Using Bookmaker Odds to Predict the Final Result of Football Matches

2 authors, including:

Jacek Grekow
Bialystok University of Technology
30 PUBLICATIONS   **60** CITATIONS

SEE PROFILE

Using bookmaker odds to predict the final result of football matches

Karol Odachowski, Jacek Grekow

Faculty of Computer Science, Bialystok University of Technology,
Wiejska 45A, Bialystok 15-351, Poland
odach@wp.pl, j.grekow@pb.edu.pl

Abstract. There are many online bookmakers that allow betting money in virtually every field of sports, from football to chess. The vast majority of online bookmakers operate based on standard principles and establish the odds for sporting events. These odds constantly change due to bets placed by gamblers. The amount of changes is associated with the amount of money bet on a given odd. The purpose of this paper was to investigate the possibility of predicting how upcoming football matches will end based on changes in bookmaker odds. A number of different classifiers that predict the final result of a football match were developed. The results obtained confirm that the knowledge of a group of people about football matches gathered in the form of bookmaker odds can be successfully used for predicting the final result.

Keywords: bookmaker odds, feature extraction, classification, forecasting, sports betting.

1 Introduction

The purpose of this paper is to investigate the possibility of predicting how upcoming sporting events will end based on changes in bookmaker odds. Football was the sport chosen for observation of changes in bookmaker odds. It should be assumed that if a gambler risks his own money, he has reasons to place such a bet. The greater the amount of gambled funds, the greater the change of the odds and greater possibility that the bet was based on factual knowledge about the competing teams, the status of the players, games played, etc. Predictions of the result can be based on such types of information. If the research should provide promising results, one might be tempted to build a decision-making system that could allow predicting final results based on observation of fluctuations of odds.

2 Previous Works

There are several papers that have dealt with similar problems of analysis and prediction of sporting event results. They are based on various types of data such as expert knowledge, results of previous matches, rankings of teams or bookmaker odds.

A group of papers directly referring to this paper are those addressing the problem of using data mining techniques to predict the final result of a sports match. An analysis of data of National Basketball Association (NBA) seasons was used to develop the expert system, which predicts the winner in a sport game [1]. The analyzed data contained detailed statistics of each game played during a season. The best accuracy (67%) was achieved by a classifier built using a multinomial logistic regression model with a ridge estimator. Miljkovic et al. [2] presents a system that uses data mining techniques in order to predict the outcomes of basketball games in the NBA league. To predict the game result the Naive Bayes method is used. Besides the actual result, the system calculates the spread for each game by using multivariate linear regression. Each game was described with attributes composed of the standard basketball statistics (field goals made, field goals attempted, 3 pointers, free throws, rebounds, blocked shots, fouls, etc), and information about league standings (number of wins and losses, home and away wins, current streak etc). The system correctly predicted the winners of about 67% of the matches. McCabe and Trevathan [3] used Artificial Neural Networks to predict games. They used attributes that indicate the quality of a particular team and achived 54.6% correct predictions for the English Football Premier League and 67.5% for Super Rugby. Smith et al. [4] used the Bayesian classifier to predict Cy Young Award winners in American baseball. The model was crated based on player statistics data collected for baseball seasons from 1967 to 2006. The accuracy of the Bayesian classifier was more than 80% correct.

3   Classic Football Bets of 1-X-2 Type

This work focuses on classic football bets of 1-X-2 type. For example, the result is a win of the first team, second team or a draw. Because of the three possible endings for the match, a three-way 1-X-2 bet, where "1" is an odd for the home team, "X" a draw, "2" odds for the away team, is used in football betting. The home team is the one that plays at home, while the visiting team is the away team. The classic bets are regarded as winning if the selected result is correct.

For example, a football match: Tottenham Hotspur vs. Chelsea:

– Bet on Tottenham Hotspur (type 1) will be settled as a win if Tottenham Hotspur wins. If Chelsea wins or there is a tie, the bets will be settled as a loss. It would be the same with a bet on Chelsea (type 2).
– A bet on a draw (type X) will be settled as a winning bet only in the event of the match ending in a tie.

4   Input Data

The input data describing the changes in bookmakers odds was obtained from the PinnacleSports [5] website, which makes public any information about sporting events in a clear form of an XML document. The XML file can be found at

`http://www.xml.pinnaclesports.com/pinnacleFeed.asp`. This is a static file, updated every 10 minutes. The process of importing data from an XML file consisted in tracking its contents for the last 10 hours preceding the football game and recording data on the changing odds. Additionally, the input data had to be supplemented with the final result of the matches. Due to the fact that Pinnacle Sports does not provide such data, we imported it from another source: `Betfair.com` [6]. We collected the input data for a period of six months, from a total of 2615 matches.

4.1 Feature Extraction

Every game, which is an independent instance included in the input data of the decision-making system, was described by a set of features. They reflected significant changes in bookmaker odds, which may affect the final result of the match.

We analyzed the overall level of changes in bookmaker odds of football games, which could determine the path of further research. For this purpose sample graphs showing the odds over time were analyzed. A period of 10 hours of sampling before the match was taken into consideration, because in this period the greatest fluctuations of the odds occurred. The time interval between successive samples was 10 minutes.
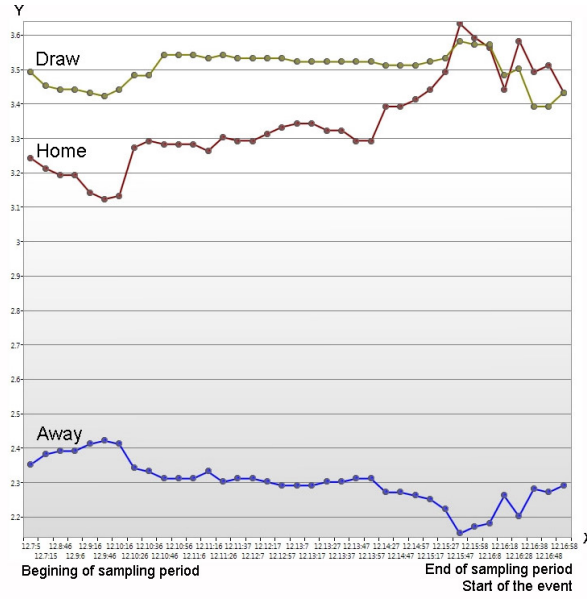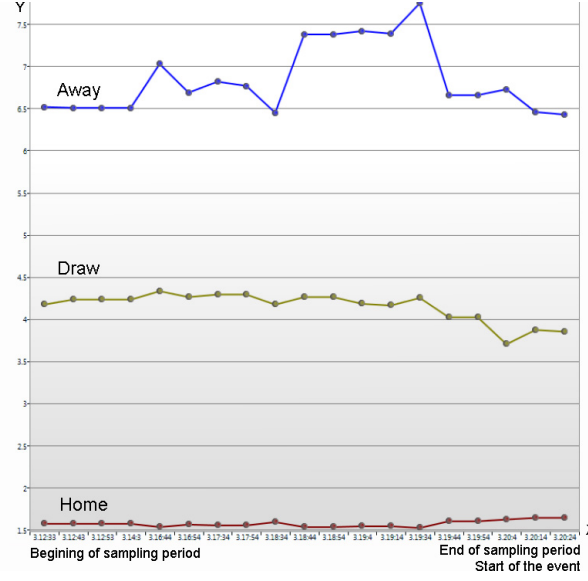

Fig. 1. Sample chart of odds changes (home, away, draw) in 1-X-2 type bets

Fig. 2. Odd changes for the Racing Genk vs. Loceren (2:1) match

We observed that the closer to the start of the match, the more changes in the odds occurred. Figure 1 illustrates such a situation. This is a chart of values of the odds for the home team, the visiting team, and a draw over time (Y axis) during the last 10 hours (X axis) before the Tottenham Hotspur vs. Chelsea match, which was held on 12th December 2010 and ended with a 1-1 draw. Figure 2 presents another example of changes in bookmaker odds for the Racing Genk vs. Loceren (2:1) match, which was held on 3th April 2011.

We decided that it would be justified to divide the sampling period into several smaller ones, because the irregularity of the distribution of the changes may indicate that the entire sampling period does not have the same effect on the final result. For each period we generated the same set of features. Additionally, the entire sampling period was also taken into account. This allowed us to extract general information about the match. Figure 3 shows a schematic diagram of such a division.
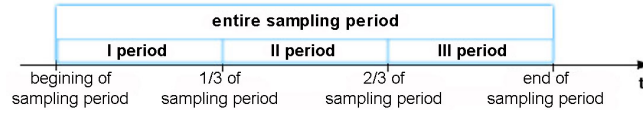
Fig. 3. Division of the sampling period

We examined three data sets:

1. Set of odds for a home team win;
2. Set of odds for a away team win;
3. Set of odds for a draw.

Regardless of the data set, we investigated four sampling periods:

− the entire 10-hour sampling period prior to the match;
− 1st sampling period correlating to the first 3 hours and 20 minutes;
− 2nd sampling period beginning on the 21st minute of the third hour and also lasting 3 hours and 20 minutes;
− 3rd sampling period correlating to the last 3 hours and 20 minutes before the start of the match.

For each set of data in each sampling period we generated a set of 24 *standard features*, which include: the minimum value; the minimum value given as a percentage; the maximum value; the maximum value given as a percentage; the value of the arithmetic mean; arithmetic mean given as a percentage; the number of odds with different values; standard deviation; initial value; initial value given as a percentage; final value; final value given as a percentage; the difference in the initial and final value; the difference in the initial and final values given as a percentage; angle between the horizontal line and a line drawn from the initial to the maximum value; angle between the horizontal line and a line drawn from the initial to the minimum value; minimum value of the derivative; maximum value of the derivative; arithmetic mean of the derivative; standard deviation of the derivative; initial value of the derivative; final value of the derivative; the difference between the initial and the final values of the derivative; the number of different values of the derivatives.

Additionally, a single sampling period contained eight *general features* that apply to all three data sets simultaneously: minimum limit of money; maximum limit of money; arithmetic mean of the limit of money; nominal feature which based on the arithmetic mean value of the odds determines the favored team; nominal feature determining the favored team at the beginning of the sampling period; nominal feature determining the favored team at the end of the sampling period; nominal feature determining the team that recorded the biggest odds drop between the beginning and the end of the sampling period; nominal feature determining the team that recorded the biggest odds drop between successive sampling periods.

The number of features determined for a single sampling period is equal to 80. This is the sum of the *general features* (8) and the product of the number of features included in the set of *standard features* (24) with the amount of data sets (3). We get a total of 320 features from the four sampling periods.

4.2   Data in ARFF Format

To use the collected input data about the matches in the decision-making process, the values of all the features describing a particular match were determined and later recorded in the ARFF file format. The last declared attribute (feature)

in this file is the decision class, which is the result of the match and adopts the nominal values from the set: Win-home, Win-away, Win-draw. It defines the final outcome of the match. For the input data prepared in such a manner, classifiers were developed allowing to predict the final result. To analyze the data and the development of classifiers, a data mining task software WEKA [7] was used. Cross-Validation Folds 10 (CV-10) were used to evaluate the classifiers.

5 Experiment Results

We constructed three variants of classifiers in order to thoroughly test the data on football matches.

5.1 Standard Data Set Classification

To make the data collected from the PinnacleSports and Betfair sites useful for data mining purposes, they had to go through pre-treatment in the form of transformation and cleaning of the collected information. The overall objective was to minimize so-called GIGO (garbage in - garbage out) - the reduction of "garbage" that enters the model so that the model could minimize the number of incorrect results [8]. For this purpose, the study included only those events that had odds in the full 10-hour sampling period and had not been postponed. An equal number of matches for each decision-making class was included in order to offset the number of instances from each class [9]. Thus a total of 1116 sample football games were selected, including: 372 matches that ended with a win for the home team; 372 matches that ended with a win for the away team; 372 matches that ended with a draw.

Six classification algorithms were selected: BayesNet, SMO, LWL, Ensemble-Selection, DecisionTable and SimpleCart [7]. For attribute selection the following attribute evaluators and search methods were used: CfsSubsetEval with Best-First, CfsSubsetEval with LinearForwardSelection and PrincipalComponents with Ranker. The highest accuracy rate of 46.51% was achieved by the DecisionTable algorithm. The confusion matrix for the created model is presented in Table 1.

Table 1. Confusion matrix of classifier for a win for the home team, the away team or a draw

a	b	c	← classified as
260	65	47	a = Win-home
154	154	64	b = Win-away
173	94	105	c = Win-draw

Matches that ended with a win for home team (Win-home class) are classified very well in comparison with the two other classes. Most of the matches which ended with a win for the away team were classified slightly worse. In this case

a big mistake occurred due to a mistaken classification as a win for the home team. The worst is the classification of matches that ended in a draw, which are mostly classified incorrectly as a win for the home or the away team. This is because a draw is a middle class between the two results.

5.2 Binary Classification

For better detection of the match result, we decided to build binary classifiers [10] for each type of match result: win for the home team (Win-home), win for the away team (Win-away), and a draw (Win-draw). The binary classifier focuses on one problem and it can perform a better classification than a classifier that has to identify three classes.

Binary Classifier for a Win for the Home Team. When developing a classifier for the home team win, just as before (section 4.1), we used 1116 sample football matches. Matches which ended with a win for the home team remained unchanged, but the matches that ended with a win for the visiting team and a draw were combined to form a new class. Then, we randomly discarded 372 matches to make the number of the instances in each class equal. Below is the size of the two classes: 372 matches that ended with a win for the home team (Win-home class); 372 matches that ended with a win for the away team or a draw (Win-no-home class).

Six classification algorithms were selected: BayesNet, SMO, LWL, Bagging, DecisionTable, and LadTree. For attribute selection the following attribute evaluators and search methods were used: CfsSubsetEval with BestFirst, ConsistencySubsetEval with GreedyStepwise, WrapperSubsetEval (classifier: Bagging) with BestFirst. The highest accuracy rate of 70.56% was noted by the Bagging algorithm, which obtained this result after feature selection (WrapperSubsetEval with BestFirst) and after discretization of attributes. The confusion matrix for the created model is presented in Table 2.

Table 2. Confusion matrix of binary classifier for a win for the home team

a	b	← classified as
229	143	a = Win-home
76	296	b = Win-no-home

Binary Classifier for a Win for the Away Team. The accuracy of predicting a win for the away team proved to be a bit more difficult than predicting a win for the home team. The classifiers achieved worse results, but as in previous studies, a positive influence of feature selection and data discretization was observed. The highest accuracy rate of 65.46% was noted by the Bayesian NaiveBayes algorithm. The confusion matrix for the created model is presented in Table 3.

Table 3. Confusion matrix of binary classifier for a win for the away team

a	b	← classified as
244	128	a = Win-no-away
129	243	b = Win-away

Binary Classifier for a Draw. Same as with the evaluation of classifiers of the standard data set (section 4.1), draws proved to be very difficult to predict. In many cases, the classifiers could not perform a correct classification, which resulted in obtaining accuracies which were not satisfactory. It can be concluded that the values of features describing matches that ended in a draw are very similar to those relating to the win of the home and away teams. The Ensemble-Selection classifier proved to be the most accurate which after feature selection (without discretization) achieved an accuracy of 56.99%. The confusion matrix for the created model is presented in Table 4.

Table 4. Confusion matrix of binary classifier for a draw

a	b	← classified as
196	176	a = Win-no-draw
144	228	b = Win-draw

5.3   Classification of Data Without Draws

Due to the fact that predicting a draw is difficult, we decided to perform additional tests on data that do not contain instances of matches ending in a draw. This allowed creating a classifier that could enable predicting a win for the home or the away team. This information can be used to place *Asian handicap* bets, where in the case of a draw the betting amount is returned.

Matches that ended in a draw were discarded from the 1116 football matches sample set. Matches that ended with a win for the home or the away team were left unchanged. Below is the size of the two classes: 372 matches that ended with a win for the home team (Win-home class); 372 matches that ended with a win for the away team (Win-away class).

Six classification algorithms were selected: BayesNet, VotedPerception, Ibk, Bagging, DecisionTable, and LADTree. For attribute selection the following attribute evaluators and search methods were used: CfsSubsetEval with BestFirst, ConsistencySubsetEval with BestFirst, WrapperSubsetEval (classifier: Naive-Bayes) with BestFirst.

Removal of matches that ended in a draw from the sample data set proved to be very beneficial. Classifiers predicting a win for a home or away team obtained the highest accuracy taking all the conducted studies into account. The classifier that proved to be the most accurate was an algorithm based on the Bayesian

network: BayesNet, which after feature selection conducted after discretization achieved an accurancy of 70.30%. The confusion matrix for the created model is presented in Table 5. The best BayesNet algorithm correctly classified more than 80% of Win-home and 60% of Win-away class matches.

Table 5. Confusion matrix of classifier for win for the home or the away team

a	b	← classified as
298	74	a = Win-home
147	225	b = Win-away

5.4   Summary of Classification of 1-X-2 Type Bets

The evaluation performed on the classifiers built for 1-X-2 type bets showed that a draw is the most difficult to predict. This study confirms the reality of football, because the draw class determines the intermediate odd between a win for the home and the away team. Tests showed that features describing a draw contain many similarities to those relating to a win for the home or the away team. Matrices of classification errors in the study of the standard data set show that most matches which ended in a draw are incorrectly classified as a win for the home team. This is due to the fact that in most cases, the home team is the favorite (has the lowest odd).

In the case of binary classifiers, the accuracy of predicting a win for the home team and the away team is promising. The classifier of a win for the home team achieved an accuracy of 70.56%. Once again the classifier of a draw had the worst results. The best independent classifier was the classifier of a win for the home or away team; the accuracy did not deteriorate with matches which ended in a hardly recognizable draw. The achieved accuracy of this classifier is very satisfying. This classifier can be used for *Asian handicap* bets, where in the case of a draw the betting amount is returned.

In most cases, feature selection resulted in increasing the accuracy of classification. We observed that the features were selected from all the sampling intervals. A selection frequently used features concerning the *minimum* and *maximum values*, *angles to these values*, *derivatives*, the *differences* between the first and last samples in the interval, and the *largest drops* in the value of odds between adjacent samples. This indicates that these features were most important. *Discretization* in most cases also had a very positive influence on the results of classification. Below are the best classification algorithms that have been selected to predict the final results of new football matches. A summary of accuracy of the developed classifiers is presented in Table 6.

Table 6. Classifying algorithms selected for predicting 1-X-2 type bets

Type of classifier	Algorithm	Accuracy
Standard data set	DecisionTable	46.51%
Win for home team	Bagging	70.56%
Win for away team	NaiveBayes	65.46%
Draw	EnsembleSelection	56.99%
Win for home and away team	BayesNet	70.30%

6   Conclusions

The results obtained, an effectiveness of 70%, are quite satisfactory and prove the existence of a relationship between changes in the bookmaker odds values and the outcome of the football match. These results confirm that the knowledge of a group of people about football matches gathered in the form of bookmaker odds can be successfully used for predicting the final result. Based on our research results, one could build a decision-making system that could allow predicting final results based on observation of fluctuations of odds. In further work on the system, new features describing changes of the odds should be investigated, which would probably contribute to improving the accuracy of the system.

References

1. Zdravevski, E., Kulakov, A.: System for Prediction of the Winner in a Sports Game. ICT Innovations 2009, Part 2, pp. 55–63 (2010)
2. Miljkovic, D., Gajic, L., Kovacevic, A., Konjovic, Z.: The use of data mining for basketball matches outcomes prediction. 8th International Symposium on Intelligent Systems and Informatics, pp. 309-312 (2010)
3. McCabe, A., Trevathan, J.: Artificial Intelligence in Sports Prediction. Proceedings of the Fifth International Conference on Information Technology: New Generations, IEEE Computer Society, pp. 1194-1197 (2008)
4. Smith, L., Lipscomb, B., Simkins, A.: Data Mining in Sports: Predicting Cy Young Award Winners. Journal of Computing Sciences in Colleges, vol. 22, issue 4, Consortium for Computing Sciences in Colleges, pp. 115-121 (2007)
5. Pinnacle Sports, `http://www.pinnaclesports.com`
6. Betfair, `http://www.betfair.com`
7. Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I.H.: The WEKA Data Mining Software: An Update. SIGKDD Explorations, Volume 11, Issue 1 (2009)
8. Larose, D.T.: Discovering Knowledge in Data: An Introduction to Data Mining. Wiley Interscience, pp. 28 (2005)
9. Hand, D., Mannila, H., Smyth, P.: Principles of Data Mining. MIT Press (2001)
10. Witten, I.H., Frank, E.: Data Mining: Practical machine learning tools and techniques. Morgan Kaufmann, San Francisco, CA, USA (2005)