UDC 004.852
DOI: 10.15587/1729-4061.2023.276977

FRAMEWORK BASED ON CONFORMAL PREDICTORS AND POWER MARTINGALES FOR DETECTION OF FIXED FOOTBALL MATCHES

Oleg Chertov
Corresponding author
Doctor of Technical Sciences, Professor*
E-mail: chertov@i.ua
Ivan Zhuk
Postgraduate Student*
*Applied Mathematics Department
National Technical University of Ukraine
«Igor Sikorsky Kyiv Polytechnic Institute»
Peremohy ave., 37, Kyiv, Ukraine, 03056

One of the difficult problems that arises during football competitions is match-fixing. In terms of negative effect, such shameful phenomena are commensurate with the problem of doping. This paper has analyzed known methods for the possible detection of match-fixing, including sociological analysis of participants in match-fixing, methods for predicting the outcome of the match, analysis of bets and performance of the player or team during the match. It is noted that the assessment of match-fixed results in the considered methods is carried out based on the analysis of a large amount of data. But such information is not always available. Given the insufficient formalization of the problem area, it is relevant to conduct research that does not require a large amount of non-publicly available data but, at the same time, makes it possible to effectively identify potentially suspicious matches regarding a fixed result. The description of the input data is formalized in the form of a data structure containing a chronological history of the results of football seasons, the ranking of teams and matches of the season depending on the overall result of the teams in the season. A method for detecting suspicious football matches with a fixed result has been built using conformal predictors and power martingales within which a new measure of non-conformity has been introduced to determine atypical football matches. To obtain a generalization of the statistics of atypical matches, a power submartingale was used. Evaluation of the effectiveness of the developed method for detecting suspicious football matches was carried out based on precision and recall of the classification metrics using data on the 2013–2014 season of the French II League. The quality of work of the developed method reaches 85 % in terms of precision metric, 96 % in terms of recall metric, and 0.853 in terms of metric F1

Keywords: football match, fixed result, measure of nonconformity, p-value for conformity, degree of difference

Received date 12.01.2023
Accepted date 23.03.2023
Published date 28.04.2023

How to Cite: *Chertov, O., Zhuk, I. (2023). Framework based on conformal predictors and power martingales for detection of fixed football matches. Eastern-European Journal of Enterprise Technologies, 2 (4 (122)), 6–15. doi: https://doi.org/10.15587/1729-4061.2023.276977*

1. Introduction

Matches with a fixed result (match-fixing), as well as doping, are called cancer of sports [1]. To combat doping, high-tech tests are used, and in 1999 a specialized international organization WADA (the World Anti-Doping Agency) was founded. These tests make it possible to detect with almost absolute accuracy the presence of prohibited drugs in the athlete's body. Unlike the doping situation, the successes in the fight against match-fixing are much more modest. Although matches with a fixed result, perhaps, exist as much as the sports themselves. At least, we have historical information that such shameful acts took place at the ancient Olympic Games in Greece [2].

Suppose that an agreement on a specific final outcome of a match or certain characteristics of it (number of goals scored, goal difference, etc.) is the purpose of obtaining an illegal benefit on bets in betting companies. In this case, it is potentially possible to track through financial transactions or the unnatural distribution of betting volumes on the corresponding match results [3, 4]. The main disadvantage of this approach is that the data on the number, size, and time of the corresponding rates are non-public information, and the known methods of privacy-preserving data mining [5] are of little use here.

Nevertheless, even the availability of data from betting companies will not help in certain situations. This is possible if there is a behind-the-scenes agreement to lose one team to another with some financial or other compensation through independent channels [2]. It is also possible when there is material stimulation by a third party of one team against another. It is almost impossible to establish such facts, at least without conducting a special police investigation. The fact is that the very nature of sports competitions ensures that a stronger team does not always win, and many relevant examples can be given in any sport. Therefore, the related literature [2–4] only states that the outcome of certain matches is suspicious, atypical, illogical, abnormal, etc. But such information is also useful and important since it acts as an additional indicator that a certain team may be related to the violation of the sports principles of fair play.

Currently, almost all known [3, 4] cases of detection of match-fixing are associated with the use of data from bookmakers. However, bets are not accepted for all matches in sports tournaments, and if they are accepted, information about them is a trade secret, or bets can generally be accepted on illegal sports betting platforms. Therefore, studies aimed at finding potentially suspicious results of sports matches based on the processing of exclusively publicly available data should be considered relevant.

2. Literature review and problem statement

The problems of match-fixing research are multidimensional and include the search for the causes of their occurrence and distribution, taking into account national characteristics, economic background, legal issues, connection with the criminal world, analysis of features in various sports, etc. [1, 2]. In the last decade, sociologists have been actively involved in the discussion of these issues [6, 7]. Competitions, where the game is not considered authentic, cannot be expected to continue to receive public support [8, 9]. The network of participants involved in the organization of match-fixing itself can have a different structure: from horizontal between weakly connected temporary participants (as in Koriopolis, when at least 40 matches were fixed during the 2009/2010 Greek football season) [10] to a strict vertical hierarchy (as in the «Şike Davası» scandal during the 2010/2011 Turkish football season), where the legal business of a football club operated under an illegal scheme [11].

A separate category of research in the field of sports competitions, in particular football, are mathematical methods for detecting sports matches with a fixed result. The search for such anomalies in sports competitions, at least in the field of football, is reduced to predicting the outcome of the match, analyzing bets or actions of the participants of the match throughout the game [3, 8, 12–14].

Predicting the outcome of a football match has long been a popular research problem. Methods for obtaining a solution to this problem include statistical methods [12] (in particular, Bayesian networks) and machine learning methods [13]. In [12], the Bayesian network was proposed to predict the outcome of the match in terms of a draw or win of a host or guest team, the accuracy of which reached 75 %, which at the time of publication of the results was one of the highest in solving problems in this way. To predict the result, the model uses 18 attributes for each match. The advantage of this method is the possibility of numerical probabilistic determination of the influence of the selected attributes on the outcome of the match, which makes it possible to formulate probabilistic rules for determining the outcome of the match. The disadvantage of this method is the need to use a significant number of additional attributes about each match to predict the result more accurately.

In [13], a neural network of deep learning is proposed to predict the winner of a football match. To predict the outcome of the current match, information about the last 35 matches of each of the teams participating in the match is used as input. From this information, the values of 9 attributes are considered for each match. For each of the attributes, 6 statistical indicators are formed on the basis of the last 35 matches. Thus, for the team participating in the match, a vector with 54 coordinates is formed. A combined vector with 108 coordinates is an input data set for the selected neural network. The result of the neural network is 3 values from the range [0; 1], which in total give 1 and are interpreted as the probability that the match will end in a draw or win of one of the teams. Using this model, a forecasting accuracy of 62 % was achieved. The disadvantage of the previous method is characteristic of this model as well. Also, a disadvantage of this model is the impossibility of determining the patterns by which it would be analytically possible to determine the outcome of a particular match.

Each of the considered models, subject to the achievement of high accuracy of forecasting, in the future can be used to determine the «anomaly» of the result of a particular match. The disadvantages of these models are the use of a significant number of additional match attributes, which are not always available, and the inability to obtain analytical patterns to predict the result. These shortcomings necessitate research into developing algorithms that use more accessible match attributes and make it possible to build appropriate analytical patterns to predict the result.

To identify matches with a fixed result, methods based on betting analysis are used [8, 15]. In [8], the authors used special models to predict the volume of bets and compare the real volume of bets with the projected using statistical methods. The volumes of these bets were compared at the time the match was going on. The models used for forecasting are based on attributes such as the passage of time, awarding red cards, and current score. If during a match the difference between the actual bet volume and the projected one is statistically significant, the match is considered fixed. In [15], a model with goals and bookings was introduced. This model includes the type of team (host, guest), events about the goal scored, and a yellow or red card. The model is based on Weibull's event process model. These models are based on probabilities and are aimed at predicting the outcome of the match. However, a recent study [16] showed that most cases of match-fixing are not related to bets.

Methods for analyzing the performance of a player or team in a game have been significantly developed [14, 17, 18]. These studies used data from players' positions to evaluate their performance. The analysis of the dominant area diagram using Voronoi diagrams [14, 17], and analysis of the trajectories of players with the possibilities of intra-team interactions were introduced [18]. These methods aim to evaluate the performance of the player and the team based on the player's actions in the game and his positioning. As part of these studies, the trajectories of players during the match are considered. To assess the quality of the player's performance, the trajectories of this player's movement during different matches starting at the same playing position are compared. Based on this comparison, it is possible to make an assumption about the fixedness of the outcome of the match, based on a significant difference in the «work» of the player in this match and other matches.

As follows from the above review, in the articles under consideration the assessment of match-fixing is carried out based on an analysis of a large amount of data: on the time and size of bets [8], events on the field during the game [15], or on the routes of movement of players during the match [14, 17, 18]. Such information is not always available for analysis. And this predetermines our research to identify potentially suspicious matches regarding a fixed outcome, which does not require a large amount of non-publicly available data.

3. The aim and objectives of the study

The purpose of this study is to identify suspicious matches, in terms of the fixed outcomes, in football tournaments using conformal predictors and power martingales based on the use of data only on the results of matches played. This will make it possible to select the specified category of matches simply and quickly from all sports matches held during a round-robin football tournament, that is, when each team plays at least one match with the other.

To achieve the set aim, the following tasks have been solved:

– to formalize the description of input data in the form of a data structure containing a chronological history of the results of the football season, the ranking of teams and matches of the season depending on the overall result of the teams in the season;

– to devise a method for detecting suspicious football matches with a fixed result using conformal predictors and power martingales;

– to investigate and evaluate the effectiveness of the developed method of detecting suspicious football matches based on classification metrics that are basic in machine learning.

4. Materials and research methods

4. 1. The object and hypothesis of the research

The object of our study is the process of identifying football match-fixing. The subject of the study are methods for finding suspicious football matches in terms of fixed outcomes, based on conformal predictors and power martingales.

The main idea of the method is to detect football match-fixing by changing the value of the corresponding power martingale. The value of the martingale for the match is formed on the basis of the results of the conformal predictor for the current and previous matches. The conformal predictor determines the level of how much the current match corresponds to the entire group of matches based on data on the success of the participating teams and the results of both the current match and previous matches of the tournament.

The source of input data is the publicly available relational database BeatTheBookie [19]. The employed formalization of the data description was proposed in [20].

4. 2. Terminology

Conformal predictors are a class of machine learning methods that predict the belonging of an object to a certain class based on the measure of difference (conformity) of the current object from previous objects [21].

Power martingale is a mathematical quantity designed to test the exchangeability hypothesis of random observations [21].

Exchangeability: the probabilistic distribution F of random events $z_1, z_2, ..., z_n \in Z$ has the property of exchangeability, if for any permutation π of events $z_1, z_2,..., z_n$ for probability p we have the following equality: $p(z_1, z_2,..., z_n)=p(\pi(z_1), \pi(z_2),..., \pi(z_n))$ [21].

For example, assume we have the problem of classifying images $Z=\{z_1, z_2, ..., z_n\}$ and we have such a distribution Q on the set Z that any order of appearance of different images has the same chance, that is, for any permutation $\pi(Z)$ the probability is $p(z_1, z_2, ..., z_n)=p(\pi(z_1), \pi(z_2), ..., \pi(z_n))$. Then such a distribution $p(z_1, z_2, ..., z_n)=p(\pi(z_1), \pi(z_2), ..., \pi(z_n))$ Q has the property of exchangeability.

In general, the monograph [21] constructed a mathematical apparatus of conformal predictors and power martingales for solving problems of classification and regression, in particular, for the classification of images and linear regression. The advantages of this mathematical apparatus are to combine the process of learning and forecasting in one stage and, as a result, to use the results of the classification (or regression) of previous objects to predict the result for the current object. Also, this mathematical apparatus makes it possible to look for anomalies in the data set. In particular, [22] shows how this mathematical apparatus can be practically used to improve the results of classification by methods of support vectors and nearest neighbors, and to detect changes in data flows.

4. 3. Selected binary classification metrics

To evaluate the results of the study, the distribution of matches by suspicion in terms of a fixed result is used according to the principle of positive-negative (binary classification), where the positive match is a match for which the characteristic «Potentially suspicious» is 1. Important for further analysis is the following elements of the confusion matrix: true positives (TP), false positives (FP), and false negatives (FN). True positives equal the number of potentially suspicious matches, and the algorithm found them to be such. False positives equal the number of matches that are not potentially suspicious, but the algorithm found them to be so. False negatives equal the number of potentially suspicious matches, but the algorithm mistakenly missed them. These characteristics are used to calculate the metrics of precision (P), recall (R), as well as measure F_1:

$$P=\frac{TP}{TP+FP}, \tag{1}$$

$$R=\frac{TP}{TP+FN}, \tag{2}$$

$$F_1=\frac{2}{1/P+1/R}=\frac{2TP}{2TP+FP+FN}. \tag{3}$$

The selected characteristics are the basic characteristics of the analysis of the effectiveness of algorithms that are used to solve binary classification problems. The characteristic (1) shows how often the algorithm detects as «positive» those objects that should belong to the «positive» class, or how rarely the algorithm refers to the «positive» class those objects that do not belong to it. The characteristic (2) shows how «fully» the algorithm detects objects that must belong to a positive class, that is, how often the algorithm refers to positive objects to a positive class and not objects of another class. The characteristic (3) is a harmonic average of (1) and (2) and is a more balanced assessment of the effectiveness of the classification algorithm, which makes it possible to generalize both the precision and recall of the algorithm. All 3 characteristics take values from the range [0; 1] and have a similar interpretation of the results: the closer the characteristic value to 1, the more efficient the algorithm is in terms of this characteristic.

5. Results of the use of conformal predictors and power martingales to identify football match-fixing

5. 1. Results of ranking teams and formalizing the description of input data

The rank i_k (or j_k) of the k-th team can be determined based on the success of the team (its place in the overall standings, the number of points scored, etc.) in this or previous seasons. Subsequently, the success rate $s(t)$ of team t indicates the number of points scored in the last completed season: $s(t)=3w(t)+d(t)$, where $w(t)$ and $d(t)$ are, respectively, the number of wins and draw games for team t. The results of the formalization of the description of the input data were considered on the example of the 2013–2014 season of the French League II. That season was chosen because there were public statements about the falsity of the results of some matches of two teams; but this information could not be legally proved by the relevant authorities. The teams are arranged in descending order of their final points that season (Fig. 1) and by conducting one-dimensional clustering by the K-means algorithm to divide into clusters of teams, close in terms of the number of points scored.

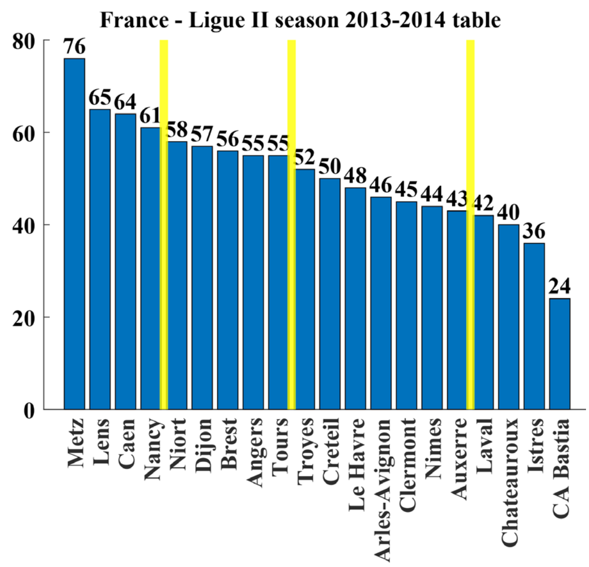

Fig. 1. Success of teams of the 2013–2014 season of the French League II

As a result, 4 clusters of teams were formed for that season (Fig. 1). The composition of the clusters is as follows: teams No. 1–4 belong to cluster No. 1, teams No. 5–9 belong to cluster No. 2, teams No. 10–16 belong to cluster No. 3, teams No. 17–20 belong to cluster No. 4. Thus, the smaller the cluster number, the higher the place in the final table taken by the teams of this cluster. Therefore, it is logical to call the cluster number the rank of the teams included in it. Match classes are built as ordered pairs of home and guest team ranks. Thus, for the 2013–2014 season of the French League II, a set of 16 classes of matches was built $C=\{(i,j)|\ i,j\in\{1, 2, 3, 4\}\}$. This set is the value domain of the $class(k)$ function, which defines the match class with the match number k.

5. 2. Devising a method for detecting football match-fixing

The unit of input for the method is the observation z_k, which describes the match of the football season, k is the serial number of the match in the season. Observations z_k is a set of values $z_k=(i_k, j_k, \alpha_k, \beta_k, T_k)$, where i_k and α_k are, respectively, the rank and result of the host team of this match, and j_k and β_k are the rank and result of the guest team of the match, T_k is the date of the match.

A further set of algorithms for detecting football match-fixing is developed on the basis of a general method for constructing a conformal predictor proposed in [21], by its following adaptation:

– introduction of a new measure of nonconformity (difference), adapted to the peculiarities of rating football matches and its participants;

– using the standard p-value for conformity and power martingale;

– introduction of decision rules according to which, based on the characteristics used, a decision is made about the potential suspicion of a football match.

For each new match z_k:

1) the measure of difference a_k is calculated, which is the first step in calculating the conformal predictor:

$$a_k = 1.5^{1-\operatorname{sgn}\left(avg_{k(i,j)}*(\alpha_k-\beta_k)\right)}\left|(\alpha_k-\beta_k)-avg_{k(i,j)}\right|, \tag{4}$$

where $avg_{k(i,j)} = \underset{class(z_k)=class(z_l)}{\text{mean}}\{\alpha_l-\beta_l\}$, k is the number of the current match in chronological order, l is the number of another match that belongs to the same class as the current match, i, j is, respectively, the number of the host and guest team of the current match. The function $\operatorname{sgn}(x)$ indicates the sign of the number x and is equal to 1 if $x>0$, 0 if $x=0$, and -1 if $x<0$. Characteristics $avg_{k(i,j)}$ is the average result of the observation group used in this study as the expected result of the game in a given class of matches. It is calculated from the entire set of observations z_k, that is the entire set of observations is considered known and available for research in full. Because of this, the algorithm was termed an offline algorithm.

The introduction of such a measure of nonconformity provides an opportunity to compare the actual outcome of the match with the results of all other matches of the class. Also, this measure takes into account both the absolute results of the match teams and the difference in the results of the actual and predicted results, and this difference has a more significant impact than the absolute results of matches. Due to this, for example, matches with an actual victory and a predicted defeat are more non-conforming than matches with an actual and predictable victory, which differ in the number of goals scored.

2) Using the measures of difference between the current k-th match and previous matches of the same class, the degree of difference p_k of the match z_k from the set of observations $\{z_1, z_2,..., z_{k-1}, z_k\}$, which is the result of the conformal predictor, is calculated:

$$p_k = p(z_1, z_2,..., z_k) = \frac{\#\{i : a_i \geq a_k, 1\leq i\leq k\}}{k}, \tag{5}$$

where operation $\#A$ returns the number of elements in set A. For example, for a set of integers $\{1, 2, 5, 10, 15, 17\}$ operation $\#\{1, 2, 5, 10, 15, 17\}=6$. In (5), the numerator contains the numbers of those observations whose measure of difference is the same or greater than that of the current observation, including the number of the current observation. Therefore, the number of elements in the set from the numerator of this formula takes values in the range $[1; k]$.

In (5), the degree of difference p_k is calculated without the use of random variables, so such an algorithm is deterministic.

3) based on the currently available series of degrees of difference $\{p_k\}$, a modified power martingale $M_k^{(\varepsilon)}$ is calculated for an arbitrary value $\varepsilon\in[0; 1]$:

$$M_k^{(\varepsilon)} = \varepsilon\prod_{i=1}^{k} p_i^{\varepsilon-1}. \tag{6}$$

To level the dependence on the value of ε, the integral martingale is calculated:

$$M_k = \int_0^1 M_k^{(\varepsilon)}\mathrm{d}\varepsilon. \tag{7}$$

4) the set of suspicious matches S, that is, matches for which there is a suspicion of a fixed result, is formed according to one of the following rules:

$$S = \{z_k \mid M_k > M_{k-1}\}, \tag{8}$$

$$S = \{z_k \mid (p_k < p_{k-1}) \vee ((z_{k-1}\in S)\wedge(p_{k-1} < p_k + \Delta))\}, \tag{9}$$

where the sign $\vee$ is a disjunction (logical operation OR), the sign $\wedge$ is a conjunction (logical operation AND); $\Delta \in (0; 1)$ is the permissible difference between the p-value for conformity p_{k-1} of the previous and p_k of the current object. In this study, this parameter was chosen depending on how the p-value for conformity for all objects changed.

The rule described by (8) analyzes the change in the value of martingale in neighboring observations, and therefore in the future this rule will be termed a submartingale rule.

Using rule (9), the degree of difference between current and previous observations is analyzed, and therefore in the future this rule will be termed a p-value rule [21]. Its meaning is that the outcome of the current match is considered suspicious of the fixation if either:

a) the outcome of the current match is more suspicious than the outcome of the previous match;

b) the result of the previous match was suspicious, and the result of the current one is not much different from it.

5. 3. Analysis and evaluation of the effectiveness of the method for detecting match-fixing

First, the work of the proposed method was demonstrated for matches of class (1, 4) (Table 1). This group includes matches in which the host team belongs to class 1, that is, it is one of the most successful that season, and the guest team belongs to class 4, that is, it is characterized by one of the lowest values of success. Suspicious matches in terms of fixed outcomes in Table 1 are determined on the basis of the deviation of the result of the match from the expected result for this class. If the result was one or more goals higher, or the result of the match in nature (victory, defeat, draw) did not correspond to the expected, then such a match is considered potentially suspicious. Therefore, in the column «Potentially suspicious» it is marked by 1 and, additionally, its row in the table is highlighted in gray. The average result for the group $avg_{k(i,j)}$ is 1.25. So, the expected result of the match is to win by the host team at the level of 1 or 2 goals. All matches, the result of which deviates from the expected, are considered suspicious.

Table 1

Matches of class (1, 4) of the 2013–2014 French League II season

No.	Host team	Guest team	Outcome	Date of the event	Potentially suspicious
1	Metz	Laval	1:0	02-Aug-2013	0
2	Lens	CA Bastia	1:0	04-Aug-2013	0
3	Metz	Chateauroux	1:0	04-Oct-2013	0
4	Caen	Istres	4:0	08-Nov-2013	1
5	Nancy	Chateauroux	2:0	08-Nov-2013	0
6	Lens	Chateauroux	2:0	23-Nov-2013	0
7	Lens	Istres	1:2	19-Dec-2013	1
8	Nancy	Laval	2:1	20-Dec-2013	0
9	Caen	Laval	2:1	17-Jan-2014	0
10	Nancy	CA Bastia	0:1	17-Jan-2014	1
11	Lens	Laval	0:0	01-Feb-2014	1
12	Caen	Chateauroux	1:1	21-Mar-2014	1
13	Metz	Istres	2:1	07-Apr-2014	0
14	Metz	CA Bastia	1:0	18-Apr-2014	0
15	Caen	CA Bastia	6:1	02-May-2014	1
16	Nancy	Istres	3:1	02-May-2014	0

Further, the work of the offline deterministic algorithm is demonstrated. Each football match is a separate observation z_k, which are sequentially processed by the algorithm. First, for the current observation z_k, the measure of difference a_k (Fig. 2) is calculated using (4). The values of this measure show how much the outcome of the match differs in value and character from the expected result, which for this algorithm is the average result for the class of matches. The greater the value of the measure of difference, the more this match stands out from the rest in terms of the expected result.

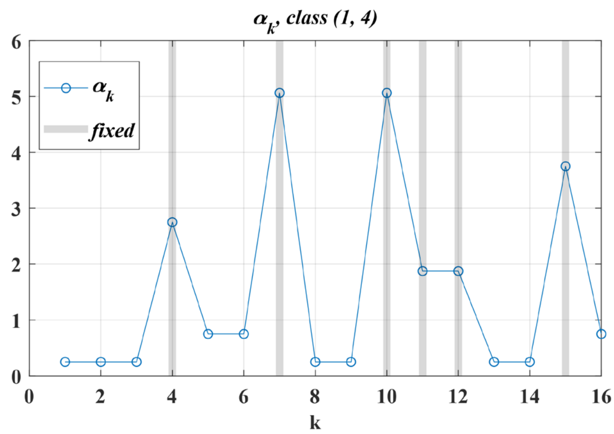

Fig. 2. Characteristics α_k for matches of class (1, 4). Gray columns highlight potentially suspicious matches

Further, according to the set of matches $\{z_1, z_2, ..., z_{k-1}, z_k\}$ for the current observation z_k, the measure of difference p_k (Fig. 3) is calculated using (5). This quantity takes values in the range $[1/_k; 1]$ and characterizes the proportion of such matches in the set $\{z_1, z_2, ..., z_{k-1}, z_k\}$, which are more different from the current match, or the same as the current match. Further, this characteristic can be analyzed according to the rule described by (9). According to this rule, a match is suspicious in which the degree of difference p_k is less than the degree of difference p_{k-1} of the previous match or for which the previous match has already been identified as suspicious and the degree of difference p_k of the current match is greater than the degree of difference p_{k-1} of the previous match by not more than Δ. Fig. 3 shows the results of detecting suspicious matches for the class (1, 4) at Δ=0. Also in this figure, dashed lines highlight matches that, using (9), are true positives (TP, green) and false negatives (FN, yellow). Based on the values of these characteristics, precision metric (1), recall metric (2), and measure F_1 (3) are calculated. So, for the class of matches (1, 4), using (9), at Δ=0, the offline deterministic algorithm for the recall metric worked by 67 %: most of the expected suspicious matches were detected. According to the precision metric, the algorithm worked 100 %: all expected matches suspicious of the match-fixing result were detected, and there were no mistakes in other matches. The measure F_1 for the class (1, 4), in this case, is 0.8, that is, the algorithm as a whole for the class (1, 4) worked well but there is room for improving the performance.

Fig. 4 shows the results of detecting suspicious matches for class (1, 4) using (9) at Δ=0.2. The introduction of a non-zero Δ-corridor even visually improved the results somewhat, but led to false positives (FP), which are indicated in the same figure with red dashed lines. So, for the class of matches (1, 4), using (9) at Δ=0.2, the offline deterministic algorithm, according to the recall metric worked by 100 %: all expected suspicious matches were detected. According to the precision metric, the algorithm worked by 75 %: all expected matches were detected but such matches were also detected that are not considered suspicious. The measure F_1 for the class (1, 4), in this case, is 0.86, that is, the algorithm as a whole for the class (1, 4) worked well and better than at $\Delta=0$, but there is still room for improving the result.

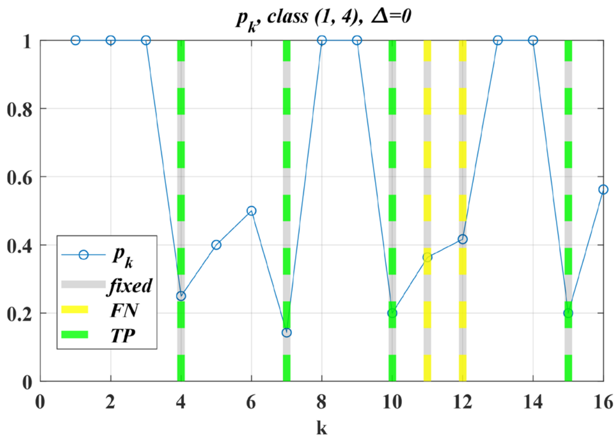

Fig. 3. Characteristics p_k for matches of class (1, 4) and results of detection of suspicious matches at $\Delta=0$ according to the *p*-value rule

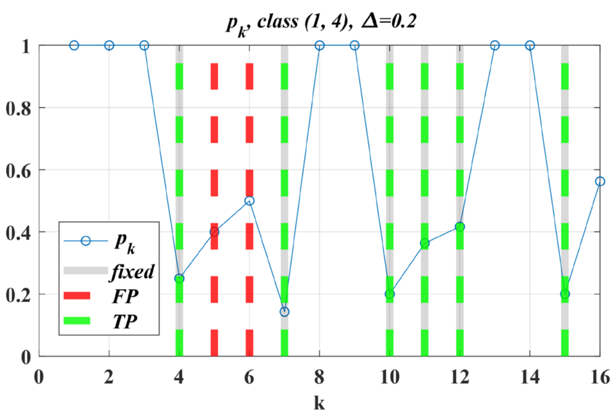

Fig. 4. Characteristics p_k for matches of class (1, 4) and results of detection of suspicious matches at $\Delta=0.2$ according to the *p*-value rule

The last step of the method is the calculation of the power integral martingale M_k. Fig. 5 shows the results of its calculation for the class of matches (1, 4). To reduce the range of values along the vertical axis, Fig. 5 shows not the characteristic M_k itself but the natural logarithm from it, that is, $\ln(M_k)$. Further, this characteristic is analyzed according to the principle described by (8) from the description of the algorithm: the points at which the value of the martingale increases compared to the previous ones correspond to matches that are suspicious of the match-fixing result. In Fig. 5, dashed lines also highlighted matches that, by (8), are true positives (TP, green), false positives (FP, red), and false negatives (FN, yellow). Based on these detections, as well as for the rule (9), the precision metric (1), recall metric (2), and measure F_1 (3) are calculated. So, for the class of matches (1, 4), using (8), the offline deterministic algorithm for the recall metric worked 100 %: all expected suspicious matches were detected. According to the precision metric, the algorithm worked by 67 %: all expected matches were detected, but such matches that are not considered suspicious have also been identified. The measure F_1 for the class (1, 4) is 0.8, that is, the algorithm as a whole for the class (1, 4) worked well but there is room for improving the result.

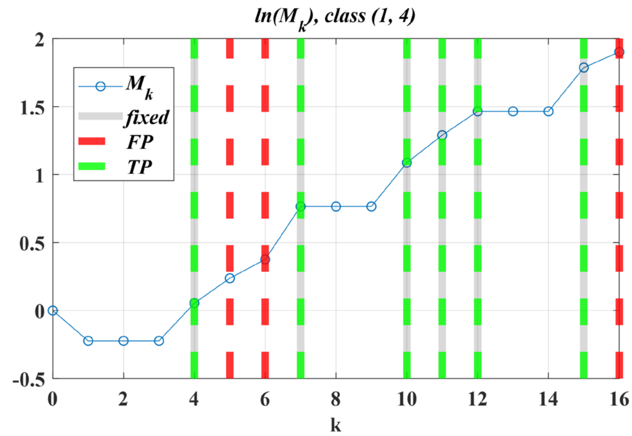

Fig. 5. Characteristics $\ln(M_k)$ for matches of class (1,4) according to the submartingale rule

Next, we considered the work of the offline deterministic algorithm on a symmetric class of matches (4, 1). The average result of the class $avg_{k(i,j)}$ is −1.5. So, the expected result of the match is to win the guest team at the level of 1 or 2 goals. All matches, the result of which deviates from the expected, are considered as suspicious; in Table 2, the lines corresponding to these matches are painted gray. Most of the suspicious matches in this class have either an overestimated or zero result.

Table 2

Matches of class (4, 1)
of the 2013–2014 French League II season

No.	Host team	Guest team	Outcome	Date of the event	Potentially suspicious
1	Laval	Caen	1:2	09-Aug-2013	0
2	CA Bastia	Nancy	1:2	10-Aug-2013	0
3	Laval	Lens	0:2	24-Aug-2013	0
4	Chateauroux	Caen	0:2	18-Oct-2013	0
5	Istres	Metz	1:2	01-Nov-2013	0
6	CA Bastia	Metz	0:2	22-Nov-2013	0
7	CA Bastia	Caen	1:5	13-Dec-2013	1
8	Istres	Nancy	0:0	13-Dec-2013	1
9	Chateauroux	Metz	0:1	14-Mar-2014	0
10	Chateauroux	Nancy	0:3	11-Apr-2014	1
11	Istres	Caen	2:3	11-Apr-2014	0
12	Chateauroux	Lens	1:1	21-Apr-2014	1
13	Istres	Lens	1:6	06-May-2014	1
14	Laval	Nancy	1:0	06-May-2014	1
15	CA Bastia	Lens	0:2	16-May-2014	0
16	Laval	Metz	0:0	16-May-2014	1

A plot of the measure of difference a_k of each match in this class is shown in Fig. 6. Unlike the situation with class (1, 4), in this class of matches, the measure of the difference between each potentially suspicious match is greater than the measure of the difference between other matches. Thus, for this class, a simplified principle of searching for suspicious matches could be applied – by the difference in the value of this measure from its most frequent value on a given set of observations. But here this principle would work quite qualitatively only because a sufficient number of ordinary matches appeared before suspicious observations. However, the more general principles set by (8), and (9) will also work perfectly in this situation.

Fig. 7 shows the results of detecting suspicious matches for class (4, 1) according to the principle formulated for the degree of difference (9) at Δ=0.2. All potentially suspicious matches have been identified and there are no false positives. So, for the class of matches (4, 1), using (9) at Δ=0.2, the offline deterministic algorithm according to the recall metric worked 100 %: all expected suspicious matches were detected. According to the precision metric, the algorithm worked 100 %. The measure F_1 for the class (4, 1) is 1, which is a sign of the distinctive operation of the algorithm on this class of matches.

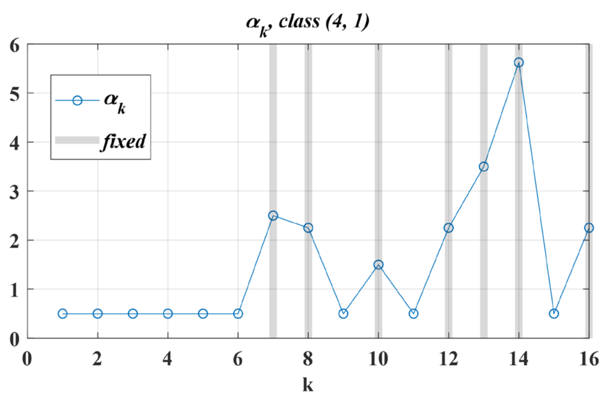

Fig. 6. Characteristics α_k for matches of class (4, 1). Gray columns highlight potentially suspicious matches

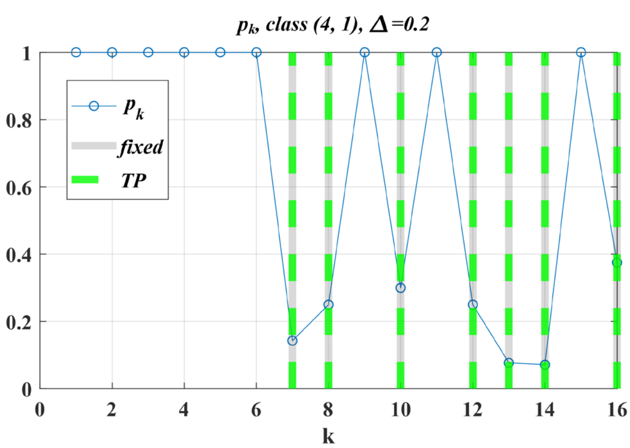

Fig. 7. Characteristics p_k for matches of class (4, 1) and results of detection of suspicious matches at Δ=0.2

Fig. 8 shows the results of detecting suspicious matches for class (4, 1) according to the principle formulated for the power martingale (8). The results are completely similar to the results from Fig. 7, that is, the offline deterministic algorithm both in terms of the recall metric and the precision metric worked by 100 %. Measure F_1 for class (4, 1) is 1, which is a sign of the distinctive operation of the algorithm on this class of matches.

Estimates of the effectiveness of applying the offline deterministic algorithm to different classes of matches according to the principle described (8) are given in Table 3. Cells in the columns of precision metrics P, R, F_1 have a range of four-color coloring. Cells with values from the range [0,4; 0,6) or [40 %; 60 %) are painted in red. Cells with values from the range [0,6; 0,75) or [60 %; 75%) are painted orange. Cells with values from the range [0.75; 0.9) or [75 %; 90 %) are painted in yellow. Cells with values from the range [0,9; 1] or [90 %; 100 %] are green.

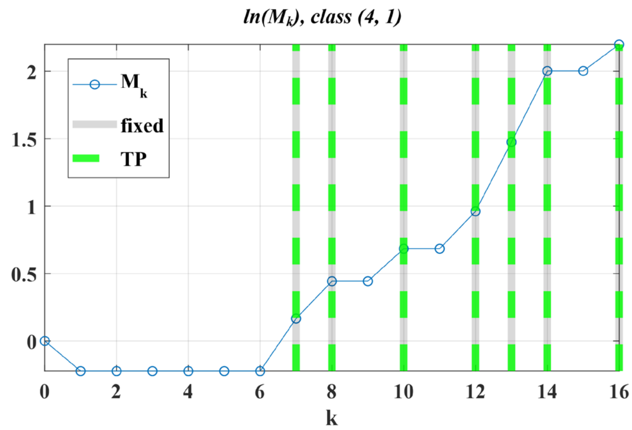

Fig. 8. Characteristic ln(M_k) for matches of class (4, 1)

Table 3

Evaluation of the effectiveness of the algorithm with the submartingale rule using (8)

Class	TP	FN	FP	P	R	F_1
(1, 1)	3	2	1	75 %	60 %	0.667
(1, 2)	8	1	0	100 %	89 %	0.941
(1, 3)	14	1	0	100 %	93 %	0.966
(1, 4)	6	0	3	67 %	100 %	0.8
(2, 1)	7	0	4	64 %	100 %	0.778
(2, 2)	9	0	4	69 %	100 %	0.818
(2, 3)	12	0	11	52 %	100 %	0.686
(2, 4)	9	0	0	100 %	100 %	1
(3, 1)	9	0	7	56 %	100 %	0.72
(3, 2)	17	0	4	81 %	100 %	0.895
(3, 3)	18	1	9	67 %	95 %	0.783
(3, 4)	15	0	5	75 %	100 %	0.857
(4, 1)	7	0	0	100 %	100 %	1
(4, 2)	13	0	3	81 %	100 %	0.897
(4, 3)	12	0	9	57 %	100 %	0.727
(4, 4)	6	0	0	100 %	100 %	1

The precision metric (P) in 3 situations reaches 52–57 %, although the recall metric is 100 %.

In the vast majority (9 out of 16) situations, the precision metric takes values from the range [75 %; 100 %], that is, according to the precision metric, the algorithm works quite efficiently.

According to the recall metric (R), the algorithm works perfectly in the vast majority (75 %) of situations, that is, suspicious matches are recognized as suspicious in 12 out of 16 classes of matches. According to the measure F_1, the algorithm also works well in 75 % of situations, i.e., $F_1 \geq 0.75$ in 12 out of 16 classes of matches.

Estimates of the effectiveness of applying the offline deterministic algorithm to different classes of matches with results according to the principle described by (9) are given in Table 4. The results are given for the values of 0 and 0.2 of the parameter Δ. Cells in the columns of precision metrics P, R, F_1 have the same color design as in Table 3.

Table 4

Evaluation of the effectiveness of the algorithm with the p-value rule using (9)

Δ	Δ=0						Δ=0.2					
Class	TP	FN	FP	P	R	F1	TP	FN	FP	P	R	F1
(1, 1)	2	3	1	67 %	40 %	0.5	3	2	2	60 %	60 %	0.6
(1, 2)	5	4	0	100 %	56 %	0.714	6	3	0	100 %	67 %	0.8
(1, 3)	9	6	0	100 %	60 %	0.75	12	3	0	100 %	80 %	0.889
(1, 4)	4	2	0	100 %	67 %	0.8	6	0	2	75 %	100 %	0.857
(2, 1)	6	1	2	75 %	86 %	0.8	6	1	3	67 %	86 %	0.75
(2, 2)	8	1	1	89 %	89 %	0.889	9	0	1	90 %	100 %	0.947
(2, 3)	10	2	3	77 %	83 %	0.8	12	0	4	75 %	100 %	0.857
(2, 4)	7	2	0	100 %	78 %	0.875	7	2	0	100 %	78 %	0.875
(3, 1)	6	3	5	55 %	67 %	0.6	8	1	6	57 %	89 %	0.696
(3, 2)	13	4	0	100 %	76 %	0.867	17	0	0	100 %	100 %	1
(3, 3)	15	4	2	88 %	79 %	0.833	16	3	2	89 %	84 %	0.865
(3, 4)	8	7	2	80 %	53 %	0.64	13	2	3	81 %	87 %	0.839
(4, 1)	6	0	1	86 %	100 %	0.923	7	0	0	100 %	100 %	1
(4, 2)	8	5	0	100 %	62 %	0.762	12	1	0	100 %	92 %	0.96
(4, 3)	9	3	4	69 %	75 %	0.72	12	0	6	67 %	100 %	0.8
(4, 4)	4	2	0	100 %	67 %	0.8	5	1	0	100 %	83 %	0.909

The precision metric (P) in only one situation reaches 55–57 %, as opposed to the results when using (8).

At the same time, the precision metric at Δ=0.2 in 12 out of 16 situations takes values from the range [75 %; 100 %], that is, in the vast majority of situations, according to the precision metric, the algorithm works quite efficiently.

Unlike the results using (8), in this situation, according to the recall metric (R), the algorithm works perfectly in 6 out of 16 classes of matches at Δ=0.2, but it works well ($R\geq 75$ %) in 14 out of 16 classes of matches, while with Δ=0 the algorithm works well only in 8 out of 16 situations. According to the measure F_1, the algorithm also works well in 75 % of situations, i.e., $F_1\geq 0.75$ in 12 out of 16 classes of matches.

As a result of the use of conformal predictors and power martingales, templates were obtained for identifying fixed football matches, the quality of which can reach high values in terms of the precision metric, recall metric, and measure F_1.

6. Discussion of results of using conformal predictors and power martingales to identify match-fixing

We analyzed the results of the measure of difference for matches of class (1, 4) from Fig. 2. For example, matches No. 1, No. 4, No. 7, and No. 15 were compared. The result of match No. 1 is equal to 1, that is, it does not differ significantly from the expected result of 1.25 both in the nature of the result and in the value. The result of match No. 4 is 3, that is, it differs significantly from the expected result of 1.25 in value. The result of match No. 7 is –1, i.e., it is significantly different from the expected result of 1.25 in character. The result of match No. 15 is 5, that is, it differs significantly from the expected result of 1.25 in value more than the result of match No. 1. A measure of difference for match No. 1, a_1=0.25; for match No. 4, a_4=2.75; for match No. 7, a_7=5.06; and for match No. 15, a_{15}=3.75. Thus, for this characteristic, several patterns can be distinguished:

a) if the outcome of the match does not differ from the expected result in character and differs by no more than one goal, a_k takes the value from the range [0; 1];

b) if the outcome of the match does not differ from the expected result in character and differs by more than one goal, a_k is determined only by the difference between the current result and the expected one;

c) if the result of the current match differs from the expected result in nature, α_k is also fined by multiplying by a penalty multiplier of 1.5 or 1.5^2, depending on the difference in the nature of the current and expected results;

d) the measure of difference for matches whose outcome differs from the expected result both in character and value (by 1 goal) is greater than the degree of difference between matches whose outcome differs only in value by 1, 2, or 3 goals.

The following is an analysis of the results from Table 3. For classes (2, 3), (3, 1), (4, 3), the algorithm shows the lowest efficiency in terms of precision metric. There are quite a few matches in these match classes that are not suspicious. At the same time, their measure of difference takes the values lying between the value that characterizes the most frequent ordinary match of this class and the values that characterize suspicious matches. At the same time, the martingale, which is calculated using (6), has the following property: if $p_{k-1}\neq p_k$, $p_{k-1}<1$, $p_k<1$, then $M_{k-1}^{(\varepsilon)}<M_k^{(\varepsilon)}$. In other words, if the degrees of difference between the previous and the current match are less than 1 and are not the same. This causes an increase in the martingale value for the current match compared to the previous one, that is, it makes us consider more matches suspicious of the fixed result.

Regarding the analysis of the results from Table 4. For class (3, 1), (9) resulted in the classification of normal matches as suspicious. In this class, there are many ordinary matches whose measure of difference takes the values that lie between the value that characterizes the most frequent ordinary match of this class and the values that characterize suspicious matches. This explains such frequent detection of regular matches as suspicious in the class (3.1).

In most other situations, on the one hand, with the introduction of the Δ-corridor, there is an increase in the efficiency of detecting suspicious matches, but in many classes, it is not sufficient for the excellent operation of the algorithm according to the precision metric. This happens in a situation similar to the one shown in Fig. 3: the value of the degree of difference of the current match p_k can be greater and much greater than the value p_{k-1}+Δ. This corridor can be increased but then it is possible to reduce efficiency in terms of precision metric.

In general, with the introduction of the Δ-corridor, as confirmed by the results given in Table 4, the recall of the proposed classifier does not deteriorate (never in Table 4), and the value of measure F_1 does not decrease (in all cases from Table 4, except for one). According to the precision metric, there was a decrease in the effectiveness of the algorithm in 5 groups of matches. In other situations, the precision metric has not deteriorated or improved.

Unlike [12–15, 17, 18], the developed method directly solves the problem of detecting matches suspicious of a fixed result. The methods considered in [12–15, 17, 18] solve other problems, in particular, forecasting the result or rates. On the other hand, the tasks under consideration can act as part of the task of detecting suspicious matches and represent possible ways to improve the developed method.

Unlike [12–15, 17, 18], our method uses input data of a smaller dimensionality: only 5 data attributes containing a historical description of the match. This allows it to be applied, having historical data on football matches that are more accessible than complete information about each football match. At the same time, the method devised can be supplemented with the models proposed in [12, 13] to improve the accuracy of predicting the outcome of the match.

Our method of detecting match-fixing makes it possible to decide on the suspicion of the match only on the basis of data on all matches of the corresponding season. Thus, the prediction of the result of a regular match is limited only to data on the current season and does not take into account the data of previous seasons of the corresponding tournament. This restriction can be eliminated by grouping the teams of all other seasons into groups in the same way and combining the corresponding groups of individual seasons into one group for all seasons.

Also, the method devised depends on the division of the teams of the season into groups, which, although performed on the basis of the rating of teams, is performed by the K-means clustering method, which depends on the number of partitions (groups) specified by the user. Such a limitation can be eliminated by using methods for finding the optimal number of partitions (for example, the Elbow method) or advanced clustering methods that do not depend on the user number of groups.

The disadvantage of the method built is the identification of match-fixing based on the search for matches that differ significantly from the expected result. If a fixed match has a result that is close to the predicted one, then it cannot be detected as suspicious. To search for such matches, you need to use additional data (for example, bets, and player trajectories) that are not publicly available for any matches, unlike the result and other historical attributes of the match.

Taking into account the chronology of matches, the proposed offline deterministic classification algorithm can be turned into an online algorithm that can «learn» and improve the precision of detecting suspicious matches over the course of the season's history. Such further development of this method is due to the need for the automatic processing of information online to help experts in decision-making.

In general, our method of detecting football matches that are suspicious of the presence of a fixed result can be applied to other sports competitions. This is relevant because the problem of match-fixing is an important one in different sports. However, such an expansion of the scope of application will require, at least, the construction of specialized measures of difference and, most likely, additional rules for making decisions on determining a suspicious match or event.

7. Conclusions

1. The formalized description of the input data used in the current work in the form of a data structure minimizes the number of data attributes and is based on information on the results of sports competitions that is always available in open sources. A data structure is formed containing a chronological history of the results of football seasons, the ranking of teams, and matches of the season, depending on the overall result of the teams in the season. The suspicion of the match for the presence of a fixed result is determined by the success of the participating teams and the outcome of the match.

2. A feature of the proposed method is the use of a specially developed measure of nonconformity and generalized statistical characteristics: the p-value for conformity and power submartingale. The introduction of a new measure of nonconformity provided an opportunity to compare the actual result of the match with the results of all other matches of the group and made it possible to take into account both the absolute results of the teams and the difference in the results of the actual and predicted results. The use of submartingale made it possible to obtain generalized statistics on the p-value for conformity of previous and current matches.

3. The proposed decision submartingale rule for detecting match-fixing allows such matches to be detected only based on an increase in the value of the submartingale. The quality of the algorithm according to this rule on the selected data is on average 78 % for the precision metric, 96 % for the recall metric, and 0.846 for the metric F_1. The proposed p-value rule makes it possible to take into account the proximity of the p-value values of the previous and current matches. The best quality of the algorithm according to this rule on the selected data is achieved by entering the threshold of 0.2 and, on average, is equal to 85 % for the precision metric, 88 % for the recall metric, and 0.853 for the metric F_1. If you need to get a high quality indicator in terms of recall metric, then it is advisable to use the submartingale rule. The p-value rule, in turn, makes it possible to get a high quality indicator in terms of the precision metric of detecting suspicious matches with a sufficiently high level of the recall metric.

Conflicts of interest

The authors declare that they have no conflicts of interest in relation to the current study, including financial, personal, authorship, or any other, that could affect the study and the results reported in this paper.

Funding

The study was carried out within the framework of research work No. 2310p «Information-analytical system for mathematical modeling and management of social risks with application in engineering and medicine» (State registration number – 0120U102216).

Data availability

The data will be provided upon reasonable request.

Acknowledgments

The authors would like to express their deep gratitude to the staff of the Applied Mathematics Department of the Faculty of Applied Mathematics at the National Technical University of Ukraine «Igor Sikorsky Kyiv Polytechnic Institute» for their support in conducting the study within the framework of research work No. 2310p «Information-analytical system for mathematical modeling and management of social risks with application in engineering and medicine») (State registration number – 0120U102216).

References

1. Lilley, E. (2015). A Review of the recommendations of the 'Report of the Sports Betting Integrity Panel' in assessing the progress towards tackling Match-fixing in Sport. Laws of the Game, 1 (1). Available at: http://journals.staffs.ac.uk/index.php/lotg/article/view/79/172
2. Huggins, M. (2018). Match-Fixing: A Historical Perspective. The International Journal of the History of Sport, 35 (2-3), 123–140. doi: https://doi.org/10.1080/09523367.2018.1476341
3. Constandt, B., Manoli, E. (2022). Understanding match-fixing in sport: Theory and practice. Routledge. doi: https://doi.org/10.4324/9781003162681
4. Hill, D. (2009). How Gambling Corruptors Fix Football Matches. European Sport Management Quarterly, 9 (4), 411–432. doi: https://doi.org/10.1080/16184740903332018
5. Chertov, O., Tavrov, D. (2014). Microfiles as a Potential Source of Confidential Information Leakage. Studies in Computational Intelligence, 87–114. doi: https://doi.org/10.1007/978-3-319-08624-8_4
6. Tzeng, C.-C., Lee, P.-C. (2020). Understanding match-fixing from the perspective of social capital: A case study of Taiwan's professional baseball system. International Review for the Sociology of Sport, 56 (4), 558–577. doi: https://doi.org/10.1177/1012690220917060
7. Tzeng, C.-C., Ohl, F. (2022). Examining the fabrics of match-fixing: The underground sport betting system. International Review for the Sociology of Sport, 58 (1), 188–207. doi: https://doi.org/10.1177/10126902221095688
8. Forrest, D., McHale, I. G. (2019). Using statistics to detect match fixing in sport. IMA Journal of Management Mathematics, 30 (4), 431–449. doi: https://doi.org/10.1093/imaman/dpz008
9. IOC targets match fixers. Available at: https://www.bbc.co.uk/blogs/davidbond/2011/03/match_fixing_is_now_a.html
10. Manoli, A. E., Antonopoulos, G. A. (2014). 'The only game in town?': football match-fixing in Greece. Trends in Organized Crime, 18 (3), 196–211. doi: https://doi.org/10.1007/s12117-014-9239-3
11. Yilmaz, S., Manoli, A. E., Antonopoulos, G. A. (2018). An anatomy of Turkish football match-fixing. Trends in Organized Crime, 22 (4), 375–393. doi: https://doi.org/10.1007/s12117-018-9345-8
12. Razali, N., Mustapha, A., Yatim, F. A., Ab Aziz, R. (2017). Predicting Football Matches Results using Bayesian Networks for English Premier League (EPL). IOP Conference Series: Materials Science and Engineering, 226, 012099. doi: https://doi.org/10.1088/1757-899x/226/1/012099
13. Anfilets, S., Bezobrazov, S., Golovko, V., Sachenko, A., Komar, M., Dolny, R., Kasyanik, V. et al. (2020). Deep multilayer neural network for predicting the winner of football matches. International Journal of Computing, 19 (1), 70–77. doi: https://doi.org/10.47839/ijc.19.1.1695
14. Narizuka, T., Yamazaki, Y., Takizawa, K. (2021). Space evaluation in football games via field weighting based on tracking data. Scientific Reports, 11 (1). doi: https://doi.org/10.1038/s41598-021-84939-7
15. Titman, A. C., Costain, D. A., Ridall, P. G., Gregory, K. (2014). Joint Modelling of Goals and Bookings in Association Football. Journal of the Royal Statistical Society Series A: Statistics in Society, 178 (3), 659–683. doi: https://doi.org/10.1111/rssa.12075
16. Van Der Hoeven, S., De Waegeneer, E., Constandt, B., Willem, A. (2019). Match-fixing: Moral challenges for those involved. Ethics & Behavior, 30 (6), 425–443. doi: https://doi.org/10.1080/10508422.2019.1667238
17. Gudmundsson, J., Wolle, T. (2014). Football analysis using spatio-temporal tools. Computers, Environment and Urban Systems, 47, 16–27. doi: https://doi.org/10.1016/j.compenvurbsys.2013.09.004
18. Janetzko, H., Sacha, D., Stein, M., Schreck, T., Keim, D. A., Deussen, O. (2014). Feature-driven visual analytics of soccer data. 2014 IEEE Conference on Visual Analytics Science and Technology (VAST). doi: https://doi.org/10.1109/vast.2014.7042477
19. Historical analysis of closing odds. Available at: https://github.com/Lisandro79/BeatTheBookie
20. Chertov, O., Zhuk, I., Serdyuk, A. (2021). Search of the Deviation from the Natural Process Using Stepanets Approach for Classification of Functions. 2021 11th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS). doi: https://doi.org/10.1109/idaacs53288.2021.9660997
21. Vovk, V. (2014). The Basic Conformal Prediction Framework. Conformal Prediction for Reliable Machine Learning, 3–19. doi: https://doi.org/10.1016/b978-0-12-398537-8.00001-8
22. Ho, S.-S., Wechsler, H. (2010). A Martingale Framework for Detecting Changes in Data Streams by Testing Exchangeability. IEEE Transactions on Pattern Analysis and Machine Intelligence, 32 (12), 2113–2127. doi: https://doi.org/10.1109/tpami.2010.48