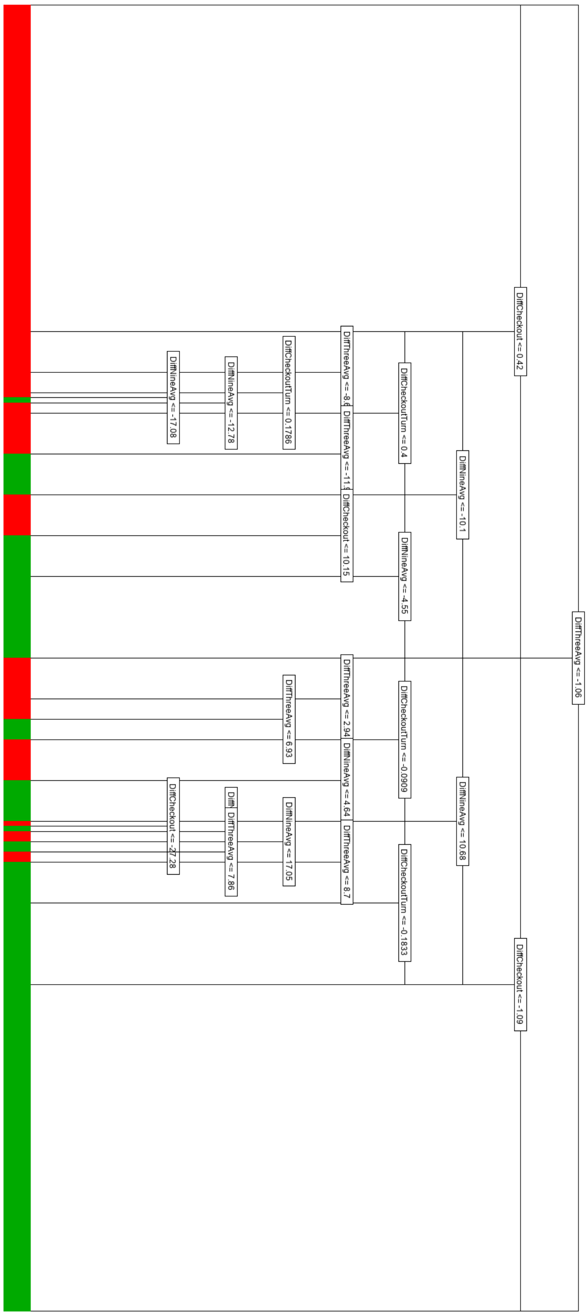

Figure 2. Descriptive Model - Difference

Table 1. Performance statistics for the eight descriptive and four predictive decision trees produced in this study.

Model Type	Features	Tree size	Leafs	Accuracy	Precision	Recall	F1
Descriptive	All Variables	73	37	0.912	0.952	0.974	0.963
Descriptive	Difference	47	24	0.989	0.904	0.892	0.898
Descriptive	All Variables No Diff	129	65	0.874	0.879	0.870	0.874
Descriptive	Play A's Statistics Only	129	65	0.817	0.797	0.852	0.824
Descriptive	Three Average	3	2	0.811	0.818	0.801	0.809
Descriptive	Nine Average	3	2	0.767	0.781	0.744	0.762
Descriptive	Checkout	15	8	0.827	0.831	0.822	0.826
Descriptive	Checkout Turn	41	21	0.786	0.792	0.779	0.785
Predictive	Leg-Level First Three Throws	89	45	0.654	0.676	0.869	0.761
Predictive	Match-Level First Leg	5	3	0.581	0.581	0.584	0.583
Predictive	Match-Level First Two Legs	7	4	0.663	0.657	0.685	0.670
Predictive	Match-Level Past Five Matches	21	11	0.633	0.637	0.633	0.632

complexity of the tree, it does not yield particular thresholds at which players achieve for each variable that result in a win. Rather the model is increasingly complex suggesting that thresholds do not exist at a high level of accuracy at which players should aim for to win a match. Consistent with the above finding, this suggests that variability of darts players performance on these indicators is high, which is why the Difference indicators outperform the others. The remaining descriptive models listed in Table 1 shows the impact of using one of the four attributes alone. Notable from the table is that checkout and three dart average have reasonable predictive power, while the nine dart average has the lowest one. Our findings contradict Liebscher and Kirschstein (2017) hypothesis that checkout percentage has a higher descriptive power than checkout percentage per turn.

The above discussion shows that these four attributes are indeed important. Knowing them for a particular match, one can infer the winner of that match. However, it is important to recognise that each of these attributes in the above analysis is calculated at the conclusion of a match, when the outcome is already determined. As such, while they carry information about the outcome, waiting to the end of the match reduces the viability of prediction since the actual outcome becomes a known.

3.2 Predictive Models

In this section, we turn our attention to prediction. The first model attempts to look at the first three throws in a leg to predict who won the leg. This anticipatory model has above-chance accuracy of 65%. The remaining predictive models focus on predicting the outcome of a match. The first model attempts to do so after the first leg, while the second model uses information from the first two legs. Both models achieved accuracy of 55% and 66%, respectively. When the first leg is completed, our ability to predict the outcome of the whole match is slightly better than chance (55%). This suggests that the results of the first leg are inconclusive, while waiting to the end of the second leg in a six-leg match, in two out of three matches, the prediction is likely to be correct; that is, the winner or loser of the match can be predicted with accuracy 66%..

The fourth predictive model shifts attention to one of the key motivations for this paper, are the statistics displayed before a match sufficiently informative to predict the outcome of a match. We used statistic of five matches prior to the one being played to predict the outcome of the one being played. While an accuracy of 62% exceeds chance, it still suggests that these statistics are not enough to calculate the odds. These statistics correlate well within a game, but they do not necessarily have strong predictive power for a future game. This also suggests significant variation in darts players performance match-to-match, given both players predictions were involved in this model.

4 Conclusion and Future Work

In this paper, we tested the hypothesis of whether information presented to viewers at the start of darts matches can actually predict the outcome of the match. We developed descriptive models which confirmed the associ-

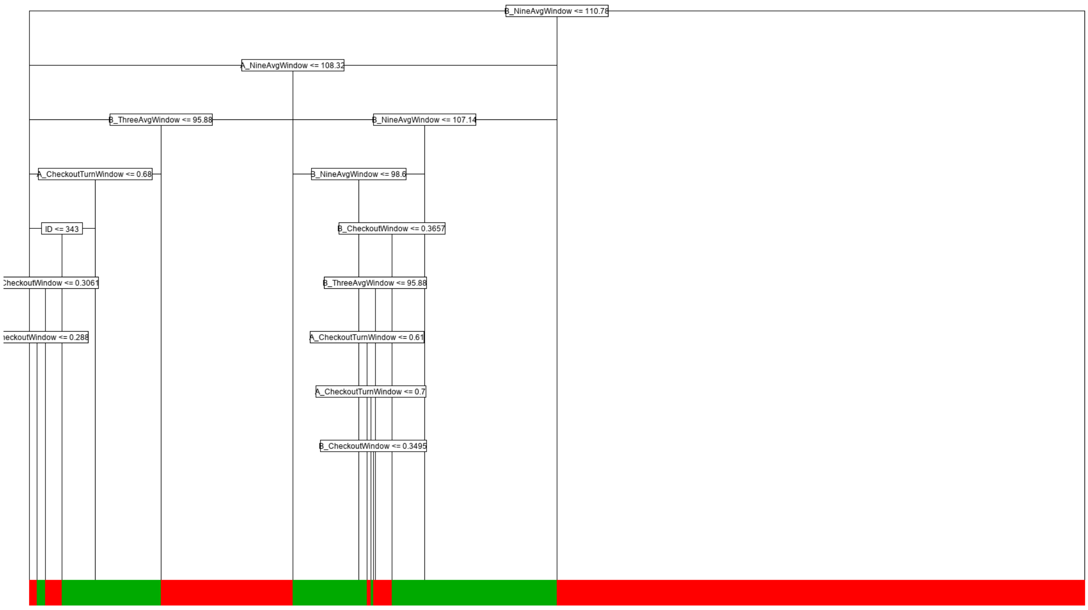

Figure 3. Predictive Model - Match-Level Past Five Matches

ation between the outcome of a match and this information. However, when attempting to predict the outcome of the match, we discovered that these indicators have little predictive power and are insufficient to achieve high predictions. Our future work will explore alternative features and statistics on darts to improve prediction. While perfect prediction is clearly unrealistic to expect, we hypothesise that a player's history needs to be combined with early throws in a match to correct for changes on the day.

REFERENCES

Duffy, L. (2002). *Gender differences in target throwing skills and dart playing performance: Evidence from elite dart players*. Ph. D. thesis, Middlesex University.

Duffy, L. J., B. Baluch, and K. A. Ericsson (2004). Dart performance as a function of facets of practice amongst professional and amateur men and women players. *Int. J. Sport Psychol 35*, 232–245.

Duffy, L. J., A. K. Ericsson, and B. Baluch (2007). In search of the loci for sex differences in throwing: The effects of physical size and differential recruitment rates on high levels of dart performance. *Research quarterly for exercise and sport 78*(2), 71–78.

Haugh, M. B. and C. Wang (2024). An empirical bayes approach for estimating skill models for professional darts players. *Journal of Quantitative Analysis in Sports 20*(4), 385–404.

Kiliçarslan, F. and M. İ. Özant (2024). Examination of flow experience and enjoyment levels of physical activity in dart playing youth in terms of some variables. *Pakistan Journal of Life and Social Sciences 22*(1).

Liebscher, S. and T. Kirschstein (2017). Predicting the outcome of professional darts tournaments. *International Journal of Performance Analysis in Sport 17*(5), 666–683.

Ötting, M., C. Deutscher, S. Schneemann, R. Langrock, S. Gehrmann, and H. Scholten (2020). Performance under pressure in skill tasks: An analysis of professional darts. *PLoS One 15*(2), e0228870.

Quinlan, J. R. (2014). *C4. 5: programs for machine learning*. Elsevier.

Tibshirani, R. J., A. Price, and J. Taylor (2011). A statistician plays darts. *Journal of the Royal Statistical Society Series A: Statistics in Society 174*(1), 213–226.