Detection of match-fixing in football matches using a conformal anomaly detector

Oleg Chertov and Ivan Zhuk

Igor Sikorsky Kyiv Polytechnic Institute, 37, Prospekt Beresteiskyi, 03056, Kyiv, Ukraine

Abstract

A complex problem that threatens the integrity and authority of football in many countries of the world, including Ukraine, is fixed matches, as they are also called – matches with a fixed result. The results of fixed matches, related to the winning of bets, can be considered atypical, or abnormal, which allows formalization of a search for such matches. To check the current match for a fixed result, mathematical methods of football analytics, such as prediction of the match result, and analysis of bets or actions of the match participants throughout the game, are used. Their advantage is the speed of decision-making, and the disadvantage is the need to use a huge amount of data, that is not publicly available. An approach when the decision about the fixedness of the match is made after the end of the season, based on the results of the games played by all teams, can be considered as an alternative. This approach allows to formalize the search of matches, suspicious for a fixed result, as the detection of contextual anomalies. Statistical non-parametric histogram methods are the most adequate for the considered task of identifying suspicious for a fixed result matches, according to the results of the whole season. However, for effective use, these methods require a significant volume of the sample, which is not performed for the considered task. A new method of finding anomalies in data is a conformal anomaly detector. It does not require knowledge of the distribution laws of the input data and also allows entering estimates of guaranteed accuracy for the obtained solutions. A method of detecting suspects for a fixed result of football matches based on the results of the entire season, using a conformal anomaly detector, has been developed. To evaluate the effectiveness, main classification metrics were used: precision, recall, and F_1 metrics. The peculiarities of using the method, based on the conformal anomaly detector, according to the data of individual classes of the model season are considered. A comparative analysis of the developed and histogram methods was carried out based on the data of the model season. Proposed detection method based on conformal anomaly detector provides a gain in detecting potentially suspicious fixed-score matches compared to the known histogram method by 13%-17% in the precision metric, 13%-21% in by the recall metric and 0.15-0.23 by the F1 metric.

Keywords[1]

Match-fixing, conformal prediction, nonconformity measure, contextual anomaly, football season, goal difference, anomaly threshold

1. Introduction

Matches with a fixed result are a real problem that threatens the integrity and authority of football in many countries of the world, including Ukraine. The European Commission defines match-fixing as the manipulation of sports results, which includes agreements on the course or outcome of a sports competition or any of its events (e.g., a match, a race) to obtain financial benefit for oneself or others,

Information Technology and Implementation (IT&I-2023), November 20-21, 2023, Kyiv, Ukraine
EMAIL: oleg.r.chertov@gmail.com (A. 1) ivan.sergeyevich.zhuk@gmail.com (A. 2)
ORCID: 0000-0003-0087-1028 (A. 1); 0000-0001-8440-427X (A. 2)

and with the aim of completely or partially eliminate the uncertainty that is usually associated with the results of competitions [1]. Even though there is still no single authoritative definition of match-fixing, in its basic form it can be defined as losing or playing to a predetermined result in sports matches by illegally manipulating the results in one's favor [2]. A fixed match is characterized by the fact that its result and/or a certain course of events (penalty award, a player receiving a warning or expulsion, etc.) are predetermined, i.e., fixed. Today, such matches are qualified as a criminal offense from a legal point of view.

The French sports agency Sportradar, which specializes in monitoring sports events, in its annual report for 2022 [3] noted that the field of football is the most vulnerable to match-fixing: in 2022, 775 fixed football matches were detected worldwide, which accounts for 64% of all fixed-score matches across sports. The largest number of fixed sports competitions was recorded in Europe (630 matches), Asia (240 matches), and South America (225 matches).

According to the UN classification, two groups of matches are distinguished: (1) fixed matches to win bets and (2) fixed matches to get sports results. In fixed matches related to winning bets, the goal is to get a match result that is different from the expected one to make the most of the bet. Therefore, the results of such matches can be considered atypical, or anomalous, which allows the formalization of the search for matches with a fixed result.

2. Related Work

It should be noted right away that attackers cannot use classic methods of preserving data publishing [4-6] to hide the agreement of a football match because, in the absolute majority of cases, the main statistical data about the match (place and date of the match, the score after the first half and the final score, etc.) are generally known and cannot be distorted.

To check the current match for a fixed result, mathematical methods of football analytics are used, such as prediction of the match result, and analysis of bets or actions of the match participants throughout the game [7-15]. Statistical methods [9] (in particular, Bayesian networks) and machine learning methods [10-11] are used to predict the outcome of a football match. These methods can be used to identify "anomalies" in match results. But their disadvantage is the need to use a significant number of match attributes, which are not always available, and the lack of the possibility of obtaining analytical regularities for predicting the result

Methods based on betting analysis are also used to detect matches with a fixed result [7-8]. If during a match the difference between the actual betting volume and the predicted volume is statistically significant, the match is considered fixed. However, rate information is also often not publicly available.

Approaches of analyzing the performance of a player or a team in a game have gained significant development [12-15]. To assess the quality of a player's work, the trajectories of this player's movement during different matches starting with him in the same playing position are compared. Based on such a comparison, it is possible to assume the fixedness of the result of the match, based on the significant difference in the "work" of the player in this match and in other matches. However, these methods also require large amounts of data that are not publicly available.

An alternative can be considered an approach where the decision about the fixedness of the match is made based on the results of the whole season. At the same time, public information about the results of the games played by all teams is publicly available, which allows for the formalization of the search for suspicious matches with a fixed result as the detection of contextual anomalies.

The peculiarity of the task of identifying suspects for a fixed result of matches based on the results of the whole season is the lack of marking of normal and abnormal classes of data, which makes it necessary to consider it as a task of unsupervised learning. The most adequate for the considered task of identifying suspicious for a fixed match result for the results of the entire season, there are statistical non-parametric histogram methods [16]. This is because the input data is characterized by a small number of discrete numerical values. Also, the laws of the probability distribution of input values are unknown. However, for effective use, these methods require a significant volume of the sample, which is not performed for the considered problem.

The mathematical apparatus of conformal predictors is a new promising direction of finding anomalies in data. The advantages of this mathematical apparatus are the combination of the learning and forecasting process in one stage and independence from the probability distribution, of which the data is generated. Also, this approach allows entering estimates of guaranteed accuracy for the obtained solutions. In [17], based on the theory of conformal predictors, ***a conformal anomaly detector*** was proposed, which is a general algorithm for checking the anomaly of the current object using a measure of nonconformity and a set level of significance. It should be noted that in the works [17-18], the conformal anomaly detector and conformal predictor [19-20] were used to detect anomalies in the data in the online mode, that is, to check the current data in real time. At the same time, when constructing the measure of non-conformity, the predicted values of the current data were used based on the data obtained up to the current moment in time.

Therefore, the development of a method for detecting suspicious fixed match results, based on the results of the entire season using a conformal anomaly detector, based on the processing of exclusively publicly available public data, should be considered an urgent scientific task.

3. The method of detecting football matches suspicious for fixed result using a conformal anomaly detector

To identify contextual anomalies, one of the main steps is to extract contextual and behavioral attributes. We will use the goal difference as a behavioral attribute of a football match because it allows us to simply and unambiguously establish the result of the corresponding match. However, a match can be a contextual anomaly for one value of the goal difference (for example, the victory of a weak team over a strong one with a difference of three goals), but with the same value of this difference, that is, a behavioral attribute, be considered normal in another context (with the victory of a strong team over weak). As contextual attributes, we will take the attributes "team strength" and "type of game" - away or home. According to strength, teams are divided into groups. Groups are determined by one-dimensional or two-dimensional clustering. One-dimensional clustering is based on the number of scored points, and two-dimensional clustering is based on the number of scored points and the difference between scored and conceded goals of the teams in the season. Clustering makes it possible to distinguish ***groups of homogeneity of teams*** based on the results of the season. Based on contextual attributes, it will then be possible to divide the matches of the tournament into ***classes,*** and in each class of matches to use behavioral attributes to determine anomalous matches.

The unit of input data is an observation $z_{k,}$ describing a match of a football season, k is the ordinal number of a match in the season. Observation 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 group (rank) and the result of the host team of this match, and j_k and β_k is the group (rank) and the result of the visiting team of the match, T_k is the date of the match, and in the further, we will write the class of the match as an ordered pair (i_k, j_k).

Detection of matches suspicious for a fixed result using a conformal anomaly detector consists of the following stages:

1. for each match $z_k \epsilon G_{ij}$from the sequence $G_{ij} = (z_1, \dots, z_k, \dots, z_N)$the degree of non-conformity is calculated ($a_1, \dots, a_i, \dots, a_N$)in relation to all other objects:

$$a_k = |\alpha_k - \beta_k - avg(i,j)|,$$

$$a_k = A_N(\{z_1, \dots, z_{k-1}, z_{k+1}, , \dots, z_N\}, z_k),$$

$$a_N = A_N(\{z_2, \dots, z_{N-1}\}, z_N)$$

where $A_N$is a function that depends on the set of the form $\{z_1, \dots, z_{k-1}, z_{k+1}, \dots, z_N\}$and the object z_k, and sets a real number in accordance with these arguments: $A_N: \boldsymbol{Z}^{N-1} \times \boldsymbol{Z} \to \boldsymbol{R}$.

The degree of non-conformity $a_k$is calculated according to one of the formulas (1)-(2), which is the first stage in the calculation of the conformal predictor:

$$a_k = |\alpha_k - \beta_k - avg(i,j)|, \quad (1)$$

$$a_k = \left|\alpha_k - \beta_k - round\big(avg(i,j)\big)\right|, \quad (2)$$

2. using measures of the difference between the current k-th match and all other matches of the same class, the degree of conformity (difference) (p–value) of the match from the set of observations $\{z_1, \dots, z_k, \dots, z_{N-1}, z_N\}$ is calculated**:**

$$p_k = p(z_1, z_2, \dots, z_k, \dots, z_N) = \frac{\#\{i: a_i \geq a_k, 1 \leq i \leq N\}}{N}, \tag{3}$$

where the operation $\# A$ returns the number of elements in the set A. For example, for the set of integers {1, 2, 5, 10, 15, 17} operation # {1, 2, 5, 10, 15, 17} = 6. In formula (4), the numerator contains a set containing the numbers of such observations (matches), 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 of the numerator of this formula takes values in the range [1; N]. Accordingly, the value p_k takes a value in the range of $[\frac{1}{N}; 1]$.

3. Based on the degree of conformity of the match p_k a decision is made regarding the observed object class according to the following rule:

1. if

$$p_k < \varepsilon, \tag{4}$$

then the object $z_k$is considered conformally anomalous;

2. if

$$p_k \geq \varepsilon, \tag{5}$$

then the object $z_k$is considered normal, where $\varepsilon \in [0; 1]$is ***the abnormality threshold*** (anomaly threshold).

The set of all matches, for which condition (4) is satisfied, is called **a conformal anomalous predictor** and is denoted as $\Gamma^{\varepsilon}(z_1, z_2, \dots, z_k, \dots, z_{N-1}, z_N)$.

Let's analyze the computational complexity of this algorithm. Let there be n matches in the class of matches. The calculation of the average result of the matches consists of n operations and is performed once, since the entire sample is available from the beginning. Calculation of the measure of non-conformity for one match consists of 3 arithmetic operations and is repeated for each object of the sample. Calculating the p-value for one of the matches generally requires n comparison operations and 1 arithmetic operation and is also repeated for each sample object. Thus, in general, we have the following number of operations:

$$n + 3n + (n + 1)n = n^2 + 4n + n = n^2 + 5n = O(n^2)$$

that is, in the basic version, the considered algorithm of the conformal anomalous predictor has a computational complexity of $O(n^2)$. When using constant sorting of the array of nonconformity measure values after calculating each new a_k value, saving the number of repetitions of each unique value and its number in the array of unique values of this measure, the step of calculating the nonconformity measure over the entire sample, on average, will have $O(n\ log(n))$ of operations, and the p-value calculation stage – respectively, $2n$ operations: 1 operation of determining the number of required values of the non-conformity measure and 1 arithmetic operation, which are repeated for each object of the sample. Then we have the following number of operations:

$$n + O(n \log(n)) + 2n = O(n \log(n)) + 3n = O(n \log(n))$$

The definition of a conformal anomaly is consistent with the statistical definition of an outlier by Hawkins [21]. A conformal anomaly is an object z_k that deviates so much from z_1,..., $z_{k-1}, z_{k+1}, \dots, z_N$in terms of unconformity, that it is suspected that this object was formed by a mechanism different from that by which the other objects in the sample were formed.

It is shown [22] that conformal prediction, as well as its extension in the form of a conformal anomaly detector provide coverage guarantees for the degree of conformity p_k , namely: if the assumption of interchangeability or independence and identity of the distribution of sample objects is fulfilled $z_1, \dots, z_N$and the condition that one object falls on the detector per unit of time is fulfilled also, then for any degree of non-conformity $A_N$and $k \geq 1$the probability of an error in making a decision that object is not normal, does not exceed ε[22]:

$$P(p_k < \varepsilon) \leq \varepsilon. \quad (6)$$

Thus, the parameter εregulates the sensitivity of the conformal anomaly detector to the detection of anomalous objects [23]: this parameter is the proportion of anomalous objects that are detected as conformal anomalies. Setting this parameter also affects the detection precision, which is equal to the relative number of anomalous objects among those detected as conformal anomalies. A high parameter value ε can increase the sensitivity of the detector, but at the same time will reduce the detection precision and increase the frequency of false detections. Although achieving high sensitivity is important, it has been argued that the limiting factor in detecting anomalies is a reduction in precision [2 4]. This problem is called the base-level fallacy, and it consists of the fact that the precision of detection begins to yield to the frequency of false decisions about the anomaly, which occurs due to the low frequency of anomalous objects.

Therefore, the parameter should be adjusted ε depending on the level of precision acceptable in a specific application.

In the case of the operation of the conformal anomaly detector in the uncontrolled mode, it can be argued that *the value of the parameter ε should be set close to the a priori probability of the appearance of anomalous objects λ in order to achieve a good balance between sensitivity and detection precision* [2 3]. Indeed, assuming the existence of such ***an ideal*** measure of nonconformity A_N, that $a_i > a_j$ for any objects $z_i$and z_j, belonging to the anomalous and normal classes, respectively, it is intuitive that setting the parameter $\varepsilon = \lambda$ will result in a detection precision close to 1.

However, setting $\varepsilon < \frac{1}{N}$regardless of the degree of nonconformity $A_k$should always be avoided , as the sensitivity to anomalous objects would then be zero. To demonstrate this fact, suppose that we observe an abnormal object $z_k$such that $a_k \gg a_i \ \forall i = 1, \ldots, N$. It follows from the formula for $p_k$that $p_k = \frac{1}{N}$. Therefore, if $\varepsilon < \frac{1}{N}$, then the object $z_k$will not be classified as anomalous, even if it looks very extreme in terms of nonconformity.

To compare the proposed methods for detecting matches suspicious for a fixed result, you can use the well-known histogram method [16]: checking for anomalousness of the match based on ***the histogram of the goal differences*** for the current class of matches (i, j) by the level of abnormality p_A($D_{ij}^{(N)} = \emptyset$):

1. the value of the goal difference is selected $\tilde{d}$, which, according to the histogram of the current class of matches, (i, j) has the highest frequency of appearance $h_d$among those values $d$that do not belong to$D_{ij}^{(N)}$;

2. a value d is $\widetilde{\text{added}}$ to the set$D_{ij}^{(N)}$;

3. the total frequency of occurrence of all values from the set $D_{ij}^{(N)}$is calculated:

$$p_{ij}^{(N)} = \sum_{d \in D_{ij}^{(N)}} h_d \, ;$$

4. if $p_{ij}^{(N)} \geq 1 - p_A$, go to step 5, otherwise go to step 1;

5. values of possible goal differences$d^* \notin D_{ij}^{(N)}$form a set $D_{ij}^{(A)}$ of abnormal differences of the class of matches (i, j);

6. among all matches in the current class of matches, we define abnormal matches according to the following rule:

- a match is abnormal if the goal difference in it is $d \in D_{ij}^{(A)}$.

4. Analysis of the method based on the conformal anomaly detector using the data of individual classes of the model season

Methods for detecting football matches suspicious for fixed results can be considered as binary classifiers, which return a value 1 if the match is "potentially suspicious for fixed result" and 0 - otherwise. The following elements of the confusion matrix of the binary classifier will be important for further analysis: the number of correct activations (true positives, TP), the number of false positives (FP), the number of false negatives (FN). TP is equal to the number of matches that are potentially suspicious and were detected as such by the classifier. FP is equal to the number of matches that are not potentially suspicious but are considered as such by the classifier. FN equals the number of matches that are potentially suspicious but were mistakenly missed by the classifier. According to these characteristics, the metrics of precision (*P*), recall (sensitivity, *R*), and their harmonic average are calculated - measure F_1:

$$P = \frac{TP}{TP + FP} \tag{7}$$

$$R = \frac{TP}{TP + FN} \tag{8}$$

$$F_1 = \frac{2}{\frac{1}{P} + \frac{1}{R}} = \frac{2TP}{2TP + FP + FN}. \tag{9}$$

The selected characteristics are the basic characteristics of the analysis of the effectiveness of the algorithms that are used to solve binary classification problems. All three characteristics take values from the range [0; 1] and have a similar interpretation of the results: the closer the characteristic value is to 1, the more efficient the algorithm is from the point of view of this characteristic.

The analysis of the method was carried out using a simulation model [23]. A feature of the fixed-scoring soccer season simulation model used is that teams are divided into groups according to their strength based on season total points. Accordingly, the probability of scoring goals by a team during a match is calculated by groups, and not by the entire season. Also, when calculating this probability, the type of game is taken into account – home or away. This allows you to take into account the characteristics of the home and away team's game.

With the use of the mentioned statistical model, a model season was created. Determination of anomalous goal differences was carried out based on histograms of goal differences for each class of matches at the data anomaly level p_A =of 0.2. Histograms were constructed for 100 model seasons according to the method, introduced in [26]. After the determination of abnormal goal differences, 10 fixed matches were introduced in the current season according to the algorithm for the formation of fixed matches from [26]. All entered contractual matches were assigned class 2 in the "Potentially Suspicious Match" characteristic. Also, based on the determined abnormal goal differences, the marking of the matches of the season was carried out for their abnormality. All matches that were formed before the introduction of match-fixing and where the goal difference was abnormal were assigned class 1 in the "Potentially Suspicious Match" characteristic. An example of the class of matches after marking and entering contractual ones is shown in the table. 1 on the example of the match class (1, 4). Matches, that were entered as fixed, are marked in blue and have a value of 2 in the "Potentially Suspicious" column. Matches that have been simulated with an abnormal result are shown in gray in the table and have a value of 1 in the "Potentially Suspicious" column. All other matches, i.e., matches with an expected score, have a value of 0 in the "Potentially Suspicious" column.

First, let's consider the work of the proposed methods for detecting matches suspicious for a fixed result on the class of matches (1, 4) (Table. 1). This group includes matches in which the host team belongs to group 1, i.e., is one of the most successful in this season, and the away team belongs to group 4, i.e., it is characterized by one of the lowest success values. The average result for the group *avg (i, j)* is equal to 1.125. Therefore, the expected result of the match is a win for the home team with a goal difference of 1 or 2 goals.

Let's consider the results obtained when anomalous matches are detected by ***the goal difference histogram*** for the current class. Fig. 1 shows a histogram of goal differences for the class of matches (1, 4), on which abnormal goal differences are determined by the level of abnormality $p_A = 0{,}2$. In this case, 3, 4, and 5 turned out to be abnormal differences in balls. Also fig. 1 shows the results of detecting anomalous matches according to this histogram: the dashed lines highlight the matches that according to the goal difference histogram are correct activations (true positives, TP, green color). Based on these findings, the metrics of precision (7), recall (8), and measure F_1 (9) are calculated. So, for the class of matches (1, 4), the detection method based on the histogram of the goal differences of the current class of matches according to the recall metric (sensitivity) worked for 75 %: most of the expected suspicious matches were detected. According to the precision metric, the algorithm worked 100 %: all expected matches were detected, and there were no false detections. The measure of F_1 for the class (1, 4) is 0.86, that is, the histogram of ball differences for the class (1, 4) gave good results, but there is room for improvement.

Table 1
Matches of class (1, 4) of the model season

No	The host team	Guest team	Result	Potentially suspicious
1	team 1	team 15	0:0	0
2	team 2	team 15	6:2	2
3	team 3	team 15	3:2	0
4	team 4	team 15	1:2	1
5	team 1	team 17	2:2	0
6	team 1	team 18	2:1	0
7	team 1	team 20	1:1	0
8	team 2	team 17	2:0	0
9	team 2	team 18	2:2	0
10	team 2	team 20	2:2	0
11	team 3	team 17	2:2	0
12	team 3	team 18	6:1	1
13	team 3	team 20	4:1	1
14	team 4	team 17	2:0	0
15	team 4	team 18	3:3	0
16	team 4	team 20	2:1	0

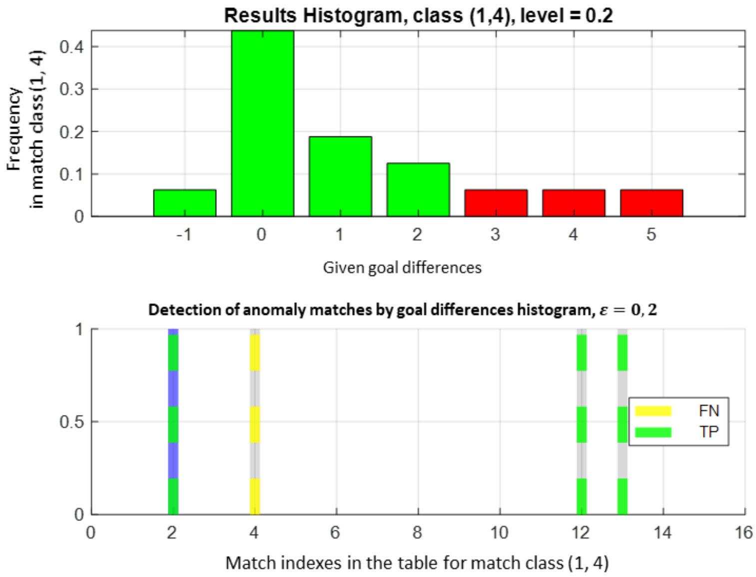

Figure 1: Histogram of ball differences for matches class (1,4) and suspicious match detection results from this histogram

Next, the operation of the method for detecting matches suspicious for a fixed result, based on a conformal predictor, is demonstrated. Each football match is a separate observation z_k which is sequentially processed by the algorithm. First for current observation z_k the difference measure is calculated a_k according to one of the formulas (1) – (2). Let's consider the results obtained when calculating the measure of non-conformity according to formula (1) (Fig. 2), that is, without rounding the average result by group. The values of this measure show how much the result of the match differs in value from the expected result, which for this method is the average result for the class of matches. The higher the value of the difference measure, the more this match stands out from the others in terms of the expected result. In fig. 2 and all subsequent gray columns highlight potentially suspicious matches according to the marking principle (those matches for which the value in the "Potentially suspicious" column is equal to 1), and blue - fixed matches that were created using the method of [23] (those matches for which the value in the "Potentially Suspicious" column is 2). The nonconformity measure of each potentially suspicious match is greater than the nonconformity measures of the other matches.

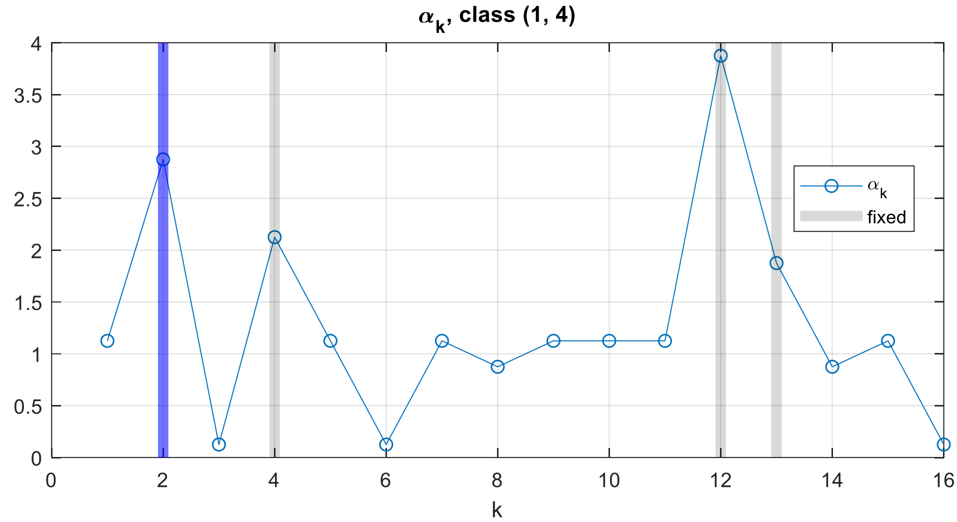

Figure 2: Characteristics α_k, calculated by formula (1) for matches of class (1, 4)

Further, by the set of matches $\{z_1, \ldots, z_k, \ldots, z_{N-1}, z_N\}$ and the obtained values of the measure of nonconformity α_k for each observation z_k the degree of conformity p_k is calculated (Fig. 3) according to (3). It takes values in the range [1/k; 1] and characterizes the proportion of such matches in the set $\{z_1, \ldots, z_k, \ldots, z_{N-1}, z_N\}$, that are more different than the current match or the same as the current match. Further, this characteristic can be analyzed according to ***the conformal anomaly detector rule*** (4). According to this rule, a match with a degree of conformity p_k smaller, than the abnormality threshold ε , is suspicious . Fig.3 shows the results of detecting suspicious matches for class (1, 4) at $\varepsilon = 0{,}2$. Also, in this figure, dashed lines highlight matches that, according to (4), are correct activations (true positives, TP, green color) and false passes (false negatives, FN, yellow color). Based on the values of these characteristics, the metrics of precision (7), recall (8) and measure F_1 (9) are calculated. Therefore, for the class of matches (1, 4) by (4) at $\varepsilon = 0{,}2$ the conformal anomaly detector by the recall metric (sensitivity) worked for 75 %: most of the expected suspicious matches were detected. According to the precision metric, the algorithm worked 100%: all expected matches suspected of fixing the result of the match were detected, and there were no errors in other matches. F_1 measure for the class (1, 4) in this case is equal to 0.86, that is, the algorithm generally worked well for the class (1, 4).

Now we will similarly consider the results obtained when calculating the measure of non-conformity, but according to formula (2) (Fig. 4), i.e., with rounding the average result by class of matches. The value of this measure, as well as the measure (1), shows how much the result of the match differs in value from the expected result, which for this method is the average result for the class of matches. The higher the value of the difference measure, the more this match stands out from the others in terms of the expected result. The only difference is that the match class average is now an integer. The average result for the class of matches (1, 4), taking into account rounding, is equal to 1. Based on the obtained values of the measure of non-conformity on this class of matches, it is possible to separate clearly the matches, that are normal, from those that are anomalous in their result. Thus, for this class, a simplified principle of searching for suspicious matches could be applied - by checking whether the measure of non-conformity is greater than 1. Further, by the set of matches $\{z_1, \ldots, z_k, \ldots, z_{N-1}, z_N\}$ and the obtained values of the measure of nonconformity α_k for each observation z_k the degree of conformity p_k is calculated (Fig. 5) according to (3). Fig. 5 shows the results of detecting suspicious matches for the class (1, 4) by (4) at $\varepsilon = 0{,}2$. At this abnormality threshold, two of the required matches will not be detected. There are also no false positives. So, for the class of matches (1, 4) by (4), the $\varepsilon = 0{,}2$conformal anomaly detector according to the recall metric (sensitivity) worked by 50%: half of the expected suspicious matches were detected. According to the precision metric, the algorithm worked at 100 %: among the detected matches, there are only expected matches. Measure F_1 for class (1, 4) in this case is equal to 0.67, that is, the algorithm in general for class (1, 4) worked well, but there is room for improvement. Compared to the results of the conformal anomaly detector, obtained using the nonconformity measure (1), the current results were worse in the recall measure: 1 less expected suspicious match was detected than when using the nonconformity measure (1).

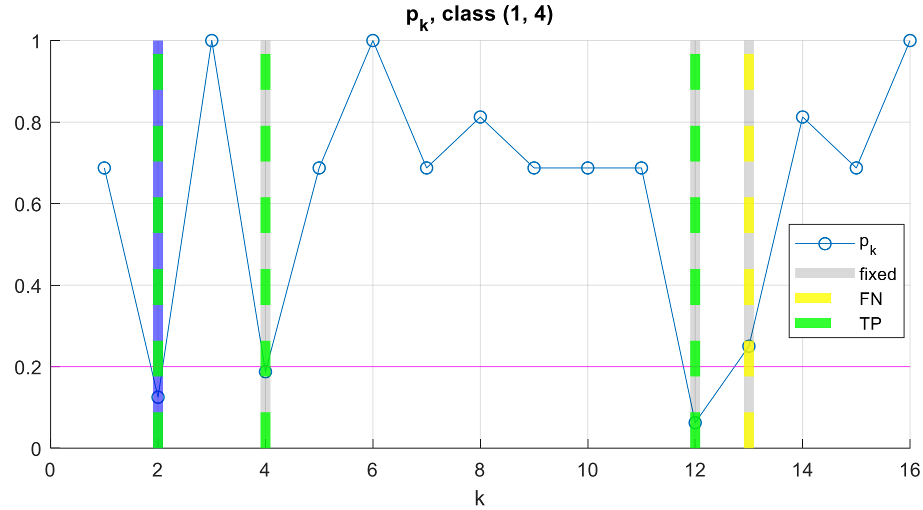

Figure 3: Characteristics *pk* _ for matches of class (1, 4), calculated by the measure of non-conformity (1) and the results of detecting suspicious matches at $\varepsilon = 0{,}2$ according to the conformal anomaly detector rule (4)

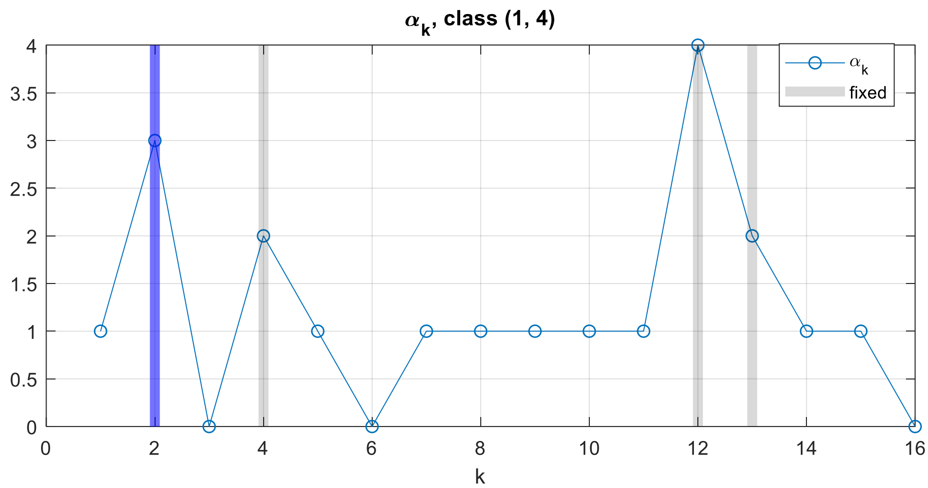

Figure 4: Characteristics $\alpha_{k,}$ calculated by the formula (2) for matches class (1, 4)

Now consider the operation of the methods on the class of matches (4, 1), that is, on the class symmetric to the previous one. Average result by class of matches $avg(i, j)$is equal to – 0.875. Therefore, the expected result of the match is a draw or a win for the visiting team with a difference of one goal. Most of suspicious matches in this class have a high match result: the goal difference is not less than 3 goals (Table 2).

Let's consider the results obtained when anomalous matches are detected by the goal difference histogram for the current class. Fig. 6 shows the histogram of goal differences for the class of matches (4, 1), on which for abnormality level $p_A = 0{,}2$ abnormal goal differences are determined. In this case, the abnormal differences of the balls turned out to be –2, 1, 3. Also fig. 6 shows the results of detecting anomalous matches according to this histogram: green dashed lines highlight matches that have a result corresponding to the red columns on the histogram and are marked as truly anomalous matches (true positives, TP, green color), red dashed lines highlight matches that have a result that corresponds to the red bars on the histogram and at the same time are not marked as anomalous matches (false positives, FP, red color).

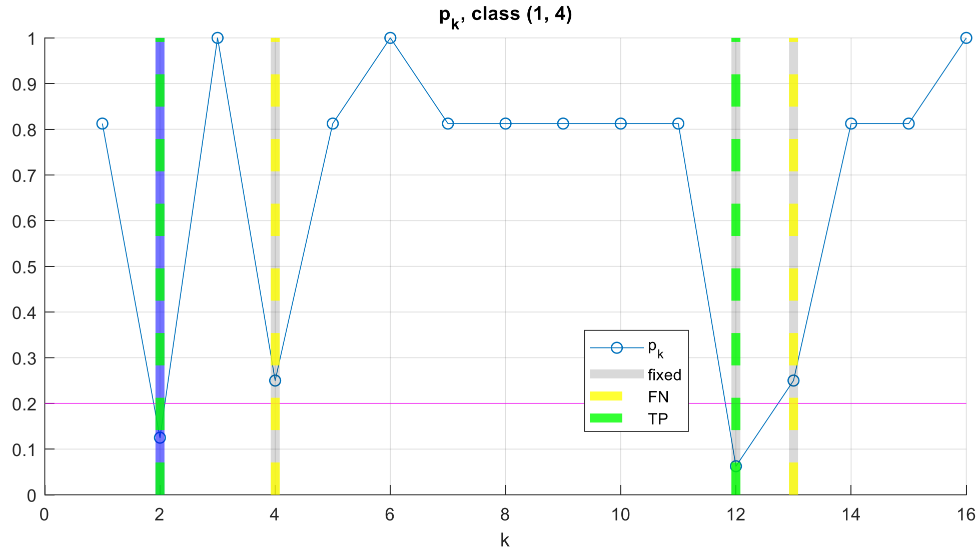

Figure 5: Characteristics *pk* _ for matches class (1, 4), calculated by the measure of non-conformity (2) and the results of detecting suspicious matches at $\varepsilon = 0{,}2$ according to the conformal anomaly detector rule (4)

Table 2
Matches of class (4, 1) of the model season

No	Home team	Away team	Result	Potentially suspicious
1	team 15	team 1	0:0	0
2	team 15	team 2	3:3	0
3	team 15	team 3	3:3	0
4	team 15	team 4	3:2	1
5	team 17	team 1	1:1	0
6	team 17	team 2	4:1	2
7	team 17	team 3	0:3	0
8	team 17	team 4	1:2	0
9	team 18	team 1	1:1	0
10	team 18	team 2	1:1	0
11	team 18	team 3	0:4	1
12	team 18	team 4	1:2	0
13	team 20	team 1	1:2	0
14	team 20	team 2	0:2	0
15	team 20	team 3	0:5	1
16	team 20	team 4	1:2	0

Based on these findings, the metrics of precision (7), recall (8) and measure F_1 (9) are calculated. So, for the class of matches (4, 1) according to the method of detection based on the histogram of goal differences of the current class of matches according to the recall metric (sensitivity) it worked for 67%: for the detected matches, most of them are the expected suspicious matches. According to the precision metric, the algorithm worked by 50 %: half of the expected matches were detected. The F_1 measure for class (4, 1) is equal to 0.571, that is, the histogram of ball differences for class (4, 1) gave results that require improvement. Graph of difference measure a_k of each match of this class, calculated by the formula (1), is shown in Fig. 7. Unlike the situation with class (1, 4), in this class there are matches whose degree of difference is on the same level as potentially suspicious matches. There is only one such match (#7) in this class of matches, but under certain conditions, such matches may cause false detections.

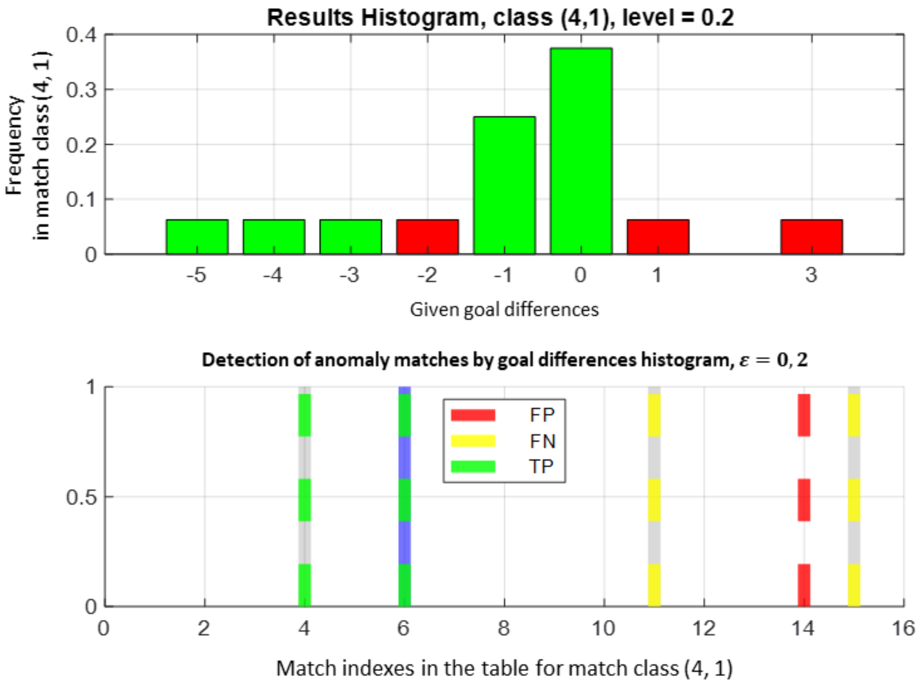

Figure 6: Histogram of ball differences for matches class (4, 1) and suspicious match detection results from this histogram

Fig. 8 shows the results of detecting suspicious matches for the class (4, 1) according to the principle formulated for *the conformal anomaly detector* (4) at $\varepsilon = 0{,}2$. In this figure, dashed lines highlight the matches which, according to (4), are correct activations (true positives, TP, green color) and false passes (false negatives, FN, yellow color). Based on these findings, metrics of precision (7), recall (8) and measure $F_{1\_}$ are calculated (9). So, at $\varepsilon = 0{,}2$ for the class of matches (4, 1), the conformal anomaly detector (4) according to the recall metric (sensitivity) worked for 75%: most of the expected suspicious matches were detected. According to the precision metric, the algorithm worked for 100 %. The F_1 measure for class (4, 1) is 0.86, which is a sign that the detector worked on this class of matches nice.

The histogram of the difference of goals for matches of classes (1,4) and (4,1) are maximally different from each other in comparison with other classes, which allows us to demonstrate more fully the peculiarities of the work of the developed method.

On both classes of matches, the developed method, when using a simple measure of non-conformity (1), made it possible to detect 75% of the expected suspicious matches, and at the same time, when using a measure of non-conformity with rounding (2) on one of the two classes of matches under consideration, only 50% of the expected suspicious matches were detected.

5. Comparative analysis of the methods based on model season data

Effectiveness estimates of the results of detecting matches, suspicious for a fixed result, by the histogram of the goal differences of the match class of the current model season by data abnormality levels $p_A = 0{,}2$ and $p_A = 0{,}3$ are given in the table 3. The considered model seasons were formed according to the algorithm considered in [23], using the data of the real season of 2013-2014 of the II League of France. Data mapping is based on match class goal difference histograms obtained over 100 model seasons. Detection, in turn, according to this method, is based on the histograms of the goal differences of match classes, built only for the current season. Cells in the columns of precision metrics P, R, and F_1 have a range of four-color paintings. Cells with values from the range [0.4; 0.6) or [40%; 60 %) are red/. Cells with values from the range [0.6; 0.75) or [60%; 75 %) are orange. Cells with values from the range [0.75; 0.9) or [75%; 90 %) are yellow. Cells with values from the range [0.9; 1] or [90%; 100%] are green.

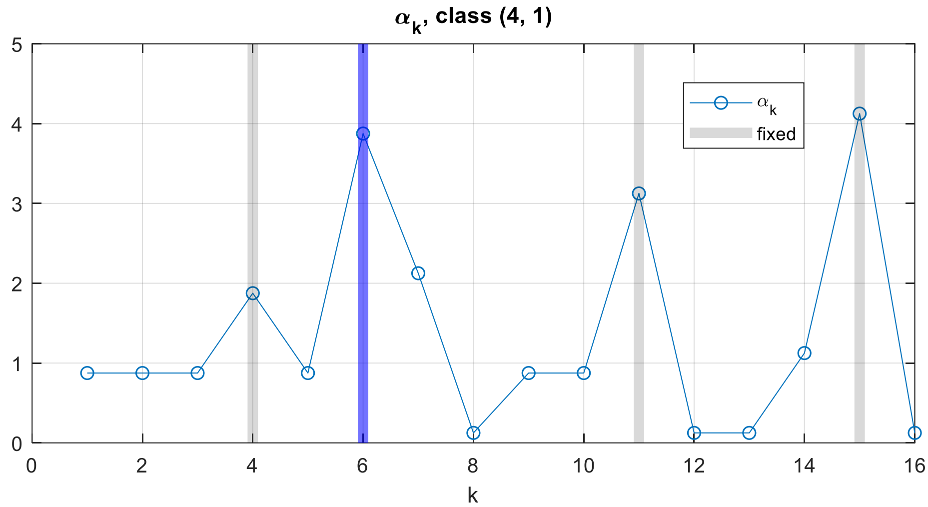

Figure 7: Characteristics α_k, calculated by the formula (1) for matches of class (4, 1)

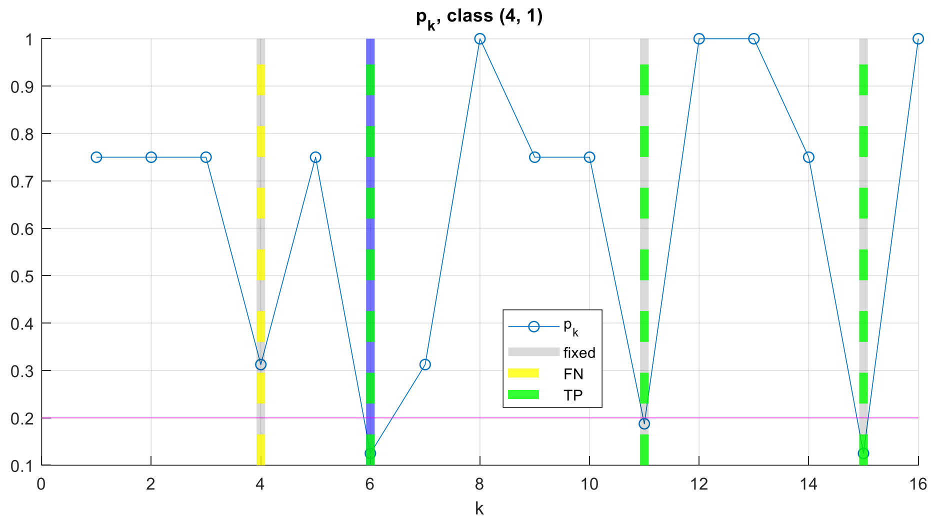

Figure 8: Characteristics *pk* _ for match class (4, 1), calculated by the measure of non-conformity (1) and the results of detecting suspicious matches at $\varepsilon = 0{,}2$ according to the conformal anomaly detector rule (4)

As can be seen from the table. 3, by levels of data abnormality $p_A = 0{,}2$and $p_A = 0{,}3$ the histogram anomaly detection method showed poor performance, with the F_1 metric showing almost the same performance in both cases. In particular, the following unique situation occurs in the results: in the class (1, 1), the algorithm did not detect any true anomalous match. From this, it can be concluded that the histogram of goal differences by class of matches, formed only for the current season, can be significantly different from such a histogram, constructed for many seasons. This, in turn, leads to the fact that anomalous matches determined by the goal difference histogram of many seasons, at a certain level of abnormality, can be considered non-anomalous by the goal difference histogram of the current season. It should also be noted that an increase in the level of abnormality resulted in an average increase in the quality of detection according to the precision metric, as well as a decrease in the quality according to the recall metric. This is because when the level of abnormality p_A increases, the number of abnormal data in the sample increases and, accordingly, the number of non-anomalous results decreases. This leads to an increase in correct detections of TP anomalies and a decrease in false detections of FP anomalies in expression (7), which leads to an increase in the precision metric. Conversely, with an increase in the sample of anomalous data, the number of false detections of normal data FN in expression (8) increases, which leads to a decrease in the recall metric. These regularities are further traced in all algorithms for detecting matches suspicious for a fixed result.

Table 3

Effectiveness estimates of the results of detecting matches, suspicious for a fixed result, based on ***the histograms of the goal differences*** of the match classes of the current season

Class	Amount samples, N	$p_A = 0{,}2$						$p_A = 0{,}3$					
		TP	FN	FP	*P*	*R*	*F1*	TP	FN	FP	*P*	*R*	*F1*
(1, 1)	12	0	2	1	0%	0%	0.00	2	0	1	67%	100%	0.80
(1, 2)	16	0	0	3	There are no abnormal matches			0	0	3	There are no abnormal matches		
(13)	35	2	6	2	50%	25%	0.33	2	6	2	50%	25%	0.33
(1, 4)	16	3	1	0	100%	75%	0.86	4	0	0	100%	100%	1.00
(2, 1)	16	2	0	1	67%	100%	0.80	2	3	1	67%	40%	0.50
(2, 2)	12	0	0	1	There are no abnormal matches			0	3	3	0%	0%	0.00
(2, 3)	32	5	0	0	100%	100%	1.00	8	0	0	100%	100%	1.00
(2, 4)	16	3	0	0	100%	100%	1.00	3	3	0	100%	50%	0.67
(3, 1)	32	6	0	0	100%	100%	1.00	6	4	0	100%	60%	0.75
(3, 2)	32	1	5	4	20%	17%	0.18	5	5	0	100%	50%	0.67
(3, 3)	56	5	0	4	56%	100%	0.71	9	0	0	100%	100%	1.00
(3, 4)	32	2	5	3	40%	29%	0.33	2	5	3	40%	29%	0.33
(4, 1)	16	2	2	1	67%	50%	0.57	3	2	1	75%	60%	0.67
(4, 2)	16	3	0	0	100%	100%	1.00	3	3	0	100%	50%	0.67
(4, 3)	32	4	2	1	80%	67%	0.73	7	0	0	100%	100%	1.00
(4, 4)	12	1	2	1	50%	33%	0.40	2	2	1	67%	50%	0.57
Average characteristics					66%	64%	0.64				78%	61%	0.66

Table 4

Effectiveness estimates of the method for detecting matches suspicious for a fixed result based on ***a conformal anomaly detector*** (4)

		$p_k < \varepsilon, \varepsilon = p_A$											
Class	Amount samples, N	$p_A = 0{,}2$						$p_A = 0{,}3$					
		TP	FN	FP	*P*	*R*	*F1*	TP	FN	FP	*P*	*R*	*F1*
(1, 1)	12	2	0	0	100%	100%	1.00	2	0	1	67%	100%	0.80
(1, 2)	16	0	0	0	There are no abnormal matches			0	0	0	There are no abnormal matches		
(13)	35	1	7	0	13%	100%	0.22	8	0	0	100%	100%	1.00
(1, 4)	16	3	1	0	75%	100%	0.86	4	0	0	100%	100%	1.00
(2, 1)	16	2	0	0	100%	100%	1.00	2	3	0	100%	40%	0.57
(2, 2)	12	0	0	2	There are no abnormal matches			0	3	2	0%	0%	0.00
(2, 3)	32	5	0	0	100%	100%	1.00	8	0	0	100%	100%	1.00
(2, 4)	16	3	0	0	100%	100%	1.00	3	3	0	100%	50%	0.67
(3, 1)	32	6	0	0	100%	100%	1.00	6	4	0	100%	60%	0.75
(3, 2)	32	1	5	4	20%	17%	0.18	5	5	0	100%	50%	0.67
(3, 3)	56	5	0	4	56%	100%	0.71	9	0	0	100%	100%	1.00
(3, 4)	32	2	5	0	100%	29%	0.44	7	0	0	100%	100%	1.00
(4, 1)	16	3	1	0	100%	75%	0.86	4	1	0	100%	80%	0.89
(4, 2)	16	3	0	0	100%	100%	1.00	3	3	0	100%	50%	0.67
(4, 3)	32	6	0	0	100%	100%	1.00	7	0	0	100%	100%	1.00
(4, 4)	12	2	1	0	100%	67%	0.80	3	1	0	100%	75%	0.86
Average characteristics					83%	85%	0.79				91%	74%	0.79

For various match classes effectiveness estimates of application of fixed matches detection method, based on ***a conformal anomaly detector*** (4), at the nonconformity measure (1) are given in the table. 4. The results are given for two cases of data marking: when marking at abnormality levels $p_A = 0{,}2$ and $p_A = 0{,}3$. The abnormality threshold value ε is chosen according to the rule $\varepsilon = p_A$ in accordance with the recommendations regarding the abnormality threshold from section 3. Cells in the columns of precision metrics *P, R, and F_1* have the same color design as in Table 3.

Increasing the level of abnormality resulted in an average 8% increase in detection quality for the precision metric and an 11% decrease in the quality in the recall metric compared to case $p_A = 0{,}2$. The indicator of the F_1 metric did not change on average. Table 5 shows the average indicators of the metrics of precision, recall, and F_1 for the considered methods of detecting suspicious about the fixed result matches, when using the measure of non-conformity (3.2) and the levels of abnormality $p_A = 0{,}2$ and $p_A = 0{,}3$. When using a simple measure of nonconformity on model season data, the proposed detection method based on ***conformal anomaly detector*** provides a gain in detecting potentially suspicious fixed-score matches compared to the known histogram method by 13%-17% in the precision metric, 13%-21% in by the recall metric and 0.15-0.23 by the F_1 metric.

Table 5
The average indicators of precision, recall, and F_1 metrics of the considered methods for detecting matches suspicious for a fixed result when using the measure of non-conformity (1)

Method	$p_A = 0{,}2$			$p_A = 0{,}3$		
	P	*R*	*F1* _	*P*	*R*	*F1* _
Histogram search method anomaly	66%	64%	0.64	78%	61%	0.66
Conform anomalous detector at $p_k < p_A$	83%	85%	0.79	91%	74%	0.79

5. Conclusions

1. Combining matches into classes based on contextual attributes allows you to use the average value of the goal difference of the corresponding class of matches as a predictive value of the numerical result of the match. The deviation of the actual result of the match from the expected one is considered as a characteristic of the abnormality of the match concerning the defined class of matches (context). Also, the introduction of the appropriate measure of non-conformity ensures the possibility of comparing the actual result of the match with the results of all other matches of the group and allows taking into account both the absolute results of the teams and the difference of the actual and predicted results.
2. The method developed based on the conformal anomaly detector for detecting suspicious for a fixed result football matches allows the detection of contextual anomalies of data in classes of matches, using the proposed measures of non-conformity, by comparing the degree of conformity (p-value) of the match with a threshold value. It belongs to the class of unsupervised learning methods and allows entering estimates of guaranteed accuracy for the obtained solutions. To achieve a good balance between detection sensitivity and precision, the threshold value should be set close to the a priori probability of the appearance of anomalous objects.
3. When using a simple measure of nonconformity on model season data, a detection method based on ***a conformal anomaly detector*** proposed will provide a gain in detecting potentially suspicious matches with a fixed result compared to the known histogram method by 13%-17% according to the precision metric, 13%-21% - according to the recall metric, and 0.15-0.23 - according to the F_1 metric.
4. In general, a method for detecting fixed football matches proposed can be applied both to other sports competitions and to other problem areas to solve the task of finding contextual anomalies (atypical transactions on a bank account, penetration into a closed network, anomalous number of messages in social networks on a certain topic, etc.).

6. References

[1] Match-fixing in sport a mapping of criminal law provisions in EU 27, Trends in Organized Crime, 18(3), 251–260, 2015, doi: 10.1007/s12117-015-9241-4.
[2] M.R. Haberfeld, D. Sheehan (Eds.). Match-Fixing in International Sports: Existing Processes, Law Enforcement, and Prevention Strategies. Springer International Publishing, 2013, doi: 10.1007/978-3-319-02582-7
[3] Sportradar. Betting corruption and match-fixing in 2022, 2023, URL: https://sportradar.com/wp-content/uploads/2023/03/Betting-Corruption-And-Match-Fixing-In-2022.pdf
[4] O. Chertov, D. Tavrov, Providing Group Anonymity Using Wavelet Transform, in: MacKinnon, LM (eds) Data Security and Security Data. BNCOD 2010. Lecture Notes in Computer Science, 2012, volume 6121. Springer, Berlin, Heidelberg. doi: 10.1007/978-3-642-25704-9_5

[5] O. Chertov, D. Tavrov, Microfiles as a Potential Source of Confidential Information Leakage, in: Intelligent Methods for Cyber Warfare. Studies in Computational Intelligence, 2015, volume 563. Springer, Cham. doi: 10.1007/978-3-319-08624-8_4

[6] A. Majeed and S. Leem, Anonymization Techniques for Privacy Preserving Data Publishing: A Comprehensive Survey," in IEEE Access, volume 9, pp. 8512-8545, 2021, doi: 10.1109/ACCESS.2020.3045700.

[7] A. C. Titman, D. A. Costain, P. G. Ridall, K. Gregory, Joint modeling of goals and bookings in association football. Journal of the Royal Statistical Society: Series A (Statistics in Society), 178(3), 659–683, 2015, doi:10.1111/rssa.12075

[8] D. Forrest, I. G. McHale. Using statistics to detect match-fixing in sport. IMA Journal of Management Mathematics, 30(4), 431–449, 2019, doi:10.1093/imaman/dpz008.

[9] N. Razali, A. Mustapha, F. A. Yatim, R. A. Aziz, Predicting Football Matches Results using Bayesian Networks for English Premier League (EPL). IOP Conference Series: Materials Science and Engineering, 226(1), 012099, 2017, doi:10.1088/1757-899X/226/1/012099

[10] S. Anfilets, S. Bezobrazov, V. Golovko, A. Sachenko, M. Komar, R. Dolny, V. Kasyanik, P. Bykovyy, E. Mikhno, O. Osolinskyi, Deep multilayer neural network for predicting the winner of football matches. International Journal of Computing, 70–77, 2020, doi:10.47839/ijc.19.1.1695

[11] A. Azeman, Football Match Outcome Prediction by Applying Three Machine Learning Algorithms. *International Journal of Emerging Trends in Engineering Research*, *8*(1.1), 73–77, 2020, doi:10.30534/ijeter/2020/1181.12020

[12] S. Fonseca, J. Milho, B. Travassos, D. Araújo, Spatial dynamics of team sports exposed by Voronoi diagrams. Human Movement Science, 31(6), 1652–1659, 2012, doi:10.1016/j.humov.2012.04.006

[13] T. Narizuka, Y. Yamazaki, K. Takizawa, Space evaluation in football games via field weighting based on tracking data. Scientific Reports, 11(1), 5509, 2021, , doi:10.1038/s41598-021-84939-7

[14] J. Gudmundsson, T. Wolle, Football analysis using spatio-temporal tools. Computers, Environment and Urban Systems, 47, 16–27, 2014, doi:10.1016/j.compenvurbsys.2013.09.004

[15] H. Janetzko, D. Sacha, M. Stein, T. Schreck, D. A. Keim, O. Deussen, Feature-driven visual analytics of soccer data. 2014 IEEE Conference on Visual Analytics Science and Technology (VAST), 13–22, 2014, doi:10.1109/VAST.2014.7042477

[16] J. Han, M. Kamber, J. Pei, Data Mining: Concepts and techniques, 3rd edition Morgan Kaufmann Publishers, 2012.

[17] R. Laxhammar, Conformal anomaly detection: Detecting abnormal trajectories in surveillance applications, 2014, URL: https://urn.kb.se/resolve?urn=urn:nbn:se:his:diva-8762

[18] R. Laxhammar, G. Falkman, Inductive conformal anomaly detection for sequential detection of anomalous sub-trajectories. Ann Math Artif Intel **74,** 67–94, 2015, doi:10.1007/s10472-013-9381-7

[19] I. Zhuk, O. Chertov, 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, 2023, doi:10.15587/1729-4061.2023.276977

[20] O. Chertov, I. Zhuk, Detection of fixed football matches based on the theory of conformal predictors using the modified Stepanets indicator function. Eastern-European Journal of Enterprise Technologies, 3 (4 (123), 22–32, 2023, doi:10.15587/1729-4061.2023.282645

[21] D. M. Hawkins, Identification of Outliers. Springer Netherlands, 1980, doi:10.1007/978-94-015-3994-4

[22] G. Shafer, V. Vovk, A tutorial on conformal prediction Journal of Machine Learning Research, 2007, doi:10.1145/1390681.1390693

[23] T. Fawcett, An introduction to ROC analysis. Pattern Recognition Letters, 27(8), 861–874, 2006, doi:10.1016/j.patrec.2005.10.010

[24] S. Axelsson, The base-rate fallacy and the difficulty of intrusion detection. ACM Transactions on Information and System Security, *3* (3), 186–205, 2000, doi:10.1145/357830.357849

[25] R. Laxhammar, G. Falkman, Sequential Conformal Anomaly Detection in trajectories based on Hausdorff distance. 14th International Conference on Information Fusion, 1–8, 2011.

[26] O. R. Chertov, I. S. Zhuk, Soccer season simulation with fixed matches, KPI Science News, no. 1–2, pp. 82–94, 2022, doi:10.20535/kpisn.2022.1-2.287916