

EVALUATION OF THE INFORMATIVENESS OF DISCRETE FEATURES OF A TRAINING SAMPLE IN CLASSIFICATION PROBLEMS

Shevchenko G. Ya. – PhD, Head of Department, Noosphere Association, Dnipro, Ukraine. ORCID: <https://orcid.org/0000-0003-3984-9266>.

Gerasimenko S. M. – Researcher, Noosphere Association, Dnipro, Ukraine. ORCID: <https://orcid.org/0009-0009-2597-9917>.

ABSTRACT

Context. Modern data mining methods are widely used to build classification and predictive models. However, when processing discrete data, the problem of quantitatively assessing the informativeness of features, which determines the accuracy and stability of classifiers, remains. The lack of a universal approach to measuring the contribution of both individual features and groups of features to the classification result complicates the process of automated feature selection and model optimization.

Objective. The aim of this paper is to develop and theoretically substantiate a method for assessing the informativeness of both individual features and arbitrary groups of discrete features, based on the relationship between the statistical characteristics of features and measures of class distinguishability.

Method. An approach is proposed that links the informativeness of both individual discrete features and arbitrary groups of such features with respect to a function characterizing the target variable. The method is based on the research results of renowned scientists Kendall and Stewart in the field of nonparametric statistics. For practical application, an algorithm for calculating the informativeness of both individual features and groups of features is introduced, suitable for implementation in data analysis software systems.

Results. It is demonstrated that the developed method enables formal and quantitative evaluation of the contribution of both individual features and arbitrary groups of features to the classification process without prior assumptions about the model type. It also enables the identification of hidden dependencies between features, which is impossible with individual assessments; i.e., it enables the identification of feature interactions. The resulting expressions provide a basis for automating feature selection when working with discrete data, improve the analytical value of the method, and offer a basis for meaningful feature selection.

Conclusions. The proposed approach enables the unification of the procedure for assessing the informativeness of both individual discrete features and arbitrary groups of features in data mining systems. It provides a formal link between the statistical characteristics of the data and the quality of the classification, which contributes to increased accuracy, robustness, and interpretability of models.

KEYWORDS: information content of features, discrete data, entropy, data mining, classification, feature selection, feature interaction, web service.

ABBREVIATIONS

IG is an information gain;
DM is a Data Mining;
OPT is an object-property table;
TS is a training samples;
DR is a decision rule;
IFG is an informative feature group.

NOMENCLATURE

x_i is a feature vector;
 y_i is a target variable (class label or numerical response);
 $\Phi(\cdot)$ is a dependence operator (mutual information, probability estimate, correlation coefficient, etc.);
 $H(Y)$ is an entropy of a target variable (before partitioning);
 $H(Y|X)$ is a conditional entropy after partitioning by feature value X ;
 $p(y)$ is a probability of each class;
 $p(x)$ is a proportion of objects with a specific value x ;
 $H(Y|X=z)$ is an entropy of a target variable within this group;
 O_{ij} is an observed (actual) number of objects with feature i and class j ;

E_{ij} is an expected number of objects with feature i and class j ;

N is a total number of objects;

p_i is a proportion of objects in class i .

$Gini_{parent}$ is $Gini_{impurity}$ of an original dataset;

$Gini_{split}$ is weighted average $Gini_{impurity}$ after splitting;

m is a number of subsets after splitting;

N_j is a number of objects in subset j ;

$Gini_j$ is a $Gini_{impurity}$ of subset j ;

function f_R describes a given partitioning of R sets of TS into classes;

x_i is a feature;

$V(x_i)$ is an informativeness of feature x_i ;

k is a number of classes;

m_Y is a number of objects belonging to class Y ;

m_{XY} is a number of objects belonging to class Y and taking the value $x = [0, \dots, k_{i-1}]$ of feature x_i ;

k_i is a number of gradations of feature x_i ;

X is a finite set of discrete features of these objects;

M is a set of object descriptions;

Y is a finite set of class labels;

Δ is an arbitrary set of feature values;

m_{Δ}^Y is a number of sample sets from the Y -class for which the relation $x_{ij}=t_{ij}$ ($j=1, \dots, \gamma$) holds;

t_{ij} is a value of feature x_{ij} in set Δ ;

Γ is a set of all sets of feature values x_{i1}, \dots, x_{ij} .

INTRODUCTION

Many modern institutions have information systems for storing a wide variety of data used by specialists to identify (diagnose) relevant processes. However, when analyzing data, discovering patterns in it, and extracting them, one must deal with the problem of dimensionality. The dimensionality of stored data, determined by the number of different features describing the state of the objects under consideration, is quite large, sometimes reaching several tens or hundreds of indicators. Therefore, the problem of reducing the dimensionality of the feature space and identifying the most informative features is highly relevant for such information systems. The following factors make this transition possible: information duplication, the low informativeness of individual features, and the construction of generalized features.

This article formulates the problem of selecting informative features as a combinatorial optimization problem with a computational complexity of $O(2^n)$, where n is the initial number of features under study. A brief overview of basic methods for reducing the feature space, such as principal component analysis and extreme feature grouping, is presented. This article examines and explores statistical methods for assessing the informativeness of discrete features used in classification problems: the Shannon method, the Gini method, the chi-square method, and nonparametric statistics. The application of these methods is demonstrated using a number of practical examples. A web service developed that implements the Gini method and a nonparametric method is described. The web service provides the user with a user-friendly interface for assessing the informativeness of features describing the state of an object, including arbitrary groups of features. This complex can be integrated into specialized information systems for diagnosing various processes.

In classification problems, it is important to understand which features are most significant for model construction. Uninformative or redundant features can reduce classification quality by introducing noise into the model, increase computational costs, complicating the training process, and cause overfitting if the model learns noise rather than patterns. Therefore, assessing the informativeness of features is a key step in building machine learning models. This problem is especially relevant for high-dimensional data, for example, in bioinformatics, credit scoring, and signal processing. In such cases, it is important to select only key features, making the task of assessing their informativeness critical [1–3].

This paper examines classical methods for assessing the informativeness of discrete features in data processing using various practical examples – IG, Gini coefficient, and the χ^2 criterion. A nonparametric method based on

assessing the relationship between features and the target variable is proposed, with the ability to account for feature interactions within groups. For each feature, informativeness values were calculated using the four methods described above. The informativeness of feature groups was also assessed using a nonparametric approach. All values are normalized in the range [0; 1].

This allows for a reduction in the dimensionality of the feature space by identifying the most informative features and informative feature groups. It is believed that features may initially be of different types; however, during the information content analysis stage, continuous numerical features can always be reduced to a discrete form using binning.

The use of DM tools begins only when prepared data exists, in the form of samples in which objects are represented by multidimensional data sets (feature descriptions). These samples are known by various names, including OPTs, TS or experimental data tables. It's important to note that selecting a feature description and preprocessing data are beyond the scope of DM, although they are present in any DM task. These issues are considered to be the domain expertise of the subject matter expert with whom the DM analyst interacts. This stage is typically referred to as the data design stage, and its main goal is to obtain high-quality OPTs or TSs, which means structuring and categorizing the data, i.e., obtaining information about the object of study.

The first step in such design is to define the target variable and the set of necessary features (parameters, factors, attributes – all synonyms) that influence them.

The next stage involves data collection. Data can be considered anything you interact with.

Determining the initial set of features that have the greatest impact on the target variable is a crucial task, as there are no rules for such a determination. Finding the set of features that have the closest relationships with the target variable requires engaging relevant scientists, engineers, and other specialists. At the initial stage of solving the problem, it is recommended to record as many features as possible that characterize the objects being studied and that can influence the outcome variable.

It is important to understand that the key factor is the interaction of certain groups of features with the target variable and the adherence to certain requirements for these features.

1 PROBLEM STATEMENT

In DM problems, when using a TS, the number of measured features is typically very large, but not all of them are equally important for constructing the DR. A large number of features complicates DR construction and leads to significant inconveniences during its use, impairs the interpretability of results, increases the volume of statistical material and the cost of obtaining data on objects, and overall reduces the quality of the entire pattern recognition procedure. To improve this, it is necessary to conduct a preliminary analysis aimed at assessing the key informative characteristics of the DR, in

particular, assessing the informativeness of features, evaluating feature values, identifying the most representative objects, informative groups of features, etc. Moreover, it is advisable to do this at the preliminary stage of analysis, after obtaining the TS, before constructing the DR itself.

In other words, it is necessary to navigate such a sample, and first, to determine the extent to which each feature is suitable for solving the main problem – constructing a DR. The degree of such suitability, or the contribution of a feature (or set of features) to explaining the target variable, essentially represents the informativeness or importance of such a feature relative to the function that determines the partitioning of the sample into classes. Formally, this is related to the reduction in the uncertainty/entropy of the target variable with knowledge of the feature (Shannon entropy, mutual information) and/or to the improvement in the quality of the model's prediction (increased accuracy, reduced error).

This entails, firstly, the task of defining formal approaches to the mathematical processing and evaluation of the informativeness of discrete features at the preliminary stage and, secondly, the task of selecting from the initial features a certain number of them and their combinations that are the most important and informative, ensuring a given level of reliability of the recognition algorithm (procedure). To find such features and their combinations, it is necessary to be able to quantitatively assess the informativeness of both the features themselves and arbitrary groups of such features.

Let's formulate the problem mathematically. Let's assume a TS:

$$D = \{(x_i, y_i)\}_{i=1}^N. \quad (1)$$

The informativeness of a feature x_j with respect to Y is determined by the functional

$$I(x_j; Y) = \Phi(p(x_j, Y)). \quad (2)$$

For a feature subset S :

$$I(S; Y) = \Phi(p(x_s, Y)). \quad (3)$$

Goal. Find a feature subset $S^* \subseteq \{1, 2, \dots, m\}$ that maximizes informativeness with respect to the target variable Y while minimizing the redundancy of the given feature set.

Since the number of subsets S grows exponentially (2^m), an exact solution to the problem is NP-hard. In practice, various methods for reducing the search effort are used, providing a compromise between accuracy and computational efficiency.

Thus, the problem of selecting informative features is formalized as an optimization problem that improves the generalization ability of the model.

In this regard, we will consider two cases – a quantitative assessment of the informativeness of individual features and a quantitative assessment of the informativeness of groups of features.

2 REVIEW OF THE LITERATURE

In modern data analysis, machine learning, and data mining, selecting informative features is a key step in building an effective model. This is especially true for discrete features that take a finite number of values (e.g., gender, category, diagnosis). Their redundancy or irrelevance can degrade the quality of the model, increase overfitting, and complicate interpretation.

From a practical perspective, feature selection aims to achieve a compromise between model accuracy, interpretability, and the robustness of the obtained results.

The most comprehensive review of modern approaches to feature selection is given in [4]. The authors distinguish three main groups of methods: filter methods, wrapper methods, and embedded methods. Filter methods use statistical measures that are independent of a specific model. For example, these include the Pearson and Spearman correlation coefficients [4], the χ^2 test [5], and several others [6, 7, 8].

The advantage of filters is their computational simplicity and scalability; however, they ignore interactions between features [9].

Wrapper methods evaluate the informativeness of features through the prediction quality of a specific model. Classic examples are sequential feature selection and recursive feature elimination [10].

These methods produce more accurate results but are computationally complex and prone to overfitting.

Embedded methods perform feature selection during model training. These include, for example, Decision Trees and Random Forests [11], which provide built-in feature importance assessment based on uncertainty reduction criteria.

For the case of preliminary analysis of training data, which is our primary interest, the works [12–16] should be noted.

For example, [12] proposed a formula for numerically assessing the usefulness of features based on cumulative frequencies, [13] proposed methods using statistical recognition theory to determine the informativeness of discrete features, [14] considered a method based on the idea of calculating the probability of individual feature values for each class of recognizable objects, [15] used the construction of a special metric based on information theory to solve this problem.

At the same time, the fundamental research of Kendall and Stewart in the field of nonparametric statistical problems should be noted. In [16], they proposed an estimate of the relationship between two categorized variables, where they showed that for some problems, in particular, predicting the categorized values of one variable from a known variable, the estimate they obtained is the best. This circumstance directly indicates the possibility of constructing on this foundation an

assessment of the information content of discrete features and their groups precisely at the stage of preliminary analysis of training information.

When it is necessary to evaluate the informativeness of only individual features, methods such as the information gain method [17], Gini [6], chi-square [5], and nonparametric methods [16, 18] are widely used. The choice of these methods is due to the following reasons: these methods are based on relatively simple algorithms for calculating the measure of informativeness, and the results of applying these methods are easily interpreted. These features are particularly important for various categories of the information system users, most of whom are not specialists in data analysis and IT technologies.

The information content of discrete features in machine learning refers to how useful each discrete feature is for solving a machine learning problem (e.g., classification, regression). Evaluation methods and approaches to working with discrete features have their own characteristics. Discrete features are features that take a finite or countable number of different values. They can be:

Categorical (nominal): Values have no natural order (e.g., color: “red”, “blue”, “green”; car type: “sedan”, “crossover”, “truck”).

Ordinal: Values have a specific order, but the distance between them can be undefined (e.g., clothing size: “small”, “medium”, “large”; satisfaction rating: “bad”, “neutral”, “good”).

Binary: Take only two values (e.g., “yes”/“no”, “0”/“1”).

The informativeness of discrete features allows for improved understanding of the data, such as identifying features that are most strongly related to the target variable, selecting the most relevant features to improve performance and reduce overfitting (i.e., improving the model), improving model interpretability (i.e., simplifying the model and making it easier to understand), and identifying data issues, such as uninformative or potentially problematic features, and determining which discrete features can be transformed or combined to create more informative features.

It is important to remember that the choice of feature informativeness assessment is influenced by the data types, machine learning tasks, the model used, and available computing resources. Furthermore, it is often useful to use multiple methods and compare their results. It is important to note that understanding feature informativeness is a key step in the process of building effective and interpretable machine learning models.

Various methods are used to assess the informativeness of discrete features of classification problems. This paper examines the specifics of a nonparametric approach to assessing the informativeness of discrete features and analyzes four approaches to selecting informative features: based on the IG method [17]; the Gini coefficient [6]; chi-square [5]; and based on nonparametric statistical methods [16, 18].

The value of such criteria or assessments is determined through their pragmatization, i.e., the construction of algorithms based on them, and this value also determines their level. The most important assessment of informativeness is the one that 1) is sufficient for developing an algorithm, 2) ensures the development of the algorithm in the shortest possible time, 3) ensures a high degree of algorithm reliability, and 4) ensures a uniform presentation of the informativeness assessment for both individual features and groups of features.

Method of evaluating the informativeness of features using IG is based on the decrease in entropy after partitioning the data by feature value. The greater the information gain, the more informative the feature.

$$IG(X, Y) = H(Y) - H(Y | X). \quad (4)$$

Entropy is defined as:

$$H(Y) = -\sum p(y) \log_2 p(y). \quad (5)$$

Conditional entropy is defined as:

$$H(Y | X) = \sum p(x) H(Y | X = x). \quad (6)$$

The higher the IG , the more informative the feature.

Note that the IG method estimates the informativeness of the feature under study as a normalized value that takes values from 0 to 1. This follows from formula (1), since the probability values p_{ik} range from 0 to 1, and the logarithm of such values is less than zero. In this case, the informativeness of feature x is said to be such that the closer $IG(x)$ is to 1, the higher the informativeness of x , and conversely, the closer $IG(x)$ is to 0, the lower the informativeness of x .

The chi-square test evaluates the statistical relationship between a discrete feature and a discrete target variable. The higher the chi-square value and the lower the p-value, the stronger the relationship.

The test calculates the difference between observed and expected frequencies using the formula:

$$\chi^2 = \sum_{ij} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}. \quad (7)$$

The expected frequencies E_{ij} are calculated using the formula:

$$E_{ij} = \frac{\sum_j O_{ij} \times \sum_i O_{ij}}{N}. \quad (8)$$

A large value indicates that the distribution of feature and class values is significantly different from random, and the feature is informative.

The Gini Index (Gini Impurity) measures the degree of uncertainty or heterogeneity in a dataset and thus allows one to evaluate the informativeness of features.

It is used to estimate the probability of misclassifying a randomly selected element. The lower the Gini Index, the better the feature discriminates between the data:

$$Gini = 1 - \sum (p_i)^2. \quad (9)$$

$Gini_{Gain}$ measures how much the Gini Index decreases after splitting the dataset by a specific feature:

$$Gini_{gain} = Gini_{parent} - Gini_{split}. \quad (10)$$

$$Gini_{split} = \sum_{j=1}^m \frac{N_j}{N} \times Gini_j. \quad (11)$$

The lower $Gini_{gain}$, the better the feature for splitting.

3 MATERIALS AND METHODS

In our view, the fundamental research of M. Kendall and A. Stewart in the field of nonparametric statistical problems [16] can serve as the basis for constructing optimal estimates of feature informativeness, including in the sense discussed above. They proposed an estimate of the relationship between two categorized variables, which, as they demonstrated, is the best for certain problems, in particular, for predicting the categorized values of another variable based on one variable.

Interpreting their approach in terms of machine learning, we can consider the predicted variable to be a finite-valued function f_R (i.e., a class number) that describes a given partitioning of R sets of TS into classes. The prediction is made based on another categorized variable, which in this case is feature x_i .

Then the estimate of the relationship between the two variables can be considered an estimate of the informativeness of feature x_i relative to function f_R . This estimate is calculated directly from the data of TS and characterizes the information that can be obtained about the function f_R , knowing the values of the feature x_i on the sets of TS.

Based on this approach, we propose the following method for calculating the informativeness $V(x_i)$ of features x_i , discussed below. The advantage of this method lies in the relative simplicity and transparency of the calculations.

The proposed formula has the following form [18]:

$$V(x_i) = \frac{1}{k} \sum_{x=0}^{k_i-1} \max\left(\frac{m_{XY}}{m_Y}\right). \quad (12)$$

The higher the value of $V(x_i)$, the more informative feature x_i is considered. It can be shown that $1/k \leq V(x_i) \leq 1$. Formula (12) is intended for the case of discrete features x_i .

Let's consider an example. We have the following table (Table 1).

For feature x_1 , we calculate its informativeness. We preliminarily define: $k=2$, $k_1=2$, $m_1=5$, $m_2=5$, $m_{01}=1$, $m_{02}=4$, $m_{11}=4$, $m_{12}=1$.

© Shevchenko G. Ya., Gerasimenko S. M., 2026
 DOI 10.15588/1607-3274-2026-2-10

Table 1

x_1	x_2	x_3	x_4	Y
0	0	1	1	1
1	1	0	0	1
1	1	0	1	1
1	0	1	1	1
1	0	0	1	1
1	1	0	1	2
0	0	0	0	2
0	1	1	1	2
0	1	0	0	2
0	0	0	1	2

$$V(x_1) = \left(\frac{1}{2}\right) \times$$

$$[(\max(m_{01}/m_1, m_{02}/m_2) + \max(m_{11}/m_1, m_{12}/m_2))] = \left(\frac{1}{2}\right) \times$$

$$[\max(1/5, 4/5) + \max(1/5, 1/5)] = \left(\frac{1}{2}\right) \times (8/5) = 0.8.$$

Proceeding in a similar manner for the remaining features we have:

$$V(x_2) = 0.6; V(x_3) = 0.6; V(x_4) = 0.6.$$

Informativeness of a group of features of the TS. Let Ω be a set of objects, $X = \{x_1, x_2, \dots, x_n\}$ a finite set of discrete features of these objects, M a set of object descriptions, and Y a finite set of class labels. For each object $\omega \in \Omega$, its feature description $\{x_1(\omega), x_2(\omega), \dots, x_n(\omega)\}$ is known – an n -dimensional vector, where the i -th coordinate of this vector is equal to the value of the i -th feature, and the class label y_k is known. The set of feature descriptions of objects and class labels from a given sample of objects $A \subseteq \Omega$ is defined as a matrix of size $|A| \times n$, called the TS. Let $I(Z)$ be the measure of the informativeness of a feature subset $Z \subseteq X$, defined on A . Among all the distinct subsets of X , we need to select a subset $Z^* \subseteq X$, such that

$$I(Z^*) = \max_{Z \subseteq X} \{I(Z)\}. \quad (13)$$

In pattern recognition theory, this problem is called feature selection [3, 4, 19]. Feature selection is computationally complex, since for $|X| = n$, searching through all distinct subsets of $Z \subseteq X$ requires $O(2^n)$ time. Even with a relatively small number of features, exhaustive search becomes impossible even with high-performance computing.

Next, we have three options. The first, described below, involves constructing new features. The second concerns feature selection. The third is based on taking into account the interactions between features.

The Feature Selection problem can be generalized by defining a transformation $Z = F(X)$ that allows a new feature space Z , $|Z| < |X|$, to be formed from X . In this formulation, the problem is called the Features Extraction problem [3]. When solving Features Extraction, new features are formed based on those already present in X .



The main methods for solving the Features Extraction problem are factor analysis and the extreme feature grouping method. Factor analysis allows one to identify generalized features (factors), each of which represents several original features at once. One method of factor analysis is the principal component analysis [19, 20]. Its essence consists of finding linear combinations of features from X and constructing, based on them, a smaller feature space $Z \subseteq X$, the information content of which is equivalent to the information content of X as a whole. As practice shows, the use of principal component analysis is most effective when all features $x_1, x_2, \dots, x_n \in X$ are of the same type and measured in the same units. Otherwise, the resulting linear combinations of the original features are difficult to interpret.

In the extreme grouping method, a correlation matrix is calculated using the “object-feature” matrix of sample A , and the set X is divided into groups such that features within a group are highly correlated, while relatively weak correlations are observed between groups. Each group of features is then replaced by a single resultant feature. A disadvantage of this method is the inability to determine the optimal number of groups [19].

The next Feature Selection approach is based on the following assumption. Estimates of the informativeness of all individual features are calculated. Based on heuristic considerations, only the most informative ones are retained, forming a group of informative features, and all further work is performed solely with this group. The drawback of this approach is obvious: it does not take into account possible interactions between features.

One of the main differences between DM and statistics is that it is not individual features that are particularly valuable, but rather their combinations, or rather, combinations of feature values that account for the interactions between features. In other words, seemingly “insignificant” features, or, more likely, some combinations of them, may prove decisive, i.e., the interactions between features may be important. The methods for assessing the individual importance of features discussed above do not take such interactions into account. Therefore, it is crucial to be able to also assess the informativeness of an arbitrary group of discrete features, which, in effect, allows for the consideration and evaluation of interactions between features. This also allows us to use this estimate to search for IFG in the TS, which allows us to talk about reduced enumeration methods.

Based on the approach of M. Kendall and A. Stewart discussed above in the field of nonparametric statistical problems, we can propose the following formula for estimating the informativeness of an arbitrary group of features, taking into account the interaction between the features:

$$V(x_{i1}, \dots, x_{ij}) = \frac{1}{k} \sum_{\Delta \in \Gamma} \max_Y \left(\frac{m_{\Delta}^Y}{m_Y} \right), \quad (14)$$

$$\Delta = t_{i1}, t_{i2}, \dots, t_{ij} \quad (0 \leq t_{ij} \leq k_{ij} - 1), j=1, \dots, \gamma \quad (1 \leq \gamma \leq n).$$

The following properties of the estimate can be noted:

1. $1/k \leq V(x_{i1}, \dots, x_{ij}) \leq 1$.
2. The information content of the features in (9) and (11) is determined relative to f_R , a function defining the partitioning of the set M of recognition objects, a feature absent from other known methods.
3. The numerical value of the score for a given feature coincides with the proportion of recognition objects that can be correctly assigned to the corresponding classes based on the values of the given feature; i.e., the score is unambiguously related to the classification errors of the recognition objects. Note that these properties follow directly from (12) and (14).
4. Score (14) is monotonically increasing.
5. Score (14) for the entire group of features can serve as an assessment of the recognition object quality; the higher the score, the better the recognition object quality. This allows: 1) a more informed approach to recognition object creation, testing its quality during its construction, and 2) the proposal of new IFG search schemes that are more practical than existing ones.

In many cases, it is desirable to have scores (12) and (14) within the range of 0 to 1. In addition to uniformity, this allows for comparison with other scores, which are most often measured within these ranges. It is easy to show that this can be achieved by multiplying scores (12) or (14) by the following weighting coefficient, given below:

$$W(x_{i1}, \dots, x_{ij}) = \frac{(V(x_{i1}, \dots, x_{ij}) \times k - 1)}{k - 1}. \quad (15)$$

In this case, scores (12) and (14) are within the range

$$0 \leq W(x_{i1}, \dots, x_{ij}) \leq 1.$$

An interesting feature of the introduced score is the ability to evaluate another type of feature information content – through its “contribution” to the information content of a group of features. This indicates the “weight” of the feature in the assessment of the informativeness of the group of features.

$$V^*(x_i) = V(x_{i1}, \dots, x_{i1-1}, \dots, x_{i1+1}, \dots, x_{ij}). \quad (16)$$

For example, the informativeness of features can be assessed differently, namely, through the “contribution” of a feature to the informativeness of the complete set of features. These properties highlight the advantage of the proposed approach to assessing the informativeness of features over other methods.

Let’s consider an example. We have the following table (Table 1).

Let’s calculate the informativeness of a group of features, let’s say x_2 and x_3 .

$$V(x_2, x_3) = \left(\frac{1}{2}\right) \times [\max(\frac{m_{10}^1, m_{00}^2}{m_1, m_2}) + \max(\frac{m_{01}^1, m_{01}^2}{m_1, m_2}) + \max(\frac{m_{10}^1, m_{10}^2}{m_1, m_2}) + \max(\frac{m_{11}^1, m_{11}^2}{m_1, m_2})] = 0.7.$$

The informativeness of other groups of features is calculated in a similar manner.

4 EXPERIMENTS AND RESULTS

Objective of the experiment. To test the effectiveness of various methods for assessing the informativeness of discrete features in classification problems and to demonstrate the capabilities of the proposed

nonparametric approach in accounting for feature interactions.

Methods Used. The following were used to assess information content:

- IG;
- Gini Coefficient;
- Chi-Square Test;
- A nonparametric method developed based on the ideas of Kendall and Stewart.

Experiment 1. For the analysis, the Mushroom Classification dataset was taken from <https://www.kaggle.com>, containing measurements of mushrooms for the purpose of classifying them as “edible” and “poisonous” (61.609 objects, 17 features, 2 classes).

Table 2 – Example of source data for calculating information content

class	cap-shape	cap-surface	cap-color	does-bruise-or-bleed	gill-attachment	gill-spacing	gill-color	stem-root	stem-surface	stem-color	veil-type	veil-color	has-ring	ring-type	spore-print-color	habitat	season
p	x	g	o	f	e	N/A	w	s	y	w	u	w	t	g	N/A	d	w
p	x	g	o	f	e	N/A	w	s	y	w	u	w	t	g	N/A	d	u
p	x	g	o	f	e	N/A	w	s	y	w	u	w	t	g	N/A	d	w

Table 3 presents the results of calculating the informativeness of features using the 4 methods discussed above.

Table 3 – Individual informativeness of features (the top 3 values are highlighted in bold, 1 in green, 2 in yellow, and 3 in red)

Feature	IG	Gini	χ^2	V
cap-shape	0.0282	0.0183	781.0306	0.5671
cap-surface	0.0419	0.0258	134.2699	0.5837
cap-color	0.0448	0.0286	69.7101	0.593
does-bruise-or-bleed	0.0003	0.0002	19.9677	0.5076
gill-attachment	0.0368	0.025	1.8485	0.5856
gill-spacing	0.0094	0.0065	1.4175	0.5375
gill-color	0.0269	0.0182	311.3171	0.5826
stem-root	0.0521	0.026	2.4546	0.5469
stem-surface	0.0628	0.0327	1236.57	0.5882
stem-color	0.0605	0.0369	674.0732	0.6148
veil-type	0.0021	0.0014	8.8159	0.5118
veil-color	0.0301	0.0148	44.3648	0.5208
has-ring	0.0024	0.0016	152.0365	0.525
ring-type	0.0411	0.0207	142.3501	0.5326
spore-print-color	0.0389	0.0206	314.9121	0.5473
habitat	0.0264	0.0154	275.753	0.5464
season	0.0098	0.0068	205.1296	0.5395

Calculations of the informativeness of feature groups using formula (14) yielded the following results:

Table 4 – Informativeness of groups of 2 features using the nonparametric method (top 10)

Groups of features	V
cap-surface + gill-color	0.732
cap-surface + stem-color	0.7319
cap-surface + gill-attachment	0.7218
gill-attachment + stem-color	0.7087
cap-surface + stem-surface	0.7078
cap-shape + cap-surface	0.7057
gill-color + stem-color	0.7012
cap-surface + cap-color	0.7
cap-shape + gill-attachment	0.6997
gill-attachment + gill-color	0.6976

Table 5 – Informativeness of groups of 3 features using the nonparametric method (top 10)

Groups of features	V
cap-surface + cap-color + gill-attachment	0.869
cap-surface + gill-attachment + stem-color	0.8654
cap-surface + gill-color + stem-color	0.8566
cap-surface + gill-attachment + gill-color	0.8566
cap-surface + gill-attachment + stem-surface	0.8494
cap-surface + gill-color + stem-surface	0.8493
cap-shape + cap-surface + gill-color	0.8393
cap-shape + cap-surface + gill-attachment	0.8329
cap-shape + cap-surface + stem-color	0.8298
cap-surface + cap-color + stem-color	0.8278

Table 6 – Informativeness of groups of 4 features using the nonparametric method (top 10)

Groups of features	V
cap-surface + cap-color + gill-attachment + gill-color	0.9433
cap-surface + cap-color + gill-attachment + stem-surface	0.9415
cap-surface + gill-attachment + gill-color + stem-color	0.9353
cap-surface + cap-color + gill-attachment + stem-color	0.9325
cap-shape + cap-surface + cap-color + gill-attachment	0.9296
cap-surface + gill-attachment + stem-surface + stem-color	0.9265
cap-shape + cap-surface + gill-attachment + stem-color	0.9265
cap-shape + cap-surface + gill-attachment + gill-color	0.9265
cap-surface + gill-attachment + gill-color + stem-surface	0.9242
cap-surface + gill-attachment + gill-spacing + gill-color	0.9215

As can be seen from Table 3 shows that all methods yield similar results in terms of the best informativeness estimates. Specifically, the “stem-surface” and “stem-color” features were ranked in the top three by all informativeness assessment methods, while the “cap-color” feature was ranked in the top three by the nonparametric method and the Gini coefficient.

In Tables 4, 5, and 6 the informativeness of various groups of features is calculated using the proposed nonparametric estimate $V(x_i)$. The most informative combinations include the features “cap-surface”, “cap-color”, “gill-attachment”, and “gill-color”. Analysis of the estimates confirms the properties of estimate (14) and the importance of considering interactions between features. The indicators demonstrate the significance of this consideration, which outweighs individual estimates, which is clearly evident for features with low informativeness.

Note: Tables 4, 5, and 6 illustrate the method’s potential for revealing hidden structures in data and constructing compact yet expressive feature models. This is particularly relevant for medical data classification tasks, where accounting for group interactions between parameters plays a critical role.

Experiment 2. For the analysis, the Student Performance in Mathematics and Portuguese dataset was taken from <https://www.kaggle.com>. This dataset contains student data for grade prediction (395 objects, 30 features, 5 classes).

Table 7 – Example of source data for calculating informativeness

school	sex	age	address	famsize	Pstatus	Medu	Fedu	Mjob	Fjob
GP	F	18	U	GT3	A	4	4	at home	teacher
GP	F	17	U	GT3	T	1	1	at home	other
GP	F	15	U	LE3	T	1	1	at home	other
GP	F	15	U	GT3	T	4	2	health	services
GP	F	16	U	GT3	T	3	3	other	other
GP	M	16	U	LE3	T	4	3	services	other
GP	M	16	U	LE3	T	2	2	other	other
GP	F	17	U	GT3	A	4	4	other	teacher
GP	M	15	U	LE3	A	3	2	services	other
GP	M	15	U	GT3	T	3	4	other	other

Table 8 presents the results of calculating the informativeness of features using the four methods discussed above.

Table 8 – Individual informativeness of features (the 3 best values are in bold, 1 in green, 2 in yellow, and 3 in red)

feature	IG	Gini	χ^2	V
school	0.0048	0.0008	2.0071	0.2202
sex	0.0054	0.0009	1.5431	0.2397
age	0.0501	0.012	0.9242	0.2876
address	0.0136	0.0039	1.6482	0.2336
famsize	0.0058	0.0009	2.1055	0.2308
Pstatus	0.0082	0.0014	0.4969	0.2314
Medu	0.068	0.0167	10.1093	0.2884
Fedu	0.0374	0.0088	5.9579	0.2705
Mjob	0.0534	0.0115	4.9202	0.3171
Fjob	0.0319	0.0068	1.1683	0.2621
reason	0.0176	0.0034	8.1294	0.2506
guardian	0.008	0.0018	1.0796	0.2185
traveltime	0.0136	0.0035	1.7932	0.2402
studytime	0.0305	0.0076	1.8293	0.256
failures	0.1058	0.0227	87.7225	0.2994
schoolsup	0.0521	0.0139	22.3592	0.2457
famsup	0.0053	0.0014	1.1126	0.2257
paid	0.0198	0.0034	5.58	0.2529
activities	0.0061	0.0014	1.6314	0.2341
nursery	0.0131	0.0018	1.174	0.2429
higher	0.0271	0.0054	0.7496	0.2308
internet	0.0102	0.0035	0.8713	0.2205
romantic	0.0286	0.003	9.5915	0.291
famrel	0.0116	0.0028	0.3601	0.245
freetime	0.0263	0.0049	0.3565	0.2775
goout	0.0454	0.0104	5.7071	0.2732
Dalc	0.0315	0.0063	3.0375	0.2582
Walc	0.0509	0.0114	10.4248	0.2996
health	0.0359	0.0106	2.7174	0.2681
absences	0.0163	0.0039	0.3917	0.2191

Calculating the informativeness of feature groups using formula (9) yielded the following results:

Table 9 – Informativeness of 2-feature groups using the nonparametric method (top 10)

Groups of features	V
Mjob + Walc	0.4459
Mjob + health	0.4073
failures + Walc	0.4034
Mjob + goout	0.4024
Medu + Walc	0.3996
age + goout	0.3922
age + Walc	0.3889
age + Mjob	0.3886
Mjob + failures	0.3876
freetime + goout	0.3865

Table 10 – Informativeness of groups of 3 features using a nonparametric method (top 10)

Groups of features	V
age + Mjob + Walc	0.5954
Mjob + Walc + health	0.5932
age + goout + health	0.5875
Mjob + goout + Walc	0.5827
age + Mjob + health	0.5824
Mjob + goout + health	0.581
Mjob + reason + Walc	0.5783
age + Mjob + goout	0.5753
age + freetime + goout	0.5739
Mjob + famrel + Walc	0.5725

Table 11 – Informativeness of groups of 4 features using the nonparametric method (top 10)

Groups of features	V
age + Mjob + goout + health	0.8235
age + Mjob + Walc + health	0.8079
age + Fedu + Mjob + goout	0.8001
age + Mjob + freetime + health	0.799
Mjob + goout + Walc + health	0.7977
age + Mjob + reason + Walc	0.7961
age + Medu + goout + health	0.7953
Mjob + freetime + Walc + health	0.7888
age + Mjob + reason + health	0.7884
Mjob + reason + Walc + health	0.7867

5 DISCUSSION

As can be seen from Table 8, all methods yield similar results in terms of the best informativeness estimates. Specifically, the “failures” feature was ranked in the top 3 by all informativeness assessment methods, the “Mjob” feature was ranked in the top 3 by the nonparametric method and information gain, and the “Walc” feature was ranked in the top 3 by the nonparametric method and the chi-square test.

In Tables 9, 10, and 11, the information content of various groups of features is calculated using the proposed nonparametric estimate $V(x_i)$. The most informative combinations include the features “age”, “Mjob”, “Walk”, “health”, and “goout”.

The results of the experiment confirm the feasibility of using various methods for assessing the informativeness of features in classification problems. A comparative analysis between classical statistical approaches and the proposed nonparametric method deserves special attention.

First, high agreement is observed between the information gain, chi-square, and Gini coefficient methods in assessing key features. This confirms the

robustness of the selected features under various criteria, increasing confidence in their selection.

Second, the proposed nonparametric assessment demonstrates an advantage in accounting for interactions between features, identifying groups with high joint significance whose informativeness significantly exceeds that of individual indicators.

This clearly demonstrates the ability of the nonparametric method to reveal hidden relationships between features, which is impossible with individual assessment.

Thus, the nonparametric approach can be considered not simply as an alternative, but as an extension of existing selection methods, allowing for the formation of more compact and expressive features, especially in conditions of high dimensionality and limited interpretability of individual parameters. It’s important to note that if a feature shows strong informativeness using only one method but low informativeness using the others, it’s worth checking: it may be sensitive to a specific statistical characteristic (for example, the chi-square test responds well to frequency skew). Furthermore, if estimates from different methods are consistent, this means that the choice of informative features will be robust to the estimation method—always a good practice.

To automate the calculation of feature informativeness, including for more complex cases where the TS contains k -valued features and the number of classes is greater than 2, the “TS Quality” web service was developed. It is located at:

<https://www.sciencehunter.net/Services/>.

This calculation helps us evaluate the quality of the entire set of TS features in terms of their separating ability, which is especially valuable when there are a large number of features.

Calculations of the informativeness of features were also carried out using the methods of IG, chi-square, Gini, and the proposed method on several samples taken from [21] with a different number of features (from 4 to 26), different types (binary, nominal, ordinal, heterogeneous), with a different number of objects (from 400 to 60.000) and with a different number of classes. For each feature set and dataset, a grid search was used to select optimal hyperparameters, followed by cross-validation. As a result, it was possible to identify the most informative features, the optimality of which was confirmed by the aforementioned selection procedure.

CONCLUSIONS

Evaluating the informativeness of features is a crucial step in constructing effective classification models and data-driven decision making. This study examined classical estimation methods – IG, χ^2 , and Gini – and proposed an original nonparametric approach with several advantages.

The proposed method for assessing the informativeness of features demonstrates high efficiency and interpretability. It enables the identification of both

individually significant features and the assessment of the information content of groups of features, taking into account interactions between them, identifying the most significant ones. This enhances the interpretability and robustness of the model. The results demonstrate the practical applicability of the proposed method for optimizing feature space, especially in high-dimensional settings.

The novelty and uniqueness of this work lies in the fact that the proposed nonparametric method identifies feature interactions.

Correlation analysis confirms the consistency of the estimates, allows for the visualization of structural dependencies, improves the analytical value of the method, and offers a basis for meaningful feature selection.

The developed web service automates calculations and makes the method accessible to a wide audience, including practitioners without ML training. This expands the applicability of the approach in medicine, industry, education, and other fields.

Thus, the combination of theoretical justification, visual analysis, and practical implementation allows us to develop a complete and scalable informativeness assessment system suitable for a wide range of tasks—from bioinformatics to industrial information systems.

All methods agree on the most informative features.

An important property of the proposed approach is the ability to normalize the informativeness assessment and its scalability for problems with an arbitrary number of classes, values, and objects. The proposed formalization allows not only for feature selection but also for using informativeness as a quality metric for the training set itself.

Further development of the proposed approach opens up broad possibilities in both methodological and applied aspects. The following promising areas can be highlighted:

– Generalization of the assessment to fuzzy and interval features, which will allow the method to be applied in situations where features are expressed in continuous rather than discrete form (e.g., medical indicators with uncertainty).

– Integration of the method into AutoML platforms will enable automatic feature selection based on their group informativeness, improving the quality of models without analyst intervention.

– Application in feature selection assessment for neural network models, particularly in explainable AI problems, where it is important to understand which features and their combinations influenced the classification.

– Development of visual tools that allow interactive exploration of the contribution of features and their groups to the final model.

– Expanding the application to multi-class problems and problems with asymmetric class distributions, where the role of group interactions increases.

– Construction of hybrid selection algorithms that combine statistical and heuristic approaches.

Thus, the proposed approach opens the way to building more accurate, interpretable, and efficient machine learning models, and can also serve as a basis for assessing the quality of the training set itself—an important aspect often overlooked in standard data analysis procedures. The use of group informativeness metrics allows for a more meaningful, analytical, rather than formal, approach to sample selection.

ACKNOWLEDGEMENTS

The work is supported by the Noosphere Association, Dnipro, Ukraine.

DECLARATIONS

Conflict of interest: The authors declare that they have no conflict of interest in relation to this research, whether financial, personal, authorship, or otherwise, that could affect the research and its results presented in this paper.

Authors' contributions: Gennadij Shevchenko: the method for calculating of the informativeness of an arbitrary group of features of TS; Stanislav Gerasimenko: experimental study of this method.

Data availability: The manuscript has associated data in a data repository <https://www.kaggle.com/>.

Software availability: The manuscript has no associated software.

Use of artificial intelligence tools: The authors confirm that they did not use artificial intelligence technologies in creating the submitted work.

REFERENCES

1. Guyon I., Elisseeff A. An introduction to variable and feature selection. *Journal of Machine Learning Research*, 2003, Vol. 3, pp. 1157–1182.
2. Theng D., Bhoyar K. K. Feature selection techniques for machine learning: a survey of more than two decades of research [Electronic resource]. *Knowledge and Information Systems*, 2023, Vol. 66, No. 3, pp. 1575–1637. DOI: 10.1007/s10115-023-02010-5.
3. Sammut C., Webb G. *Encyclopedia of Machine Learning*. New York, Springer Science+Business Media, 2010, 1059 p.
4. Pearson K. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling [Electronic resource]. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 1900, Vol. 50, No. 302, pp. 157–175. DOI: 10.1080/14786440009463897.
5. Hastie T., Tibshirani R., Friedman J. *The Elements of Statistical Learning*. 2nd ed. Berlin, Springer, 2009, 764 p.
6. Quinlan J. Induction of decision trees. *Machine Learning*, 1986, Vol. 1, No. 1, pp. 81–106.

7. Shannon C. E., Weaver W. The Mathematical Theory of Communication. Urbana, University of Illinois Press, 1949, 117 p.
8. Kira K., Rendell L. A practical approach to feature selection. *Proceedings of the International Conference on Machine Learning*, ed. by D. Sleeman, P. Edwards. Aberdeen, 1992, pp. 368–377.
9. Gretton A. et al. Kernel methods for measuring independence. *Journal of Machine Learning Research*, 2005, Vol. 6, pp. 2075–2129.
10. Kursa M. B., Rudnicki W. R. Feature selection with the Boruta package. *Journal of Statistical Software*, 2010, Vol. 36, No. 11, pp. 1–13.
11. Breiman L. et al. Classification and Regression Trees. Boca Raton, CRC Press, 1984, 366 p.
12. Spiegel M. Statistics, 4th ed. New York, McGraw-Hill, 2009, (Schaum's Outline Series), 577 p.
13. Agresti A. Introduction to Categorical Data Analysis. New York, John Wiley & Sons, 2007, 400 p.
14. Adasovsky B. I. Method for calculating the information content of multimodal features. *Doklady Akademii Nauk SSSR*, 1978, Vol. 239, No. 2, pp. 286–289.
15. Kullback S. Information Theory and Statistics. New York, John Wiley & Sons, 1959, 399 p.
16. Kendall M., Stewart A.; trans. from Engl. by L. I. Galchuk, A. T. Terehin. Statistical Inferences and Relationships. Moscow, Nauka, 1973, 900 p.
17. Rokach L., Maimon O. Data Mining Using Decision Trees: Theory and Application. Singapore, World Scientific Publishing, 2008, 267 p.
18. Vasilenko Yu. A., Shevchenko G. Ya. Analytical method for finding tests. *Avtomatika*, 1979, No. 2, pp. 3–8.
19. Jolliffe I. T. Principal Component Analysis. New York, Springer, 2002, 487 p.
20. Kim D. O., Mueller C. W., Klekka W. R. ; trans. from Engl. by A. M. Hotinskiy, S. B. Korolev Factor, Discriminant, and Cluster Analysis. Moscow, Finance and Statistics, 1989, 215 p.
21. UC Irvine Machine Learning Repository. Welcome to the UC Irvine Machine Learning Repository [Electronic resource]. *UC Irvine Machine Learning Repository*. Mode of access: <https://archive.ics.uci.edu/> (date of access: 03.12.2025). Title from screen.

Received 21.10.2025.

Accepted 07.04.2026.

Published 26.06.2026.

УДК 681.3

ОЦІНКА ІНФОРМАТИВНОСТІ ДИСКРЕТНИХ ОЗНАК НАВЧАЛЬНОЇ ВИБІРКИ В ЗАДАЧАХ КЛАСИФІКАЦІЇ

Шевченко Г. Я. – канд. техн. наук, завідувач відділу, Асоціація Ноосфера, Дніпро, Україна.
ORCID: <https://orcid.org/0000-0003-3984-9266>.

Герасименко С. М. – науковий співробітник, Асоціація Ноосфера, Дніпро, Україна Анотація.
ORCID: <https://orcid.org/0009-0009-2597-9917>.

АНОТАЦІЯ

Актуальність. Сучасні методи інтелектуального аналізу даних (Data Mining) широко застосовуються для побудови класифікаційних та прогностичних моделей. Однак при обробці дискретних даних зберігається проблема кількісної оцінки інформативності ознак, що визначає точність та стійкість класифікаторів. Відсутність універсального підходу до вимірювання вкладу як окремих ознак, так і груп ознак результат класифікації ускладнює процес автоматизованого відбору ознак і оптимізації моделей.

Мета роботи. Метою роботи є розробка та теоретичне обґрунтування методу оцінки інформативності як окремих ознак, так і довільних груп дискретних ознак, що ґрунтується на взаємозв'язку статистичних характеристик ознак із заходами помітності класів.

Метод. Запропоновано підхід, що пов'язує інформативність як окремих дискретних ознак, так і довільних груп таких ознак щодо функції, що характеризує цільову змінну. Метод базується на результатах досліджень відомих вчених Кендалла і Стюарта в галузі задач непараметричної статистики. Для практичного застосування вводиться алгоритм обчислення інформативностей як окремих ознак, так і груп ознак, що допускає реалізацію в програмних системах аналізу даних.

Результати. Показано, що розроблений метод дозволяє формально та кількісно оцінювати внесок як кожної ознаки, так і довільних груп ознак у процес класифікації без попередніх припущень про тип моделі, а також дає змогу виявити приховані залежності між ознаками, що неможливо при індивідуальній оцінці. мова йде про виявлення взаємодії ознак.

Практичні розрахунки підтверджують, що інформативність, обчислена через імовірнісні характеристики, корелює з реальною значущістю ознак при побудові класифікаторів на основі дерев рішень та байєсівських мереж. Отримані вирази забезпечують основу для автоматизації відбору ознак під час роботи з дискретними даними, покращують аналітичну цінність методу та пропонують основу для осмисленого відбору ознак

Висновки. Запропонований підхід дозволяє уніфікувати процедуру оцінки інформативності як окремих дискретних ознак, і довільних груп ознак у системах Data Mining. Він забезпечує формальний зв'язок між статистичними характеристиками даних та якістю класифікації, що сприяє підвищенню точності, стійкості та інтерпретованості моделей.

КЛЮЧОВІ СЛОВА: інформативність ознак, дискретні дані, ентропія, Data Mining, класифікація, вибір ознак, взаємодія ознак.

ЛІТЕРАТУРА

1. Guyon I. An introduction to variable and feature selection / I. Guyon, A. Elisseeff // *Journal of Machine Learning Research*. – 2003. – Vol. 3. – P. 1157–1182.
2. Theng D. Feature selection techniques for machine learning: a survey of more than two decades of research [Electronic resource] / D. Theng, K. K. Bhoyar // *Knowledge and Information Systems*. – 2023. – Vol. 66, No. 3. – P. 1575–1637. – DOI: 10.1007/s10115-023-02010-5.
3. Sammut C. *Encyclopedia of Machine Learning* / C. Sammut, G. Webb. – New York : Springer Science+Business Media, 2010. – 1059 p.
4. Pearson K. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling [Electronic resource] / K. Pearson // *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*. – 1900. – Vol. 50, no. 302. – P. 157–175. – DOI: 10.1080/14786440009463897.
5. Hastie T. *The Elements of Statistical Learning* / T. Hastie, R. Tibshirani, J. Friedman. – 2nd ed. – Berlin : Springer, 2009. – 764 p.
6. Quinlan J. Induction of decision trees / J. Quinlan // *Machine Learning*. – 1986. – Vol. 1, No. 1. – P. 81–106.
7. Shannon C. E. *The Mathematical Theory of Communication* / C. E. Shannon, W. Weaver. – Urbana : University of Illinois Press, 1949. – 117 p.
8. Kira K. A practical approach to feature selection / K. Kira, L. Rendell // *Proceedings of the International Conference on Machine Learning* / ed. by D. Sleeman, P. Edwards. – Aberdeen, 1992. – P. 368–377.
9. Kernel methods for measuring independence / A. Gretton et al. // *Journal of Machine Learning Research*. – 2005. – Vol. 6. – P. 2075–2129.
10. Kursa M. B. Feature selection with the Boruta package / M. B. Kursa, W. R. Rudnicki // *Journal of Statistical Software*. – 2010. – Vol. 36, No. 11. – P. 1–13.
11. *Classification and Regression Trees* / L. Breiman [et al.]. – Boca Raton : CRC Press, 1984. – 366 p.
12. Spiegel M. *Statistics* / M. Spiegel. – 4th ed. – New York : McGraw-Hill, 2009. – (Schaum's Outline Series). – 577 p.
13. Agresti A. *Introduction to Categorical Data Analysis* / A. Agresti. – New York : John Wiley & Sons, 2007. – 400 p.
14. Adasovsky B. I. Method for calculating the information content of multimodal features / B. I. Adasovsky // *Doklady Akademii Nauk SSSR*. – 1978. – Vol. 239, No. 2. – P. 286–289.
15. Kullback S. *Information Theory and Statistics* / S. Kullback. – New York : John Wiley & Sons, 1959. – 399 p.
16. Kendall M. *Statistical Inferences and Relationships* / M. Kendall, A. Stewart ; trans. from Engl. by L. I. Galchuk, A. T. Terehin. – Moscow : Nauka, 1973. – 900 p.
17. Rokach L. *Data Mining Using Decision Trees: Theory and Application* / L. Rokach, O. Maimon. – Singapore : World Scientific Publishing, 2008. – 267 p.
18. Vasilenko Yu. A. Analytical method for finding tests / Yu. A. Vasilenko, G. Ya. Shevchenko // *Avtomatika*. – 1979. – No. 2. – P. 3–8.
19. Jolliffe I. T. *Principal Component Analysis* / I. T. Jolliffe. – New York : Springer, 2002. – 487 p.
20. Kim D. O. *Factor, Discriminant, and Cluster Analysis* / D. O. Kim, C. W. Mueller, W. R. Klekka ; trans. from Engl. by A.M. Hotinskiy, S. B. Korolev. – Moscow : Finance and Statistics, 1989. – 215 p.
21. UC Irvine Machine Learning Repository. Welcome to the UC Irvine Machine Learning Repository [Electronic resource] // UC Irvine Machine Learning Repository. – Mode of access: <https://archive.ics.uci.edu/> (date of access: 03.12.2025). – Title from screen.