NBA Winning Team Prediction Modelling 2018

Frances Kim

March 2018

# Introduction

NBA is one of the world’s most watched games. Online betting on the sport’s winning team and/or margin is just as popular. The predictive modelling stems from the principle that certain attributes drive winning probability of a game. This causal relationship is already evident in the odds derived in sports betting markets. The objective of the codebook is to explain the data preparation, manipulation and modelling steps taken to predict the NBA game winning team and the margin by which game is won.

# Data Preparation

## Data source

There are 3 sources of input data.

* <http://zcodesystem.com/scorespredictor/>

predictive tool which provides score prediction and confidence levels. The tool is advertised as verifiable winning system and an industry standard in sports investing. The model is however a black box

* <http://zcodesystem.com/line_reversals>

indicator of major line movements (odd changes) and big money influxes throughout the day until game starts. It also displays the Vegas percentages and public betting percentage to allow anti-public betting strategies.

* <https://stats.nba.com/teams/traditional/?sort=W&dir=-1&Season=2017-18&SeasonType=Regular%20Season&LastNGames=1>

statistics of historical team performance by season. The statistics are dynamic therefore historic snapshots must be preserved in order for train dataset to contain true information.

## variables

There are 78 predictor variables and 9 dependent variables. The list of variable, description, type, format and calculation method is stated in Appedix1 (data variables). The variables are saved in excel worksheet.

Many team-level factors are accompanied by the team array which is an index of the two teams participating in a game. The arrays convert team names to either Team1 or Team2 of the game. This is to minimize modelling complexity arising when training set does not have enough fully-trained classifiers to predict test set. Combination of team names alone can create many unique records which are not re-usable for prediction of another unique and unseen test record.

Winning margin is calculated by (Team1’s score – Team2’ score) for both predictions and actuals. The sign is kept as it is an indicator of the winning team (if +, Team1 has won by the winning margin, if – , Team2 has won by the winning margin).

## missing data

This study is structured in a way not to tolerate missing data. If no information can be sourced, the specific record is excluded from modelling. This is warranted because the hybrid model includes random forest which does not accept null entries.

## data repository

data tables are saved in Excel in .xlsx format. In-season (therefore dynamic) variables 1-78 and 134 – 141 are saved in one worksheet. Raw game level team statistics are saved in different worksheets that derive team-level performance differencing statistics and Previous season statistics are also saved in a separate worksheet. Ref table is also created to convert team full names to three-letter acronyms. All necessary worksheets are read separately into R then merged in the model.

# model design

## model

The study aims to predict to objects: winning team and the winning margin. The winning team prediction is a classification problem that selects either of the two teams participating in a game and the winning margin is a regression problem that determines an integer which is the difference of Team1’s score – Team2’s score.

The classification forecast model is a hybrid of naïve bayes, random forest, support vector machine and gradient boosting. Ada boosting had been attempted but failed due to inappropriate data specification.

The regression forecast model is a random forest.

|  |  |
| --- | --- |
| Technique | Expected Output |
| Naïve bayes | Winning team array; factor {Team1, Team2} |
| Logistic regression | Probability of Team2 winning |
| Random forest (classification) | Probability of each team winning |
| Support vector machine | Winning team array; factor {Team1, Team2} |
| Gradient boosting | Probability of each team winning |
| Random forest (regression) | Numeric score (-inf, +inf) |

## code

The reader may refer to R code nba1.4.Rmd. The chunks largely separate to three sections: 1. web scraping, 2. hyper-parameter tuning, 3. Machine learning.

#### Web scraping

Rvest package & functions are used to download metrics from two zcode sources. This removed the need to manually copy and paste or write values into excel sheet. The zcode display changes based on dropdown menu that allows to select game type and play date. The conventional reading of url and embedded html does not work. Therefore, the user must first download or copy the html code of the relevant page to an offline copy and scrape values from the copy. The scraped metrics are merged into a data frame then exported to an excel sheet.

#### Hyper-parameter tuning

##### MLR package for model hyper-parameter tuning

Machine learning is sensitive to hyper-parameters which warrants optimization prior to implementing. MLR package is used to tune hyper-parameters in randomforest (classif), svm, gradient boost and randomforest (regression). The parameter values were then used to analyse & tune other specifications listed below.

##### Number of variable tuning

Several methods had been explored to optimise hyper-parameters used in the models, especially random forest. For the tuning exercise, existing train set is further divided into train (70%) and test set (30%). First, forecast error was measured against varying degree of polynomials, i.e. number of variables. For increasing degree of polynomials, random forest was fitted then the error in the cv and test sets was plotted against the degree of polynomials to create a learning curve. In Figure 1, red line denotes train error and black denotes test error. It is evident the in-sample error decreases as the number of variables increase however out-of- sample error shows less such trend. Over certain mark, the test error shows higher variance. Cubic function is fitted to the lines to find the number of features corresponding to the minimum error. The subsequent result suggested last season’s performance may not be significant in predicting current season’s game.

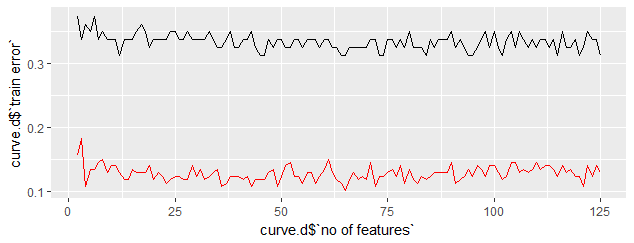


Figure 1

##### Sample balancing

The train dataset is skewed to contain more Team2 as winning teams at a ratio of 1:1.4. An analysis was conducted to find the optimal split between the counts of winning team in the train dataset that will yield the minimum prediction error. To do so, quantiles between the two counts of winning teams had been derived and those values were used to select the number of Team2 games to be trained on. The model error was then plotted against the undersampling size. It has been found the test error decreases as the sample size ratio is close to the original. Therefore no sample balancing will be applied.

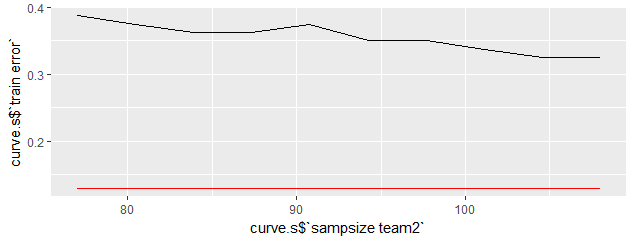


Figure 2

## Betting strategy

The rule of thumb is “the less game the user bets, the better” as the randomness increases with the number of games. If the user is confident on the betting strategy, he/she is better off betting amount on a game than splitting the bet amount to multiple games.

The final hybrid model excludes logistic regression as the technique did not yield good forecast accuracy.

To be considered a valid betting strategy, the minimum predicted winning probability must exceed 60% where the model output is given as a probability (0,1) – i.e. for random forest and gradient boosting. The user is recommended to bet only if the naïve bayes, random forest, svm and gradient boost prediction for a game had produced consistent (singular) winning team. The user is then allowed to place a moneyline or a runline with the additional margin deducted from the predicted winning team. Usually the moneyline odd is little attractive hence runline is preferred.

In the case of runline bet, the classification model is combined with the random forest regression model to generate winning team + winning margin. The margin can be used in two ways: add the margin to the losing team in expectation of the losing team excelling beyond the prediction; or deduct the margin from the winning team in expectation of the winning team excelling beyond the prediction. Based on the empirical study of historical forecast accuracies, the predicted margin is conservative therefore it is determined better off betting on the winning team minus the margin. Furthermore, it was of view that had the 4 hybrid model gave consistent prediction then it was robust enough to bet on the prediction, not the opponent. The forecast accuracy of the hybrid model is detailed in the forecast accuracy evaluation section.

## forecast accuracy evaluation

Accuracy measurement is based on 398 moneyline predictions and 248 runline prediction. Three evaluations are made: ML accuracy, RL accuracy independent of ML accuracy and combined accuracy as the final betting model. The model accuracies are compared to two benchmarks: 1. Zcode’s forecasting performance and 2. market prediction.

Table 1 displays the moneyline error in each forecast category as of 28th March 2018. As highlighted, betting on Team1 yielded in 67% accuracy and Team2 yielded in 75% accuracy. Against zcode’s forecasting, the hybrid model performed better, showing better accuracy over either team winning. Simultaneously the market performance is 67% accuracy on Team1 and 72% accuracy on Team2; the hybrid excelled again *(beat the market)*.[[1]](#footnote-1)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Proportion** | **Bet forecast** |  |  |  |
| **ML error** | **do not bet** | **Team1** | **Team2** | **Grand Total** |
| 0 | 0.00% | 67.19% | 75.16% | 40.45% |
| 1 | 100.00% | 32.81% | 24.84% | 59.55% |
| **Grand Total** | **100.00%** | **100.00%** | **100.00%** | **100.00%** |

Table 1

|  |  |  |  |
| --- | --- | --- | --- |
| **%** | **Zcode prediction** |  |  |
| **Actual win team** | **Team1** | **Team2** | **Grand Total** |
| Team1 | 57.35% | 29.01% | 38.69% |
| Team2 | 42.65% | 70.99% | 61.31% |
| **Grand Total** | **100.00%** | **100.00%** | **100.00%** |

Table 2

As of 28.03.2018, the runline accuracy is as per

|  |  |
| --- | --- |
| **RL error** | **Count %** |
| 0 | 53.52% |
| 1 | 46.48% |
| **Grand Total** | **100.00%** |

. Expectedly the accuracy drops as the predictions are more aggressive than moneyline bets. The accuracies however significantly surpasses that of zcode (Table 4) and if the odds are bigger than 1.5, the user is still able to make profit. Runline betting strategy is preferred if the bettor is willing to be more risk taking and potentially earn more profit in return through higher odds. The comparison against market prediction is not available due to non-collection of odds data that correspond to specific predicted margins.

|  |  |
| --- | --- |
| **RL error** | **Count %** |
| 0 | 53.52% |
| 1 | 46.48% |
| **Grand Total** | **100.00%** |

Table 3

|  |  |
| --- | --- |
| **Zcode RL error** | **Count %** |
| 0 | 13.67% |
| 1 | 86.33% |
| **Grand Total** | **100.00%** |

Table 4

Ultimately, RL bet is placed on top of ML bet as a leverage. The evaluation must therefore consider the final betting accuracy whereby prediction is deemed correct only when both moneyline and runline predictions are correct (i.e. an intersection); anything else is determined incorrect. Table 5 indicates the accuracy of the final betting decision. When the ultimate outcome is to bet on Team1, the historical accuracy is shown to be 56.7%; for Team2 – similar 57.3%. The accuracy drops further. Despite the degrading accuracy, the loss/profit can be offsetted of the odds of RL bets are guaranteed high.

|  |  |  |  |
| --- | --- | --- | --- |
| **Count of Error** | **Column Labels** |  |  |
| **Row Labels** | **Team1** | **Team2** | **Grand Total** |
| 0 | 56.76% | 57.29% | 57.14% |
| 1 | 43.24% | 42.71% | 42.86% |
| **Grand Total** | **100.00%** | **100.00%** | **100.00%** |

Table 5

## conclusion

This report explored how machine learning can be used to produce sports betting strategy and whether it displayed enough forecasting accuracy to make profit in real actions. The hybrid model consisting of naïve bayes, randomforest, svm and gradient boosting resulted in better than market & input data predictions.

# Appendix

## data variables

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Index | Variable Name | Description | Type | Format | Calculation |
| 1 | DataType | Metadata which determines modelling domain | Factor | {Archive, Train, Test, MLTest} | Historic records <= 3 mths old are default to Train. Older records are archived & not modelled. |
| 2 | Team1 | NBA away team name | Factor | Character - 3 | As per the team mention order in zcode |
| 3 | Team2 | NBA home team name | Factor | Character - 3 | As per the team mention order in zcode (hometeam) |
| 4 | Date | Game Date (US) | Date | Date |  |
| 5 | Index | Unique Game ID | Character | Character | Concatenation of Team1, Team2 Date. Primary key of the data table |
| 6 | Home | NBA home team name | Factor | Character – 3 | Same as Team2 |
| 7 | HomeTeam | Home team array | Factor | {Team1, Team2} | Team2 by structure |
| 8 | Favorite | Favorite team name | Factor | Character – 3 | Team with lower odd at the time of modelling |
| 9 | FavoriteTeam | Favorite team array | Factor | {Team1, Team2} |  |
| 10 | Odd1 | Odd of Team1 winning | Numeric | Continuous | Team 1’s winning odd at the point of data collection |
| 11 | Odd2 | Odd of Team2 winning | Numeric | Continuous | Team2’s winning odd at the point of data collection |
| 12 | Odddiff | Odd1 – Odd2 | Numeric | Continuous | Difference of the odds |
| 13 | Antipublic | Name of team with lower ML Public % | Factor | Character – 3 | Team with smaller portion of ML tickets sold |
| 14 | AntipublicTeam | Antipublic team array | Factor | {Team1, Team2} | Antipublic team array |
| 15 | Sharpline1 | Number of Team1’s sharpline movements | Numeric | Integer | Sum of indicators in the two charts  {sharp line move in favor of the team, sharp line move against the opponent} |
| 16 | Smartmoney1 | Number of Team1’s smart money indicator | Numeric | Integer | Sum of indicators in the two charts { smart money coming on } |
| 17 | Sharpline2 | Number of Team2’s sharpline movements | Numeric | Integer | Sum of indicators in the two charts  {sharp line move in favor of the team, sharp line move against the opponent} |
| 18 | Smartmoney2 | Number of Team2’s smart money indicator | Numeric | Integer | Sum of indicators in the two charts { smart money coming on } |
| 19 | Winpred | Zcode predicting winning team name | Factor | Character – 3 | Winning by Zcode score prediction |
| 20 | WinpredTeam | Winpred team array | Factor | {Team1, Team2} |  |
| 21 | Scorepred1 | Team1 score prediction | Numeric | Integer | Absolute winning team bet is determined by the higher score prediction value |
| 22 | Scorepred2 | Team2 score prediction | Numeric | Integer | Absolute winning team bet is determined by the higher score prediction value |
| 23 | Scorepreddiff | Scorepred1 – Scorepred2 | Numeric | Integer | Difference of the score predictions |
| 24 | Scorepredconf | Score prediction confidence level | Numeric | Decimal |  |
| 25 | Pred1 | Team1 win betting confidence | Numeric | Decimal |  |
| 26 | Pred2 | Team2 win betting confidence | Numeric | Decimal | Pred1 + Pred2 = 1 |
| 27 | GP\_ALLdiff | Difference in the number of games played | Numeric | Integer | Sum of all games played in the season. Difference is calculated by Team1 stat – Team2 stat |
| 28 | W\_ALLdiff | Difference in the number of wins | Numeric | Integer | As above |
| 29 | L\_ALLdiff | Difference in the number of losses | Numeric | Integer | As above |
| 30 | WINPER\_ALLdiff | Difference in the win percentages | Numeric | Continuous | As above |
| 31 | MIN\_ALLdiff | Difference in the minutes played | Numeric | Continuous | Mean of all games played in the season. Difference is calculated by Team1 stat – Team2 stat |
| 32 | PTS\_ALLdiff | Difference in the sum of points | Numeric | Continuous | As above |
| 33 | FGM\_ALLdiff | Difference in the number of field goals made | Numeric | Continuous | As above |
| 34 | FGA\_ALLdiff | Difference in the number of field goals attempted | Numeric | Continuous | As above |
| 35 | FGPER\_ALLdiff | Difference in the field goal success percentage | Numeric | Continuous | As above |
| 36 | THREEPM\_ALLdiff | Difference in the 3 pt field goals made | Numeric | Continuous | As above |
| 37 | THREEPA\_ALLdiff | Difference in the 3 pt field goals attempted | Numeric | Continuous | As above |
| 38 | THREEPPER\_ALLdiff | Difference in 3 pt field goals success percentage | Numeric | Continuous | As above |
| 39 | FTM\_ALLdiff | Difference in the number of free throws made | Numeric | Continuous | As above |
| 40 | FTA\_ALLdiff | Difference in the number of free throws attempted | Numeric | Continuous | As above |
| 41 | FTPER\_ALLdiff | Difference in free throw success percentage | Numeric | Continuous | As above |
| 42 | OREB\_ALLdiff | Difference in the number of offensive rebounds | Numeric | Continuous | As above |
| 43 | DREB\_ALLdiff | Difference in the number of defensive rebounds | Numeric | Continuous | As above |
| 44 | REB\_ALLdiff | Difference in the number of rebounds | Numeric | Continuous | As above |
| 45 | AST\_ALLdiff | Difference in the number of assists | Numeric | Continuous | As above |
| 46 | TOV\_ALLdiff | Difference in the number of turnovers | Numeric | Continuous | As above |
| 47 | STL\_ALLdiff | Difference in the number of steals | Numeric | Continuous | As above |
| 48 | BLK\_ALLdiff | Difference in number of blocks | Numeric | Continuous | As above |
| 49 | BLKA\_ALLdiff | Difference in the number of blocked field goal attempts | Numeric | Continuous | As above |
| 50 | PF\_ALLdiff | Difference in number of personal fouls | Numeric | Continuous | As above |
| 51 | PFD\_ALLdiff | Difference in number of personal fouls drawn | Numeric | Continuous | As above |
| 52 | PLUSMINUSdiff | Difference in plus minus | Numeric | Continuous | As above |
| 53 | GP\_LST1diff | Difference in the number of games played | Numeric | Discrete | Of last game played in the season. Difference is calculated by Team1 stat – Team2 stat |
| 54 | W\_LST1diff | Difference in the number of wins | Numeric | Discrete | As above |
| 55 | L\_LST1diff | Difference in the number of losses | Numeric | Discrete | As above |
| 56 | WINPER\_LST1diff | Difference in the win percentages | Numeric | Discrete | As above |
| 57 | MIN\_LST1diff | Difference in the minutes played | Numeric | Discrete | As above |
| 58 | PTS\_LST1diff | Difference in the sum of points | Numeric | Discrete | As above |
| 59 | FGM\_LST1diff | Difference in the number of field goals made | Numeric | Discrete | As above |
| 60 | FGA\_LST1diff | Difference in the number of field goals attempted | Numeric | Discrete | As above |
| 61 | FGPER\_LST1diff | Difference in the field goal success percentage | Numeric | Discrete | As above |
| 62 | THREEPM\_LST1diff | Difference in the 3 pt field goals made | Numeric | Discrete | As above |
| 63 | THREEPA\_LST1diff | Difference in the 3 pt field goals attempted | Numeric | Discrete | As above |
| 64 | THREEPPER\_LST1diff | Difference in 3 pt field goals success percentage | Numeric | Discrete | As above |
| 65 | FTM\_LST1diff | Difference in the number of free throws made | Numeric | Discrete | As above |
| 66 | FTA\_LST1diff | Difference in the number of free throws attempted | Numeric | Discrete | As above |
| 67 | FTPER\_LST1diff | Difference in free throw success percentage | Numeric | Discrete | As above |
| 68 | OREB\_LST1diff | Difference in the number of offensive rebounds | Numeric | Discrete | As above |
| 69 | DREB\_LST1diff | Difference in the number of defensive rebounds | Numeric | Discrete | As above |
| 70 | REB\_LST1diff | Difference in the number of rebounds | Numeric | Discrete | As above |
| 71 | AST\_LST1diff | Difference in the number of assists | Numeric | Discrete | As above |
| 72 | TOV\_LST1diff | Difference in the number of turnovers | Numeric | Discrete | As above |
| 73 | STL\_LST1diff | Difference in the number of steals | Numeric | Discrete | As above |
| 74 | BLK\_LST1diff | Difference in number of blocks | Numeric | Discrete | As above |
| 75 | BLKA\_LST1diff | Difference in the number of blocked field goal attempts | Numeric | Discrete | As above |
| 76 | PF\_LST1diff | Difference in number of personal fouls | Numeric | Discrete | As above |
| 77 | PFD\_LST1diff | Difference in number of personal fouls drawn | Numeric | Discrete | As above |
| 78 | PLUSMINUSLST1diff | Difference in plus minus | Numeric | Discrete | As above |
| 79 | Id | Order game number | Numeric | Discrete | Arbitrarily assigned for unique id assignment |
| 80 | RANK.1 | Team1’s rank | Numeric | Discrete | Of previous season |
| 81 | GP.1 | Team1’s number of games played | Numeric | Discrete | As above |
| 82 | W.1 | Team1’s number of wins | Numeric | Discrete | As above |
| 83 | L.1 | Team1’s number of losses | Numeric | Discrete | As above |
| 84 | WINPER.1 | Team1’s win percentage | Numeric | Continuous | As above |
| 85 | MIN.1 | Team1’s minutes played | Numeric | Continuous | Mean of previous season  games |
| 86 | PTS.1 | Team1’s points | Numeric | Continuous | As above |
| 87 | FGM.1 | Team1’s number of field goals made | Numeric | Continuous | As above |
| 88 | FGA.1 | Team1’s number of field goal attempts | Numeric | Continuous | As above |
| 89 | FGPER.1 | Team1’s field goal success percentage | Numeric | Continuous | As above |
| 90 | THREEPM.1 | Team1’s number of 3 pt field goals made | Numeric | Continuous | As above |
| 91 | THREEPA.1 | Team1’s number of 3 pt field goal attempts | Numeric | Continuous | As above |
| 92 | THREEPER.1 | Team1’s 3 pt field goal success percentage | Numeric | Continuous | As above |
| 93 | FTM.1 | Team1’s number of free throws made | Numeric | Continuous | As above |
| 94 | FTA.1 | Team1’s number of free throw attempts | Numeric | Continuous | As above |
| 95 | FTPER.1 | Team1’s free throw success percentage | Numeric | Continuous | As above |
| 96 | OREB.1 | Team1’s number of offensive rebounds | Numeric | Continuous | As above |
| 97 | DREB.1 | Team1’s number of defensive rebounds | Numeric | Continuous | As above |
| 98 | REB.1 | Team1’s number of rebounds | Numeric | Continuous | As above |
| 99 | AST.1 | Team1’s number of assists | Numeric | Continuous | As above |
| 100 | TOV.1 | Team1’s number of turnovers | Numeric | Continuous | As above |
| 101 | STL.1 | Team1’s number of steals | Numeric | Continuous | As above |
| 102 | BLK.1 | Team1’s number of blocks | Numeric | Continuous | As above |
| 103 | BLKA.1 | Team1’s number of blocked field goal attempts | Numeric | Continuous | As above |
| 104 | PF.1 | Team1’s number of personal fouls | Numeric | Continuous | As above |
| 105 | PFD.1 | Team1’s number of personal fouls drawn | Numeric | Continuous | As above |
| 106 | PLUSMINUS.1 | Team1’s plus minus | Numeric | Continuous | As above |
| 107 | RANK.2 | Team2’s rank | Numeric | Discrete | Of previous season |
| 108 | GP.2 | Team2’s number of games played | Numeric | Discrete | As above |
| 109 | W.2 | Team2’s number of wins | Numeric | Discrete | As above |
| 110 | L.2 | Team2’s number of losses | Numeric | Discrete | As above |
| 111 | WINPER.2 | Team2’s win percentage | Numeric | Continuous | As above |
| 112 | MIN.2 | Team2’s minutes played | Numeric | Continuous | Mean of previous season  games |
| 113 | PTS.2 | Team2’s points | Numeric | Continuous | As above |
| 114 | FGM.2 | Team2’s number of field goals made | Numeric | Continuous | As above |
| 115 | FGA.2 | Team2’s number of field goal attempts | Numeric | Continuous | As above |
| 116 | FGPER.2 | Team2’s field goal success percentage | Numeric | Continuous | As above |
| 117 | THREEPM.2 | Team2’s number of 3 pt field goals made | Numeric | Continuous | As above |
| 118 | THREEPA.2 | Team2’s number of 3 pt field goal attempts | Numeric | Continuous | As above |
| 119 | THREEPER.2 | Team2’s 3 pt field goal success percentage | Numeric | Continuous | As above |
| 120 | FTM.2 | Team2’s number of free throws made | Numeric | Continuous | As above |
| 121 | FTA.2 | Team2’s number of free throw attempts | Numeric | Continuous | As above |
| 122 | FTPER.2 | Team2’s free throw success percentage | Numeric | Continuous | As above |
| 123 | OREB.2 | Team2’s number of offensive rebounds | Numeric | Continuous | As above |
| 124 | DREB.2 | Team2’s number of defensive rebounds | Numeric | Continuous | As above |
| 125 | REB.2 | Team2’s number of rebounds | Numeric | Continuous | As above |
| 126 | AST.2 | Team2’s number of assists | Numeric | Continuous | As above |
| 127 | TOV.2 | Team2’s number of turnovers | Numeric | Continuous | As above |
| 128 | STL.2 | Team2’s number of steals | Numeric | Continuous | As above |
| 129 | BLK.2 | Team2’s number of blocks | Numeric | Continuous | As above |
| 130 | BLKA.2 | Team2’s number of blocked field goal attempts | Numeric | Continuous | As above |
| 131 | PF.2 | Team2’s number of personal fouls | Numeric | Continuous | As above |
| 132 | PFD.2 | Team2’s number of personal fouls drawn | Numeric | Continuous | As above |
| 133 | PLUSMINUS.2 | Team2’s plus minus | Numeric | Continuous | As above |
| 134 | Gamewin | Actual game win team name | Factor | Character – 3 |  |
| 135 | GamewinTeam | Actual game win team array | Factor | {Team1, Team2} |  |
| 136 | Score1 | Actual Team1’s score | Numeric | Discrete |  |
| 137 | Score2 | Actual Team2’s score | Numeric | Discrete |  |
| 138 | Scorediff | Score1 – Score2 | Numeric | Discrete |  |
| 139 | Bet.forecast | Final forecast | Factor | {Team1, Team2, Do Not Bet} | Final moneyline betting forecast |
| 140 | Runline.forecast | Final winning margin forecast | Numeric | Discrete | Final winning margin betting forecast |
| 141 | Runline.error | Accuracy of winning margin forecast | Factor | {0,1} | 0 = true if margin forecast is right sign (i.e. margin in favor of actual winning team) and less than or equal to the actual winning margin otherwise 1 = false |
| 142 | Error | Accuracy of winning team forecast | Factor | {0,1} | 0 = true if winning team forecast is actual otherwise 1 = false. If betting strategy is “Do Not Bet” error value is 1. |

1. To calculate the market predictability, in a given dataset calculate # of games where Team1/2’s odd is less than 2 and ratio of those whose actual winning team is Team1/2 respectively. [↑](#footnote-ref-1)