**UDC 004.67**

**Olena G. Zhdanova**[1], Candidate of Technical Sciences, Associate Professor of the Department Computer-aided Management and Data Processing Systems, E-mail: zhdanova.elena@hotmail.com, ORCID: 0000-0002-8787-846X
**Bohdan V. Romanchenko**[1], student of the Department Computer-aided Management and Data Processing Systems, E-mail: bohdan.romanchenko@gmail.com, ORCID: 0000-0002-9454-4978
**Maiia O. Sperkach**[1], Candidate of Technical Sciences, Associate Professor of the Department Computer-aided Management and Data Processing Systems, E-mail: sperkachmaya@gmail.com, ORCID: 0000-0003-0251-2141
[1]National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", 37, prosp. Peremohy, Kyiv, Ukraine, 03056

# PREDICTING OF SPORTS EVENTS RESULTS

***Abstract.*** *Today, the field of betting and bookmaking is popular with a wide range of sports fans. Issues of predicting the outcome of future events are and will be relevant for everyday life, sports, politics, etc. With the increasing number and quality of methods of intellectual analysis, the idea of predicting the results of sporting events became feasible. Applying different mathematical methods helps to obtain more accurate predictions of results than subjective expert estimates. The paper introduces the concept of betting and describes in general terms the task of bookmaking. The purpose of the study and the tasks that must be accomplished to achieve the goal are identified. Existing research results of different scientists who have researched this problem are analyzed. There are four basic principles for predicting the outcome of sports events. Different approaches to the task have been considered and our own way of solving it has been proposed. Methods such as Poisson distribution, simulation modeling of the Markov Monte Carlo chain, and many other research methods have been considered. The formulation of the problem is formulated and the properties of the problem are investigated. A backtesting algorithm was developed and described as a mechanism for presenting team statistics at any point in time for a particular season to collect sports event data. Correlation analysis for the selected parameters was shown to show a moderate correlation of data and the use of Google AutoML to identify patterns between the data was described. The importance of using machine learning to solve this problem is outlined. A system has been developed that collects event data and calculates statistics for each team at each point of time using the backtesting algorithm. A service has been developed to create and test the quality of the strategy. The results of experimental studies of task efficiency are presented, where we conducted experimental sets of strategies with and without adding the result of the AutoML service and for each strategy the Pearson correlation coefficient was calculated based on the results of two past seasons. The results obtained are analyzed.*

***Keywords:*** *betting, bookmaking; backtesting; data processing; filtration; structured approach; correlation analysis*

**Introduction.** Today, the concepts of betting and bookmaking are becoming increasingly popular. Using a variety of methods, people try to predict the outcome of future events. To do this, you need to be able to analyze pre-match statistics.

The word "betting" [1] is derived from the English word "bet", which translates to bet. Thus, betting is betting with the ability to win or lose money. Betting is always two sides that make a bet. The first party offers the bet and the second accepts it. Often the first side is the player and the second is the bookmaker. But sometimes the bet may not be taken by the bookie, but by the same player. This happens on the betting exchange. Both bookmakers and exchanges earn their daily betting, as they have a certain "margin" for each event. Margin is a certain numerical value that adds to the odds calculated by a bookie. Each factor contains a margin. Bookmakers put odds on the basis of sports analytics, but on the stock market it may be more interesting, because there the players themselves choose what odds to bet and take bets on.

Bookmaker is a company that bets on sports and other events based on the likelihood of their occurrence [2]. Other events include various cultural events, movie or TV series finals, as well as presidential and political elections, and more.

The bookmaker's job is to make a profit after all bets have been placed. Sometimes the offices try to simply not get the losses or cut them to a minimum, but more often than not they are left with a profit. Bookmakers make a profit by thoroughly analyzing each betting event, collecting statistics, determining the strength of the parties and other factors that can influence the outcome of the event. The bookmaker then sets the odds according to the analysis so that any event result will generate a profit. This is achieved by adding the margin described above to the probability of each result.

In order to organize the sporting events, it was decided to collect historical data from twenty leagues since 2010. A backtesting function was developed to expand information on each event. The result of the backtesting function is a characteristic

© Zhdanova, O., G., Romanchenko, B. V., Sperkach, M. O.. 2019
This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/deed.uk)

of sports strategy. This result characterizes what the results of the strategy have been in the past, and with the help of correlating analysis on backtesting of different seasons, we will be able to predict how the results of the strategy will move in the future.

Data mining techniques are used to study the processes of the wide spread nature. With their help you can solve various problems of data analysis – forecasting, classification, clustering and more.

In order to make a prediction for a particular event, it is necessary to take into account the influence of many factors in the system and external factors. Also, certain parameters can affect not only the output but also other parameters. The unpredictability and non-obviousness of the links on the system and with each other is a major prediction problem. With the development of neural networks, it became possible to take into account all these parameters and their impact on the result, which also increased the accuracy of predicting the results of certain events.

Today, machine learning and neural networks have begun to be used in many areas of life, due to their accuracy, convenience and flexibility in adjusting the automatic adjustment for specific tasks.

**The purpose** of the study is to simplify the process of choosing the events to bet on, to improve the balance statistics of the betting people.

**To achieve this purpose it is necessary to perform the following tasks:**

– review the known results of solving the task;
– develop an event data collection process;
– integrate with the service to calculate predictions for events;
– to develop a backtesting algorithm;
– to develop software implementation of the backtesting algorithm;
– analyze the results.

**Analysis of existing research results**

The directions of world research on the above topic can be divided into the following:

1) the type of data used;

2) prediction phase (that is prediction is made during the game or before the game or even the season);

3) the result to be predicted;

4) applied technology.

*The type of data used.* Most of the data used in sports is structured data. Most studies use structured game or player data [3], or structured odds data based on past odds [4]. However, there are also studies that use unstructured data, such as sentiment analysis in tweets [5] or Tumblr posts [6], but they are not effective in matches with low audience interest.

*Prediction phase.* Studies also vary in the level of "window" used for prediction. Although some studies focus on the study of direct data, such as player trajectories, and the evaluation of player performances [7; 8], others use the data of the first half of the game to predict second-half totals. This is because the bid window is still open during halftime.

*The result to be predicted.* Different studies attempt to predict different types of outcomes, such as the number of goals scored [9; 10; 11], the outcome of the game directly from a "win-win-draw" perspective [12], the measurement of the efficiency or inefficiency of the betting market [13], the behavior of football players (tipsters). The research findings are quite narrow-minded, while in this article we will look at a universal approach for each type of outcome for forecasting.

*Applied technology.* Different studies have used different methods of predicting results. While most previous studies have used methods of statistics and different distributions to predict results, such as Poisson distribution [14], Markov chain iterative modeling by Monte Carlo method [15] discrete-choice regression models for win-loss scenarios [10; 11; 13 ]. Recent studies tend to use data sharing methods such as Naive Bayes [3; 16], Bayesian belief networks [17], reference vector methods, neural and genetic optimization, or combinations of different machine learning algorithms [18]. However, as the study shows, these methods are not very effective, since the prediction accuracy with their use varies from 39 % to 59 %.

The issue of predicting sports events is a constant topic of many research works [19; 20]. In most works, attempts are made to predict the absolute magnitudes of athletic achievement by extrapolating the time series of athletic achievement, depending on the load parameters during training and competition at some future time [20].

Another area of research is related to the study of cyclical changes in results over time and their chronobiological changes over different lengths of time. In these studies, the lengths of time are usually compared with the time stages of many years of preparation.

Another promising direction of prognosis in changes in sports results is the study of relatively short intervals of time, compared with the time of competition [21].

These studies are especially relevant for sports related to the display of accuracy and the performance of the same monotonous actions over many hours: shooting, golf, darts, bowling, etc.

In [22], the ways of taking into account the correlation coefficient in determining the probability

of occurrence (absence) of certain events are presented. Considering foreign sources devoted to the issues of taking into account correlation dependencies in the definition of "weak links" of systems, we can distinguish the works [23-25]. However, the use of correlation analysis, in betting itself, to investigate correlation relationships between random phenomena still remains unresolved.

Models and computer programs for predicting the results of sports games have been developed for many years. Most of them use stochastic methods of describing uncertainty: regression and autoregressive analysis, Bayesian method in combination with Markov chains and their Monte Carlo simulation. Features of such models are: quite high complexity, a large number of assumptions, the need for a large array of statistics. Moreover, these models are not always easy to interpret. There are also models that use neural networks to predict the outcome of a football match [26].

There are a number of methods that researchers use to produce the expected result. [27] uses artificial intelligence methods to create match predictions. For the most part, this problem is referred to as a classification problem, where one class should be provided for the classifier in the classification task (win, loss or draw). In [28], the possibility of using numerical prediction is considered, where it is possible to predict the payoff segment – a numerical value. However, our solution proposes to follow strategies that have been tested before in past research. In non-team sports, researchers use machine learning models to predict player performance. For example, [29] compares neural networks and nonlinear regression to predict the length of the javelin. The purpose of the study was to determine the ability to effectively use neural network models as a means of screening athletes and to compare the results of their work with those obtained from the commonly used regression model.

In this paper, we will look at the problem of systemizing information about games and seasons and finding relationships between them.

**Formulation of the problem in general form**

There is an array of sporting events. Each event has its own parameters, such as time and odds for each party winning. At each time point, the team has its own parameters, such as the number of matches played in a draw, lost or won in succession, which are required for certain strategies.

It is necessary to develop an algorithm for gathering event data depending on the strategy and how to check the reliability of the strategy. The strategy implies a set of factors by which events are filtered. These factors can be both the odds of winning a particular bookmaker side and the average number of points scored per game, the average or maximum draws and more.

The result of the strategy is a "pick", or a selected event and market that the strategy proposes to bet on.

The task is to test the strategy created by analyzing past sporting events. As a result, we must obtain the value of the correlation between the seasons we are back-testing, the required minimum amount of investment and the profit or loss provided that we adhere to this strategy from some point in the past. For the sake of accuracy, we will only use real sports events. And since each strategy filters sports events differently, the data for each strategy will be different.

**Investigation of the properties of the problem**

$E_p$ – a sport event that took place in the past. Indexes $h$ i $a$ will be used to indicate whether the team is playing at home or away.

Each sport event has the following options:

$t$ – time of conduct;

$s_h$, $s_a$ – the number of points earned by teams playing at home and away respectively;

$l$ – a league that both teams belong to;

$T_h$ – team playing at home;

$T_a$ – the away team;

$O_h$ – the odds of winning a home team;

$O_a$ – the odds of the away team playing;

$O_d$ – the odds that the teams draw;

$st$ – the amount we put on a certain odds.

In this case, to find out the profit $P$ when betting on a team playing at home, you must use the formula:

$$P = st \cdot O_h.$$

To fine-tune your strategy, you can specify additional options for teams. They may include average points (goals) per game $T_{ppg}$, current, average and maximum series of no draws, wins or losses. We introduce the notation $avg$ to indicate the middle series, for the maximum, $no_draw$ for no draws, $no_win$ for not winning and $no_lose$ for not losing. Then the current version of no-draws for the team will be indicated as follows: $T_{no_draw}$. Number of games in the series denoted $qty$.

The stats of the team after the game to be played is calculated as follows:

$$T_{no_draw_avg} = \frac{T_{no_draw_avg} \cdot T_{no_draw_qty} + T_{no_draw}}{T_{no_draw_qty} + 1},$$


ISSN 2663-0176 (Print)
ISSN 2663-7731 (Online)

$T_{no_draw_avg}$ – the average number of no-draws for the team;

$T_{no_draw_qty}$ – the number of games played do not draw.

The formula can be used for any series. Thus, you can configure a strategy, such as "winning time", the essence of which is that if the team has not won a number of games in a row, the chance of winning increases.

**The algorithm.** We describe the algorithm for the search for backtesting. The idea behind the algorithm is to calculate at each step statistics for the team and the amount of profit or loss, starting from a certain point in the past [31].

Here is a pseudocode of the backtesting algorithm:

```
Вхід: E* // array of sporting events
      // each event carries information
      // about the odds
      F // the factors exemplified
      // above
      l* // leagues
      T* // teams
Вихід: b // bet on events
       // according to strategy
       avg_odds // avarage
       // odds
       roi // return of
       // investment
// Initialization
       b := []
       avg_odds = 0
// Events filtering
for i:=0 to E.length do
  if T* not includes E_h^i or E_a^i
    delete
  endif
  if l* not includes E_l^i
    delete
  endif
end
for i:=0 to E.length do
  // for each strategy, the bidding
  // algorithm performs
  // differently
  // b_p – what the bet is on
  // (win / lose teams)
  b := PlaceBet(l*, T*, F, E^i)
  result := Process(E_h^i, E_a^i, b_p)
  if result = 'win'
    outcome := outcome + O_[b_p] · st
  else
    outcome := outcome − O_[b_p] · st
  endif
end
e_qty := b.length
for i:=0 to b.length do
 if b.result = 'win'
  odds_out := odds_out + b.odd − 1
 else
  odds_out := odds_out − 1
 endif
end
roi := (odds_out/b.length) · 100
for i:=0 to b.length do
  total_odds := total_odds + b.odd
end
avg_odds := total_odds/b.length
```

*PlaceBet* – a function that checks whether team statistics at a point in time are consistent with a strategy.

Team statistics include: how many games were played in total and away and at home, how many points were scored for each game in general, as well as playing away and at home, average goals scored, and away and away, the number of consecutive games played at home and away, in which at least one point was earned for each team, the average number of consecutive games played, at home and at least one point for each team, the maximum number of matches in a row in general, I play and on the road and at home, which had earned at least one point for each of the teams, the number of runs for more than one game in a row all playing on the road and at home, which had earned at least one point for each of the teams. Also, the statistics of the team are the following parameters: the number of consecutive games in general, playing on the field and at home, which did not earn at least one point on any team, the average number of matches in a row, playing on the field and at home, which did not earn at least would be one point to any team, the maximum number of consecutive games played at home and away, which did not earn at least one point to any team, the number of series more than one game at a time, playing at home and away, who didn't make money, though would be one point to any of the teams, the number of consecutive games in general, playing away and away in which the team won and did not win, and also played in the draw, the average number of consecutive games in total, playing away and at home in which the team won won and played a draw, the maximum number of consecutive matches played at home and away, in which the team won and did not win, and also played a draw, the number of series more than one game in a row,

playing away and at home, in which the team retired Ala and not won, and played in a draw.

The strategy has factors that the team and event must satisfy. The factors correspond to team statistics and event parameters such as odds, leagues and teams. So, the factor of the strategy may be, for example: the event must be within a certain league and that the team playing on the field at least 2 consecutive times does not lose and the coefficient for its victory should be at least 2.

*Process* – a function that returns the result of a bet on an event: win or lose.

As a result of this algorithm, we get a filtered list of past events that characterizes the strategy, showing how its performance has moved in the past, the percentage of return of investment that is a characteristic of the strategy. An important result is the number of events that were found, because the larger the number of events, then more we can trust the data. This result can be obtained when factors too narrow the event search area. If backtesting has returned few events, then we cannot trust its results as much as backstaging has returned many events.

**Experimental studies of the effectiveness of the proposed solution**

Let's run experimental strategies with and without adding AutoML results. We present the input data for the seasons 2017-2018, 2018-2019 and the current season 2019-2020 years of European league football matches in Table 1. We will make experimental choices of strategies and describe their results in Table 2.

Table 1. Input data for experiments

| Characteristic | min | max | avg |
|---|---|---|---|
| Number of events | 6231 | 7523 | 6930 |
| Number of seasons | 3 | 3 | 3 |
| Number of leagues | 10 | 14 | 13 |
| Number of teams | 200 | 280 | 260 |

Table 2. The results of the experiments

| Indicator | With probability values | | | Without probability values | | |
|---|---|---|---|---|---|---|
| | min | avg | max | min | avg | max |
| Pearson coefficient | 0.7 | 0.78 | 0.85 | 0.1 | 0.23 | 0.6 |
| Number of events in strategy | 2532 | 3361 | 3723 | 3420 | 3782 | 4521 |
| Avaradge coefficient | 1.08 | 1.5 | 2.5 | 1.17 | 2 | 3 |
| Estimated percentage of return of investment | 4 % | 12 % | 20 % | -20 % | -5 % | 5 % |

Table 2 shows the results of the experimental strategy selection. There were 2 types of strategies considered when choosing strategies: with and without AutoML predictions. This takes into account the predictions of filtering the event by an additional parameter: what advantage does AutoML give to each party. As we can see, the estimated return of investment is better in strategies that use the prediction of filtering events.

Fig. 1 shows a graph of the backtesting result for the experimental strategy without filtering the events by probability from Google AutoML. Table 3 shows the main characteristics of the strategy. The balance is calculated on the condition that $100 was set for each selected event.

In Fig. 2 shows a graph of the backtesting result for an experimental strategy without filtering events by probability from Google AutoML. Table 4 shows the main characteristics of the strategy.


ISSN 2663-0176 (Print)
ISSN 2663-7731 (Online)

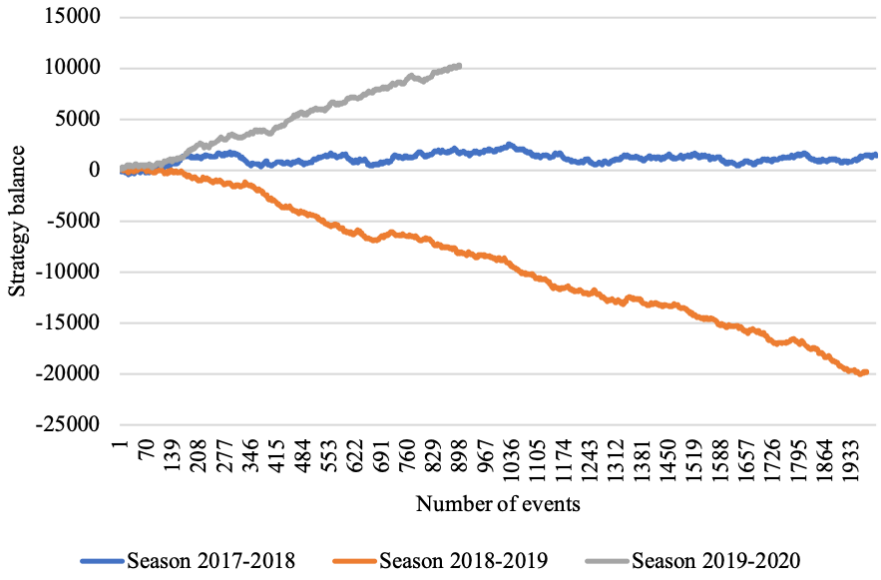



Fig. 1. Graph of a backtesting result for an experimental strategy without using Google AutoML

Table 3. Characteristics of one of the experimental strategies

| | **Number of events** | **Percentage of return of investment** | **Balance** | **Avaradge coefficient** |
|---|---|---|---|---|
| 2017-2018 | 2001 | 1.3 | 2611 | 2.12 |
| 2018-2019 | 1976 | -11 | -22647 | 2.15 |
| 2019-2020 | 900 | 12 | 10598 | 2.2 |

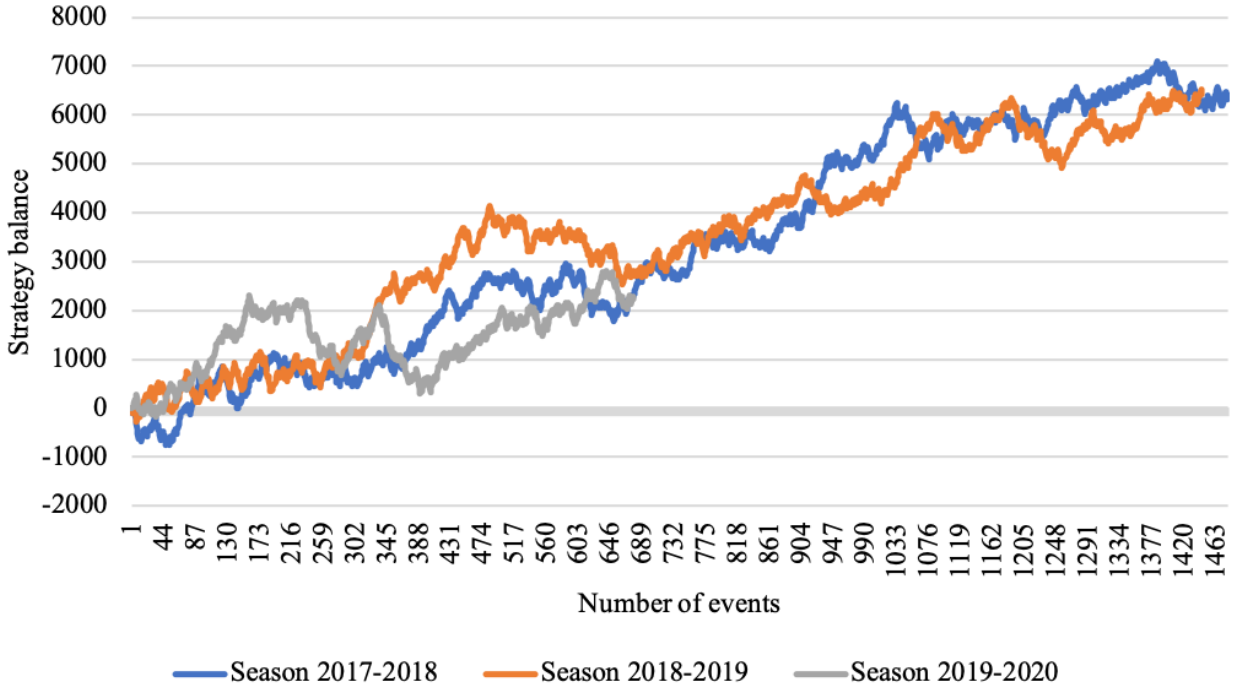



Fig. 2. Graph of a backtesting result for an experimental strategy with using Google AutoML

Table 4. Characteristics of one of the experimental strategies

| Season | Number of events | Percentage of return of investment | Balance | Avaradge coefficient |
|---|---|---|---|---|
| 2017-2018 | 1481 | 5.2 | 6823 | 1.7 |
| 2018-2019 | 1446 | 4 | 6725 | 1.5 |
| 2019-2020 | 679 | 5.5 | 3712 | 1.7 |

**Discussion of results**

If we calculate the Pearson correlation coefficient for strategy balance data without using AutoML, we can conclude that the linear relationship between the data exists, but it is not stable, since its indicator between seasons 2017-2018 and 2018-2019 is 0.48, between seasons 2018-2019 and 2019-2020 is 0.34, and between seasons 2017-2018 and 2019-2020 is -0.629. By calculating the Pearson correlation coefficients for strategy balance data using the AutoML service, we can conclude that the linear relationship between them is and is moderate, since correlation between seasons 2017-2018 and 2018-2019 is 0.73, between seasons 2018-2019 and 2019-2020 coefficient is 0.85 and the correlation coefficient is 0.92 between 2017-2018 and 2019-2020 seasons. According to experimental research, we can conclude that adding Google AutoML significantly improved the correlation coefficient and that the estimated percentage of return of investment was consistently positive and increased by an average of 12 %.

**Conclusion**

The article discusses an approach to simplifying the selection of events to bet on and a development approach that can improve the balance statistics of betters. The following tasks were performed: known results of the task were solved, event data collection process was developed, integration with the Google AutoML event prediction service was implemented, pseudocode of the backtesting algorithm was developed and presented. The results obtained confirm the effectiveness of the proposed solution.

**References**

1. “What is betting?” [Electronic resource] – Access mode: https://marketbusinessnews.com/financial-glossary/betting/ – Active link – 15.12.2019.

2. Norman Dreўper & Harry Smyt. (2007) “Prykladnoj rehressyonnyj analyz. Mnozhestvennaia rehressyia” [Applied Regression Analysis. Multiple Regressions], 3-e yzd, Moscow: Russian Federation, *Publ. Dyalektyka*, 912 p.

3. Baboota, R. & Kaur, H. (2018) “Predictive Analysis and Modeling Football Results using Machine Learning Approach for English Premier League”, *International Journal of Forecasting, Forthcoming*, doi: 10.1016/j.ijforecast.2018.01.003.

4. Dobrovec, S. (2015) “Predicting Sports Results using Latent Features: A case Study”, *38th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), 25-29 May 2015*, pp. 1267 – 1272, doi: 10.1109/MIPRO.2015.7160470.

5. Schumaker, R. P., Jarmoszko, A. T. & Labedz, C. S. (2016) “Predicting wins and Spread in the Premier League using a Sentiment Analysis of Twitter”, *Decision Support Systems,* Vol 88, pp. 76-84, doi: 10.1016/j.dss.2016.05.010.

6. Radosavljevic, V. (2014) “Large-scale World Cup 2014 Outcome Prediction Based on Tumblr posts”, *KDD Workshop on Large-Scale Sports Analytics*, Sydney: Australia.

7. DOrazio, T., Guaragnella, C., Leo, M. & Distante, A. (2004) “A new Algorithm for ball Recognition using Circle Hough Transform and Neural Classifier”, *Pattern Recognition*, Vol 37, pp. 393-408, doi: 10.1016/S0031-3203(03)00228-0.

8. Kang, Y. L., Lim, J. H., Kankanhalli, M. S., Xu, C. S. & Tian, Q. (2004) “Goal Detection in Soccer Video using audio/visual Keywords”, *IEEE International Conference on Image Processing (ICIP)*, Singapore, pp. 1629-1632, doi: 10.1109/ICIP.2004.1421381.

9. Maher, M. J. (1982). “Modeling Association Football Scores”, *Statistica Neerlandica*, Vol 36, pp. 109-118.

10. Gianluca, B. & Blangiardo, M. (2010). “Bayesian Hierarchical Model for the Prediction of Football Results”, *Journal of Applied Statistics*, Vol 37, Issue 2, pp. 253-264, doi: 10.1080/02664760802684177.

11. Karlis, D. & Ntzoufras, I. (2008) “Bayesian Modelling of Football Outcomes: Using the Skellams Distribution for the Goal Difference”, *IMA Journal of Management Mathematics*, pp. 229-244, doi: 10.1093/imaman/dpn026.

12. Goddard, J. (2005) “Regression Models for Forecasting Goals and Match Results in Association

Football", *International Journal of Forecasting*, Vol 21, Issue 2, pp. 331-340, doi: 10.1016/j.ijforecast.2004.08.002.

13. Dixon, M. J. & Coles, S. G. (1997). "Modelling Association Football Scores and Inefficiencies in the Football Betting Market", *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, Vol 46, pp. 265-280, doi: 10.1111/1467-9876.00065.

14. Forrest, D. & Simmons, R. (2000) "Forecasting Sport: the Behaviour and Performance of Football Tipsters", *International Journal of Forecasting*, Vol 16, Issue 3, pp. 317-331, doi: 10.1016/S0169-2070(00)00050-9.

15. Koopman, S. J. & Lit, R. (2013) "A Dynamic Bivariate Poisson Model for Analysing and Forecasting Match Results in the English Premier League", *Journal of the Royal Statisitical Society:* Series A, Vol 178, Issue 1, pp. 167-186, doi: 10.1111/rssa.12042.

16. Crowder, M., Dixon, M., Ledford, A. & Robinson, M. (2002) "Dynamic Modelling and Prediction of English Football League Matches for Betting", The Statistician, Vol. 51, pp.157-168, doi: 10.1111/1467-9884.00308.

17. Godin, F. (2014). "Beating the Bookmakers: Leveraging Statistics and Twitter Microposts for Predicting Soccer Results", *KDD Workshop on Large-Scale Sports Analytics*, Sydney: Australia, doi: 10.13140/2.1.2168.0000.

18. Byungho, M., Kim, J., Choe, C., Eom, H. & McKay, R. I. (2008). "A Compound Framework for Sports Results Prediction: A Football Case Study", *Knowldege Based Systems*, Vol. 21, pp. 551-562, doi: 10.1016/j.knosys.2008.03.016.

19. Korolkov, A. N. (2014). "Ultradyannыe rytmы rezultatyvnosty v holfe" [Golf's Ultra-performance Rhythms], *Vestnyk sportyvnoi nauky*, Vol. 2, pp. 34-37.

20. Hermanov, H. N. (2010) "Prohnoz dostyzhenyi rossyiskykh behunov na srednye y dlynnie distantsii na chempionatakh Evropi 2012-2014 hh. po rezultatam vistupleniy yuniorov i molodikh sportsmenov v Evropeiskykh Pervenstvakh" [Prognosis of Achievements of Russian Runners in the Middle and Long Distances at the European Championships 2012-2014 as a Result of the Performances of Juniors and young Athletes in the European Championships], *Kultura fyzycheskaia y zdorove*, Vol. 4, pp. 7-11.

21. Korolkov, A. N. (2013). "Effektyvnost trenyrovky v holfe v vyde peredatochnoi funktsyy kvazistatsionarnikh spektrov rezultativnosty" [Golf Training Effectiveness as a Transfer Function of Quasi-stationary Performance Spectra], *Teoryia y praktyka fyzycheskoi kulturi*, Vol. 10, pp. 62-64.

22. Kupalova, H. I. (2008) "Teoria ekonomichnoho analizu" [The theory of Economic Analysis], Kyiv: Ukraine, 639 p.

23. Dan M. Frangopol. (2007) "Reliability and Optimization of Structural Systems: Assessment, Design, and life-cycle Performance: Proceedings of the Thirteenth IFIP WG 7.5", *Working Conference on Reliability and Optimization of Structural Systems*, Kobe: Japan, 269p.

24. Ditlevsen, O. & Madsen, H. O. (2003) "Structural Reliability Methods". *Department of mechanical engineering. Technical University of Denmark maritime engineering*, Denmark, 323 p.

25. Hitoshi Furuta. (2002) "Reliability and Optimization of Structural Systems: Proceedings of the 10th IFIP WG7.5", *Working Conference on Reliability and optimization of structural systems*, Osaka: Japan, 276 p.

26. "Metod hruppoho uchёta arhumentov" [The Method of Group Accounting of Arguments] [Electronic resource] – Access mode: http://www.machinelearning.ru/wiki/metod hruppoho uchёta arhumentov/ – Active link – 15.12.2019.

27. "Dokumentatsiia do servisu AutoML vid Google" [Google AutoML Documentation] [Electronic resource] – Access mode: https://cloud.google.com/automl/ – Active link – 15.12.2019.

28. "Matematychne prohnozuvannia rezultativ futbolnykh matchiv" [Mathematical Prediction of Football Match Results] [Electronic resource] – Access mode: http://naub.oa.edu.ua/2015/ – Active link – 15.12.2019.

29. "Market Business News" [Electronic resource] – Access mode: https://marketbusinessnews.com/financial-glossary/betting/ – Active link – 15.12.2019.

30. "Gambling Sites" [Electronic resource] – Access mode: https://www.gamblingsites.com/sports-betting/introduction/what-a-bookmaker-does/ – Active link – 15.12.2019.

31. Zhdanova, O. H., Romanchenko, B. V. & Sperkach, M. O. (2019). "Prohnozuvannia rezultativ sportivnykh podiy" [Predicting the Results of Sports Events], *Matematychne ta imitatsiine modeliuvannia system MODS 2019*, *Chotyrnadtsiaty mizhnarodna naukovo-praktychna konferentsiia*, pp. 288-292.

Received 21.10.2019
Received after revision 08.11.2019
Accepted 28.11.2019

**УДК 004.67**

[1]**Жданова, Олена Григорівна,** кандидат технічних наук, доцент кафедри автоматизованих систем обробки інформації та управління, E-mail: zhdanova.elena@hotmail.com,
ORCID: 0000-0002-8787-846X
[1]**Романченко, Богдан Володимирович,** магістрант кафедри автоматизованих систем обробки інформації та управління, E-mail: bohdan.romanchenko@gmail.com, ORCID: 0000-0002-9454-4978
[1]**Сперкач, Майя Олегівна,** кандидат технічних наук, доцент кафедри автоматизованих систем обробки інформації та управління, E-mail: sperkachmaya@gmail.com, ORCID: 0000-0003-0251-2141
[1]Національний технічний університет України «Київський політехнічний інститут ім. Ігоря Сікорського», пр. Перемоги, 37, м. Київ, Україна, 03056

# ПРОГНОЗУВАННЯ РЕЗУЛЬТАТІВ СПОРТИВНИХ ПОДІЙ

***Анотація.*** *У сьогоденні сфера беттингу та букмекерства є популярною у широкого кола прихильників спорту. Питання прогнозування результатів майбутніх подій є і будуть актуальними для повсякденного життя, спорту, політики, тощо. З ростом кількості та якості методів інтелектуального аналізу стала здійсненною ідея прогнозування результатів спортивних подій. Застосування різних математичних методів допомагає отримати більш точні прогнози результатів, аніж суб'єктивні прогнози експертів. У роботі у загальному вигляді описано поняття беттингу та задачу букмекерства. Визначено мету дослідження й задачі, які необхідно виконати для досягнення мети. Проаналізовано існуючі результати досліджень різних науковців, які досліджували дану проблематику. Виділено основні чотири основи для прогнозування результатів спортивних подій. Розглянуто різні підходи до поставленої задачі, у тому числі такі методи, як розподіл Пуассона, імітаційне моделювання методом Монте-Карло та багато інших методів дослідження й було запропоновано власний спосіб її розв'язання. Сформульовано постановку задачі та проведено дослідження властивостей задачі. Для збору даних спортивних подій було розроблено й описано алгоритм бектестінгу, як механізм представлення статистики команди в будь-який момент часу певного сезону. Показано, що кореляційний аналіз по обраним параметрам показував помірну зв'язаність даних та було описано застосування системи Google AutoML для знаходження закономірності між даними. Експериментально доведено важливість використання сервісів з машинного навчання при розв'язанні даної задачі. Розроблено систему, яка збирає дані про події та за допомогою алгоритму бектестінгу обраховує статистику для кожної команди в кожен момент часу. Розроблено сервіс для створення й перевірки на якість стратегії. Наведено результати експериментальних досліджень ефективності задачі, де було проведемо експериментальні підбори стратегій з і без додавання результату роботи сервісу AutoML й для кожної стратегії було обраховано коефіцієнт кореляції Пірсона по результатам двох минулим сезонам. Проведено аналіз отриманих результатів.*

***Ключові слова:*** *беттинг; букмекерство; бектестінг; обробка даних; фільтрація; прогнозування; кореляційний аналіз*

**УДК 004.67**

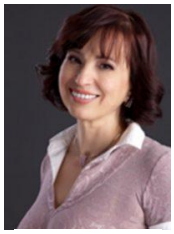
[1]**Жданова, Елена Григорьевна,** кандидат технических наук, доцент кафедры автоматизированных систем обработки информации и управления, E-mail: zhdanova.elena@hotmail.com,
ORCID: 0000-0002-8787-846X
[1]**Романченко, Богдан Владимирович,** магистрант кафедры автоматизированных систем обработки информации и управления, E-mail: bohdan.romanchenko@gmail.com, ORCID: 0000-0002-9454-4978
[1]**Сперкач, Майя Олеговна,** кандидат технических наук, доцент кафедры автоматизированных систем обработки информации и управления, E-mail: sperkachmaya@gmail.com,
ORCID: 0000-0003-0251-2141
[1]Национальный технический университет Украины «Киевский политехнический институт им. Игоря Сикорского », пр. Победы, 37, г. Киев, Украина., 03056

# ПРОГНОЗИРОВАНИЕ ИСХОДОВ СПОРТИВНЫХ СОБЫТИЙ

***Аннотация.*** *В наши дни сфера беттинга и букмекерства является популярной у широкого круга любителей спорта. Вопрос прогнозирования результатов будущих событий есть и будут актуальными для повседневной жизни, спорта, политики и тому подобное. С ростом количества и качества методов интеллектуального анализа, стала осуществимой идея прогнозирования результатов спортивных событий. Применение различных математических методов помогает получить более точные прогнозы результатов, чем субъективные прогнозы экспертов. В работе в общем виде описано понятие беттинга и задача букмекерства. Определены цели исследования и задачи, которые необходимо выполнить для достижения цели. Проанализированы существующие результаты исследований различных ученых, которые исследовали данную проблематику. Выделены главные четыре основы для прогнозирования результатов спортивных событий.*

*Рассмотрены различные подходы к поставленной задаче, в том числе такие методы, как распределение Пуассона, имитационное моделирование методом Монте-Карло и много других методов исследования. Также был предложен собственный способ ее решения. Сформулировано постановку задачи и проведено исследование свойств задачи. Для сбора данных спортивных событий был разработан и описан алгоритм бэктестинга, как механизм представления статистики команды в любой момент времени определенного сезона. Показано, что корреляционный анализ по выбранным параметрам показывал умеренную связанность данных. Было описано применение системы Google AutoML для нахождения закономерности между данными. Экспериментально доказана важность использования сервисов из машинного обучения при решении данной задачи. Разработана система, которая собирает данные о событиях и с помощью алгоритма бэктестинга рассчитывает статистику для каждой команды в каждый момент времени. Разработан сервис для создания и проверки на качество стратегии. Приведены результаты экспериментальных исследований эффективности задачи, где были проведены экспериментальные наборы стратегий-с и без добавления результата работы сервиса AutoML. Для каждой стратегии было подсчитано коэффициент корреляции Пирсона по результатам двух прошлых сезонов. Проведен анализ полученных результатов.*

***Ключевые слова:*** *беттинг; букмекерство; бэктестинга; обработка даннях; фильтрация; прогнозирование; корреляционный анализ*



**Olena G. Zhdanova,** Candidate of Technical Sciences, Associate Professor of the Department computer-aided management and data processing systems
*Scientific interest:* information technology, project management, neural networks, knowledge management



**Bohdan V. Romanchenko,** student of the Department computer-aided management and data processing systems
*Scientific interest*: information technology, neural networks, machine learning, knowledge management



**Maiia O. Sperkach,** Candidate of Technical Sciences, Associate Professor of the Department computer-aided management and data processing systems
*Scientific interest*: information technology, project management, machine learning, knowledge management