IRSTI 27.41.23 DOI: https://doi.org/10.26577/JMMCS2025125104

A.Kh. Nishanov[1*], **A.T. Tursunov**[2], **F.F. Ollamberganov**[1], **D.E. Rashidova**[1]

[1]Tashkent university of information technologies named after Muhammad al-Khwarizmi, Tashkent, Uzbekistan

[2]Tashkent pharmaceutical institute, Tashkent, Uzbekistan

[*]e-mail: nishanov_akram@mail.ru

ALGORITHM FOR CLUSTERING DIFFERENT TYPES OF DRUGS AFFECTING BLOOD PRESSURE

This article presents the development of an algorithm and software for grouping different types of drugs that affect blood pressure in humans. The results are experimentally tested on a cross-section of more than 1,100 drugs that affect human blood pressure. Using the proposed algorithm, the training sample is formed and the problem of clustering is solved. The training set consists of ten classes. Selection symbols are given in different types, they consist of nominal and value symbols. In the article, each object is examined, and the importance of the object in the sample is assessed using a criterion. This criterion contributes to the formation of the studied class of the object. The developed algorithm works taking into account both types of features. If the similarity of the object under study with any class is high, this object is transferred to this class. This process is performed sequentially several times for all objects of the class. The process stops when the position of objects remains unchanged and the degree of similarity exceeds the required percentage. The accuracy of data set classification by object classes was experimentally verified using an algorithm and software package based on neural networks.

Key words: Drugs, pattern recognition, symbols, clustering issue, algorithm and software.

А.Х. Нишанов[1*], А.Т. Турсунов[2], Ф.Ф. Олламберганов[1], Д.Е.Рашидова[1]

[1]Мұхаммед әл-Хорезми атындағы Ташкент ақпараттық технологиялар университеті, Ташкент, Өзбекстан

[2]Ташкент фармацевтикалық институты, Ташкент, Өзбекстан

[*]e-mail: nishanov_akram@mail.ru

Қан қысымына әсер ететін дәрілердің түрлі түрлерін кластерлеу алгоритмі

Мақалада адамдардың қан қысымына әсер ететін дәрілердің әртүрлі түрлерін топтастыруға арналған алгоритм мен бағдарламалық пакетті жасау зерттеледі. Нәтижелер адамның қан қысымына әсер ететін 1100-ден астам дәрі-дәрмектің көлденең қимасында эксперименталды түрде тексерілді. Ұсынылған алгоритмді пайдалана отырып, кластерлеу мәселесі шешіліп, оқыту үлгісі қалыптастырылды. Білім беру іріктеу он сыныптан тұрады. Таңдау белгілері әртүрлі түрде беріледі, олар номиналды және мәндік белгілерден тұрады. Мақалада әрбір объект зерттеліп, іріктемедегі объектінің маңыздылығы критерий арқылы бағаланады. Бұл критерий объектінің зерттелетін класын қалыптастыруға ықпал етеді. Әзірленген алгоритм белгілердің екі түрін де ескере отырып жұмыс істейді. Егер зерттелетін объектінің қандай да бір класқа ұқсастығы жоғары болса, онда бұл объект осы класқа көшіріледі. Бұл процесс кластың барлық объектілері үшін кезекпен бірнеше рет орындалады. Нысандардың орналасуы өзгермей, ұқсастық дәрежесі қажетті пайыздан асқан кезде үдеріс тоқтатылады. Оқыту объектілерінің дұрыс жіктелуі нейрондық желіге негізделген алгоритм және бағдарламалық пакетті қолдану арқылы эксперименталды түрде тексеріледі.

Түйін сөздер: дәрілер, үлгіні тану, белгілер, кластерлеу мәселесі, алгоритм және бағдарламалық қамтамасыз ету.

А.Х. Нишанов[1*], А.Т. Турсунов[2], Ф.Ф. Олламберганов[1], Д.Э.Рашидова[1]
[1]Ташкентский университет информационных технологий имени Мухаммада аль-Хорезми, Ташкент, Узбекистан
[2]Ташкентский фармацевтический институт, Ташкент, Узбекистан
[*]e-mail: nishanov_akram@mail.ru

Алгоритм кластеризации различных видов препаратов, влияющих на артериальное давление

В статье исследуется разработка алгоритма и программного комплекса для группировки лекарственных средств, выраженных в различных типах, влияющих на кровяное давление человека. Результаты были апробированы на более чем 1100 препаратах, влияющих на кровяное давление человека. С помощью предложенного алгоритма решена задача кластеризации и сформирована наборы данных. Наборы данных состоит из десяти классов. Признаки выборки были представлены различными типами, состоящими из номинальных и стоимостных признаков. В статье исследуется каждый объект, и важность объекта в выборке оценивается с помощью критерия. Этот критерий способствует формированию исследуемого класса объекта. Разработанный алгоритм работает с учетом обоих типов признаков. Если сходство исследуемого объекта с каким-либо классом высокое, этот объект переводится в этот класс. Этот процесс выполняется последовательно несколько раз для всех объектов класса. Процесс останавливается, когда положение объектов остается неизменным и степень сходства превышает требуемый процент. Правильность определения классами объектов наборы данных была экспериментально проверена с использованием алгоритма и программного комплекса, основанного на нейронных сетях.

Ключевые слова: Лекарственные препараты, распознавание образов, символы, проблема кластеризации, алгоритм и программное обеспечение.

1 Introduction

Clustering different types of antihypertensive drugs involves classifying them based on similarities in their mechanisms of action, chemical structures, or therapeutic effects. This task is critical for medical professionals to optimize treatment plans, improve patient understanding, and facilitate research. The necessary technologies and their analysis in the construction of algorithms and software for clustering drugs affecting blood pressure can be found in [1-6]. Research works [12-19] serve as the main sources of data clustering methods, algorithms and software used in medicine.

Medicines are divided into groups based on how they affect blood pressure.

Angiotensin-converting enzyme inhibitors (Inhibitory angiotensin converting enzyme (IACE)). Angiotensin-converting enzyme inhibitors (IACE) are a class of drugs used primarily to treat high blood pressure (hypertension), treat heart failure, prevent stroke, and protect the kidneys in people with diabetes [7].

Angiotensin-II receptor blockers (BRA). Angiotensin-**II** receptor blockers (ARB) are a class of drugs used to treat high blood pressure (hypertension), heart failure, diabetic nephropathy, and certain other conditions. They work by blocking the effects of angiotensin-**II**, a hormone that causes blood vessels to constrict and increase blood pressure. By blocking these receptors, ARB help relax blood vessels, thereby lowering blood pressure and reducing the workload on the heart.

Beta-blockers (BB). (*β*-adrenoblocker, BB). Beta-blockers (BB) are a class of drugs used primarily to treat arrhythmias, prevent recurrent myocardial infarction, treat hypertension, heart failure, certain types of tremors, migraines, and glaucoma. These drugs can be used to treat high blood pressure, certain types of heart arrhythmias, and migraines.

Calcium channel blockers (CCB). Calcium channel blockers (CCBs) are a class of drugs used to treat a variety of cardiovascular conditions, including hypertension (high blood pressure), angina (chest pain due to reduced blood flow to the heart muscle), and some forms of arrhythmia [9].

Diuretics: Diuretics are drugs that help to remove more fluids from the body. They work by affecting the work of the kidneys, resulting in increased urine production and excretion. Diuretics are used to treat various diseases, including heart failure, high blood pressure, swelling (edema), and others. Diuretics help control the amount of fluid in the body and can reduce swelling and pressure in the blood vessels [8-9].

Peripheral vasodilators. Peripheral vasodilators are a group of drugs that widen peripheral blood vessels, reduce the resistance of the vessel walls, and thereby lower blood pressure.

Peripheral vasodilators are used in combination with other medications to better control blood pressure and overall cardiovascular health. They can cause side effects such as dizziness, swelling, facial flushing, headache, and reflex tachycardia, so their use should be under strict medical supervision.

Selective sinus node If-channel inhibitors. Selective sinus node If-channel inhibitors are a group of drugs often used to control sinus rhythm. If channels, also called "phase inhibitors play an important role in the control of electrical impulses in cardiac myocytes. These channels are more common in myocytes in the sinus node, and by inhibiting them, sinus rhythm can be slowed, which reduces cardiac output and may be useful in controlling arrhythmias.

If-channel inhibitors can be used to ease the heart's workload in certain heart conditions, including high blood pressure, heart failure, and atrial fibrillation. One of the famous representatives of these drugs is ivabradine. Ivabradine targets only the If channels, which makes it safe and effective for heart rhythm control because it does not affect other cardiac channels and therefore has few side effects.

If-channel inhibitors work primarily by slowing sinus rhythm and reducing the pressure the heart exerts on the circulatory system. One of the advantages of these drugs is their ability to lower the heart rate, which is especially useful for patients with respiratory problems.

Medicines that improve myocardial metabolism. Medicines used to improve the metabolism of the myocardium are often used to improve the blood circulation of the heart and increase the performance of the heart muscle. Such drugs are mainly prescribed for cardiovascular diseases, such as ischemic heart disease, angina pectoris and myocardial infarction.

Antiplatelet agents. Antiplatelet agents are a group of drugs used to prevent or reduce the clumping of blood platelets. They are mainly used in the prevention and treatment of thrombosis, myocardial infarction, stroke and other vascular diseases. Antiplatelet drugs inhibit the activity of blood platelets and prevent excessive blood clotting and the formation of blood clots [10-11].

2 Clustering of given objects using symbols of different types

Let's assume that in the space of N-dimensional nominal and value symbols, given drugs of various types $x_i \in X$, $i = \overline{1, M}$, i.e., objects. So, $x_i = (x_i^1,\ x_i^2,\ \ldots,\ x_i^N)$, $i = \overline{1, M}$, drug is

given in the space of N-dimensional nominal and value characters. And all of them are X collection objects. We call it the common sample objects, and the set to which they belong is denoted by X.

Task. It is required to form educational selections on the basis of the given general selection. That is, it is required to form classes expressed in the following form $x_{p1}, x_{p2}, \ldots, x_{pm_p} \in X_p$, $p=\overline{1,r}$. Where x_{pi} is read as the i- object of class X_p in the N-dimensional nominal and valued symbol space, and it is written in the following form in the N-dimensional nominal and valued symbol space $x_{pi}=\left(x_{pi}^1,\ x_{pi}^2,\ \ldots x_{pi}^N\right)$, $i=\overline{1,m_p}$ viewed in the space of N-dimensional nominal and value symbols, $X=\bigcup_{p=1}^{r} X_p$ consisting of m_p objects $x_{p1}, \ldots, x_{p_{m_p}}$) in class X_p . In [7-14], this issue is referred to as the object clustering issue.

Let's introduce the following dimensions and designations

Let the quantity indicating the similarity of objects in the space of nominal symbols be determined by $\rho^j(x_{pi},x_{pq})$ and calculated by (1), i.e.

$$\rho_{pi}^{j}\left(x_{pi},x_{pq}\right)=\begin{cases} 1, \text{ if } \left(x_{pi}^{j}-x_{pq}^{j}\right)=0; \\ \\ 0, \qquad\quad \text{otherwise} \end{cases} \tag{1}$$

Let the quantity indicating the similarity of objects in the space of numerical symbols be determined by $\rho^j(x_{pi},x_{pq})$ and calculated by (2), i.e.

$$\rho_{pi}^{j}\left(x_{pi},x_{pq}\right)=\begin{cases} 1, \text{ if } \quad |x_{pi}^{j}-x_{pq}^{j}|\leq\varepsilon^{j} \\ \\ 0, \qquad\quad \text{otherwise} \end{cases} \tag{2}$$

where $p=\overline{1,r}$; $i\neq q=\overline{1,m_p}; j=\overline{1,N}$; The expressed quantities (1) and (2) are the parameters of the vector, which is expressed in the following form $\rho_{pi}\left(x_{pi},x_{pq}\right)=\left(\rho_{pi}^{1}\left(x_{pi},x_{pq}\right),\rho_{pi}^{2}\left(x_{pi},x_{pq}\right),\ldots,\rho_{pi}^{N}\left(x_{pi},x_{pq}\right)\right)$. So, if the considered j-symbol is nominal, the j-symbol of the vector $\rho_{pi}\left(x_{pi},x_{pq}\right)$ is calculated using (1), otherwise, that is, if the symbol is numerical, then this symbol of the vector is calculated by (2) . In this case, the ε^j- threshold values corresponding to the j-character are performed by the following formula for each character of the class objects:

$$\varepsilon^{j}=\frac{1}{M-1}\sum_{i=1}^{M-1}\left|x_{pi}^{j}-x_{pi+1}^{j}\right|,$$

where $j=\overline{1,N}; p=\overline{1,r}; i=\overline{1,M-1};$.

We present the following steps for solving the above-mentioned clustering problem:

1. Drug data is preprocessed into $x_i=(x_i^1,\ x_i^2,\ \ldots,\ x_i^N)\in X,\ i=\overline{1,M}$ In this case, missing data are filled in, anomalous data are replaced by the mean value, and quantitative signs are normalized. Standardization is carried out for all quantitative signs based on the following formula:

$$x_i^j=\frac{x_i^j-\min_i x_i^j}{\max_i x_i^j\ -\min_i x_i^j}.\ i=\overline{1,M};\ j=\overline{1,N};$$

2. Then, drugs $x_i \in X$, $i = \overline{1,M}$, objects are arbitrarily divided into r classes. That is,arbitrary are divided into x_{p1}, x_{p2}, ..., $x_{pm_p} \in X_p$, $p=\overline{1,r}$ classes;

3. Based on the equations (1) and (2) given above, all $p = \overline{1,r}$; $i \neq q = \overline{1,m_p}$; $j = \overline{1,N}$ for all parameters of the vector$\rho_{pi}(x_{pi},x_{pq})$. That is $\rho_{pi}(x_{pi},x_{pq}) = (\rho_{pi}^{1}(x_{pi},x_{pq}), \rho_{pi}^{2}(x_{pi},x_{pq}), \ldots, \rho_{pi}^{N}(x_{pi},x_{pq})$ vector symbols are calculated for $p = \overline{1,r}$; $i \neq q = \overline{1,m_p}$; $j = \overline{1,N}$;

4. The position of the i- object in the optional p-class in the remaining set of m_p-1 objects of this class is evaluated as follows:

$$\Gamma_{pi}(x_{pi},X_p) = \frac{1}{m_p-1}\sum_{q=1}^{m_p-1}\sum_{j=1}^{N}\rho^{j}(x_{pi},x_{pq}),\ p=\overline{1,r}\ ;i=\overline{1,m_p};i\neq q.$$

5. The general grade of the arbitrary p-class is calculated based on the criterion $\Gamma_p(X_p) = \frac{1}{m_p}\sum_{i=1}^{m_p}\Gamma_{pi}(x_{pi},X_p)$, $p=\overline{1,r}$. The degree of similarity of their objects is evaluated as follows

$$\nu_p(X_p) = = \frac{\Gamma_p(X_p) \quad *100\%}{N},\ p=\overline{1,r}.$$

6. Also, the i -object in an arbitrary $p-$class is evaluated by other X_q, $q=\overline{1,r-1}$ class objects: $\Gamma_{pi}(x_{pi},X_q) = \frac{1}{m_q}\sum_{k=1}^{m_q}\sum_{j=1}^{N}\rho^{j}(x_{pi},x_{qk}), p=\overline{1,r}\ ;i=\overline{1,m_p};i\neq q$. And the i-object in the p-class is transferred to the class which is highly valued by the objects of the class. If the highest marks are tied, they are kept in their own class. If the upper values are equal in two classes other than in the p-class, then this object is moved to the smaller q index.

7. Usually, the degree of similarity of objects in the formed classes is required to be

$$\nu_{\mathrm{p}}(\mathrm{X_p}) \geq \delta \geq 55\%.$$

8. These calculations are performed for all objects x_{pi}, $p=\overline{1,r}$;$i=\overline{1,m_p}$; and the resulting new x_{p1}, x_{p2}, ..., $x_{pm_p} \in X_p$, $p=\overline{1,r}$ classes gives a true clustered training sample.

Based on these clustering steps, 1116 different types of blood pressure drugs were clustered using the algorithm and software. One nominal character and nine quantitative characters were studied. They are given in the table below:

Table 1

Information about drugs that affect blood pressure (antihypertensive)											
№	The name of the drug	Percentage of active substance in mg	Melting point %	Color	Form of production	Blood concentration (max)	Distribution density	**Technological indicators of prepared compositions**			
								Spreadability, 10^{-3} kg/s	Spreading density, kg/m^3	Metabolism in the liver %	Decomposition
1	x_1	0,10	96	yellow	10	64	3,50	4,51	20,72	90	7,41
2	x_2	0,10	85	white	20	78	4,20	7,85	14,36	75	8,56
3	x_3	0,50	88	pale yellow	25	20	4	3,87	25,48	95	5,69
4	x_4	0,03	89	white	50	72	0,20	16,87	15,64	75	7,96
5	x_5	0,20	91	white	15	60	3,20	19,20	25,63	85	5,98
6	x_6	0,50	93	white	25	57	3,90	4,89	48,65	89	6,85
7	x_7	0,20	85	white	10	105	4,24	3,56	13,56	74	6,47
8	x_8	16	95	yellow	50	29	4,80	7,86	14,65	96	4,12
1111	x_{1111}	0,4	75	pink	30	96	3,5	6,52	12,69	85	7,89
1112	x_{1112}	0,2	78	gray	20	97	3,6	10,42	12,25	80	8,54
1113	x_{1113}	0,5	76	white	25	86	4,01	7,41	13,25	78	9,65
1114	x_{1114}	0,25	74	gray	14	93	4,5	3,41	10,26	60	12,48
1115	x_{1115}	0,5	72	white	28	94	3,6	4,85	12,48	85	9,85
1116	x_{1116}	0,25	76	white	30	85	3,9	4,63	7,96	86	8,69

The following table provides information about the digitization of the nominal values of drugs, that is, their colors:

Table 2

brown	1
blue	2
gray	3
red	4
black	5
white	6
fire color	7
pale red	8
pale pink	9
pale yellow	10
pink	11
yellow	12

3 Algorithm for clustering given objects in different types of character space

The algorithm for finding a solution to the problem described in the article is presented below. The algorithm consists of six clauses, and it is appropriate to apply the problem of symbol determination only to objects of a class taken separately.

First step. Drug data is preprocessed into $x_i = (x_i^1,\ x_i^2,\ \ldots,\ x_i^N) \in X,\ i = \overline{1, M}$. In this case, missing data are filled in, anomalous data are replaced by the mean value, and quantitative indicators are normalized.

Second step. Subjects of educational selection are included in the database. The initial database is formed on the intersection of all X_p, $p = \overline{1, r}$ class;

Third step. The magnitude (1) indicating the similarity of objects in the space of nominal symbols and the magnitude (2) indicating the similarity of objects in the space of numerical symbols, which are used to determine the contribution of objects of class X_p to the formation of their class, are calculated based on the formula;

Fourth step. The position of the i- object of the arbitrary p-class in the selection marks in the set of m_p-1 objects of the same class is calculated based on the formula in step 3 of solving the above-mentioned clustering problem;

Fifth step. The evaluation of the i- object in the optional p-class by other X_q, q=$\overline{1,r-1}$ class objects in the selection marks is calculated based on the formula in the step 4 of solving the above-mentioned clustering problem;

Sixth step. In selection symbols, the i- object of the arbitrary p-class is transferred to the class which is highly evaluated by the objects of the class. If the highest marks are tied, they are kept in their own class. If the upper values are equal in two classes other than the p-class, then this object is shifted to the smaller q index.

Based on the proposed theoretical research, algorithm, we will solve the problem described above. Cross-clustering of blood pressure (antihypertensive) drugs is reflected in the following table.

Table 3

Information about drugs that affect blood pressure (antihypertensive)												
№	The name of the drug	Percentage of active substance in mg	Melting point %	Color	Form of production	Blood concentration (Cmax)	Distribution density	**Technological indicators of prepared compositions**				
								Spreadability, 10^{-3} kg/s	Spreading density, kg/m^3	Metabolism in the liver %	Decomposition	Class
1	x_1	0,10	96	12	10	64	3,50	4,51	20,72	90	7,41	1
2	x_2	0,10	85	6	20	78	4,20	7,85	14,36	75	8,56	1
3	x_3	0,50	88	10	25	20	4	3,87	25,48	95	5,69	7
4	x_4	0,03	89	6	50	72	0,20	16,87	15,64	75	7,96	10
5	x_5	0,20	91	6	15	60	3,20	19,20	25,63	85	5,98	10
6	x_6	0,50	93	6	25	57	3,90	4,89	48,65	89	6,85	10
7	x_7	0,20	85	6	10	105	4,24	3,56	13,56	74	6,47	1
8	x_8	16	95	12	50	29	4,80	7,86	14,65	96	4,12	10
...												
1111	x_{1111}	0,4	75	11	30	96	3,5	6,52	12,69	85	7,89	10
1112	x_{1112}	0,2	78	3	20	97	3,6	10,42	12,25	80	8,54	10
1113	x_{1113}	0,5	76	6	25	86	4,01	7,41	13,25	78	9,65	10
1114	x_{1114}	0,25	74	3	14	93	4,5	3,41	10,26	60	12,48	10
1115	x_{1115}	0,5	72	6	28	94	3,6	4,85	12,48	85	9,85	10
1116	x_{1116}	0,25	76	6	30	85	3,9	4,63	7,96	86	8,69	10

4 A neural network-based clustering algorithm using linear regression

On the result obtained by the clustering algorithm, an experimental test was conducted on the belonging of the object to the class using a neural network. The neural network is constructed as follows:

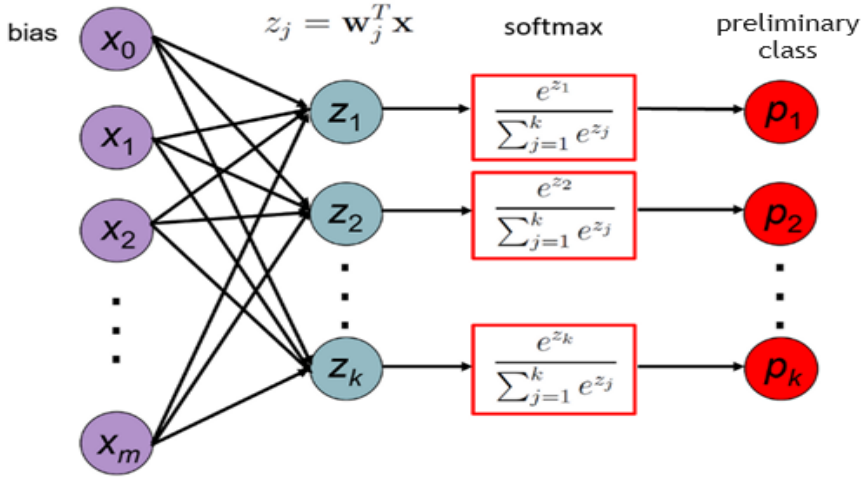

Figure 1: An exemplary structure of a neural network

A neural network based on linear regression was built for this problem.

$$z = W\overrightarrow{x} + \overrightarrow{b}$$

where W is a **weight matrix** initially filled with a set of random numbers, b is a **bias** vector initially filled with a set of random numbers. x is the **training sample** in the data set.

Activation functions are presented as follows:

$$\hat{y} = \sigma(\overrightarrow{z})$$

where $\hat{y}$ is the prediction value from the neural network. The activation functions ReLU and Softmax were used in this network.

Back-propagation and gradient descent were used to reduce errors in calculations:

$$L = -\sum (y - \log\hat{y}\ ) + (1 - y)(log1 - \hat{y})$$

$$W' = W - \alpha\frac{\partial L}{\partial W}$$

Where W'- this is a **weighting matrix** that brings the prediction closer to the true value at each step**.**

Based on the data and parameters, neural network training was carried out using the gradient descent method. The algorithm of this neural network is as follows:

First step. Import the necessary libraries and define the main parameters of the neural network.

Second step. Selection of activation functions and error elimination functions (sparse cross-entropy).

Third step. Downloading data from a file and preparing it for training and experiments.

Fourth step. Initialize neural network weights and offsets with random values.

Fifth step. Neural network training using stochastic gradient descent (SGT) and mini-batch (mini-batch) on a specific epoch (NUM_EPOCHS).

Sixth step. Class prediction for test data using a trained network.

Seventh step. Calculation of accuracy of predictions based on test data.

Eighth step. Display and visualize the accuracy of predictions on the screen.

A block diagram of this algorithm is fully described in Figure 2:

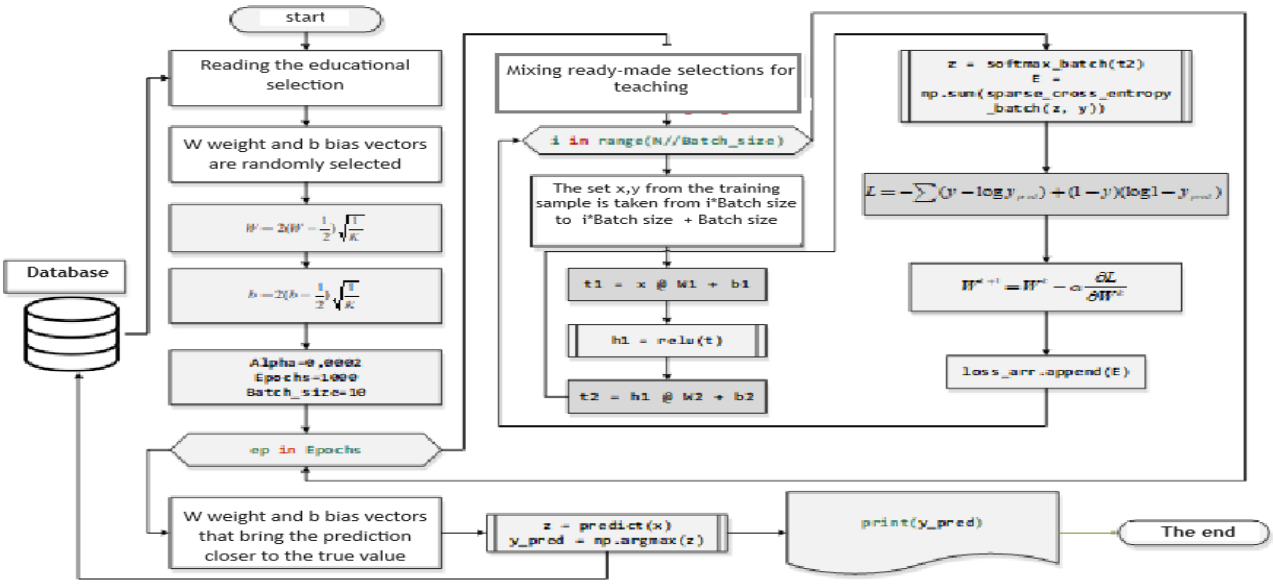

Figure 2: Block diagram of a neural network algorithm

The experimental results of the constructed neural network predicted the class of the new object with 81 - 87% accuracy.

5 A clustering model built on the KNIME platform

In addition, experiments were conducted in the **KNIME** environment, which is intended for building neural networks in the form of a **scratch-block**. A neural network model in the **KNIME** environment is shown in figure 3.

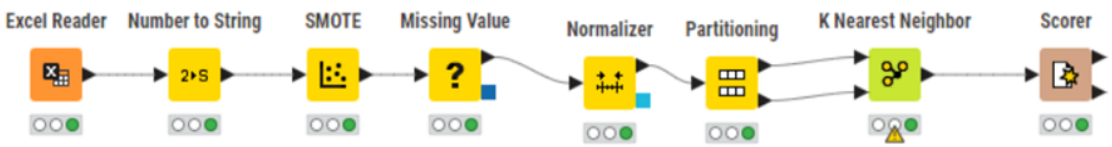

Figure 3: K-nearest neighbors model in KNIME

The functions of each **scratch block** in this model are as follows:

1. **Excel Reader** - reading and downloading data from an Excel file.
2. **Number to String** - convert numeric data to string data.
3. **SMOTE** - reproduce data of many different types using the following methods: "Synthetic Minority Over-sampling Technique".
4. **Missing Value** - filling in or removing missing data.
5. **Normalizer** - normalizing data using specific methods.
6. **Partitioning** - splitting data into test and training samples.
7. **K Nearest Neighbor** - data classification or regression with KNN algorithm.
8. **Scorer** - print summary results of model accuracy metric evaluations.

Through this model, 90% of the training sample was allocated to training and 10% to testing, and the level of clustering as a result of the model was estimated in the range of 88-94%.

The fourth table below presents a comparative analysis of the three studied algorithms.

Table 4

Class	**Recognition accuracy by neural network**	**Recognition accuracy by model generated in KNIME**	**Algorithm accuracy of classification of given objects in different types of character space**
1	82.62	94.23	95.52
2	81.4	93.48	90.18
3	84.8	88.03	97.36
4	86.23	90.26	88.07
5	82.02	94.11	95.19
6	83.79	91.48	90.27
7	81.83	93.49	94.13
8	85.61	88.43	93.47
9	87.32	94.13	94.06
10	82.55	90.88	96.47

According to the analysis, an experiment-test was conducted based on the classification of the given objects in the space of various types of symbols. Therefore, it was noteworthy that the sample was divided into clusters based on the proposed algorithm.

The first column of the Table 4 lists the names of 10 classes formed on the basis of clustering. The second column shows the accuracy of recognition by the neural network model, the third column shows the accuracy of recognition by the model created in KNIME, and the fourth column shows the accuracy of the object classification algorithm in different types of character space.

6 Conclusion

In conclusion, the article proposes a new approach to solving the problem of clustering through the character space of the objects of the sample class and an algorithm based on neural networks to correctly find the classes of the objects of the training sample.

In classification, clustering, pattern recognition and intelligent data analysis, reducing the set of symbols in pre-processing and solving the problems proposed based on them is a very important research.

An algorithm and software complex developed to clusterize drugs affecting human blood pressure has been researched. The results of the study covered more than 1,100 drugs. The tests confirmed the effectiveness of the algorithm and the ability to help.

The proposed algorithm works in solving the problem of clustering, taking into account nominal and value symbols. The selection consists of ten classes, and the information of each class is organized from nominal and value symbols. As a result of using the algorithm, a software package based on neural networks and providing the correct identification of classes was developed.

The proposed algorithm and software complex achieved high results in the clustering of drugs that affect human blood pressure, ensuring their accurate and effective grading. Techniques based on neural nets provide an opportunity to work with both nominal and value characters in mind, and this in turn increases the accuracy and efficiency of clustering.

References

[1] Pang B., & Lee, L. (2008). Opinion mining and sentiment analysis. Foundations and Trends® in Information Retrieval, 2(1-2), 1-135. DOI: 10.1561/1500000011

[2] Liu B., (2012). Sentiment analysis and opinion mining. Synthesis Lectures on Human Language Technologies, 5, 1-167. DOI: 10.2200/S00416ED1V01Y201204HLT016.

[3] Manning C.D., Raghavan P., & Schutze, H. (2008). Introduction to information retrieval. Cambridge University Press. ISBN-13: 978-0521865715

[4] Weiss, S. M., Indurkhya, N., Zhang, T., & Damerau, F. (2010). Text mining: Predictive methods for analyzing unstructured information. Springer Science & Business Media. ISBN-13: 978-1441977932.

[5] Caragea D., McNeese N., Jaiswal P., Traylor G., Kim, H., Mitra P., & Wu D. (2011). Classifying text messages for the blood pressure medication adherence challenge. Proceedings of the 2nd ACM SIGHIT Symposium on International Health Informatics - IHI '12, 71-80. DOI: 10.1145/2110363.2110377

[6] Huang X., Yang Q., & Yu Y. (2008). Towards Context-Aware Search by Learning a Very Large Variable Length Hidden Markov Model from Search Logs. Proceedings of the 17th ACM Conference on Information and Knowledge Management - CIKM '08, 1197-1206. DOI: 10.1145/1458082.1458243

[7] Кукес В.Г., Сычев Д.А. [и др.] Клиническая фармакология : учебник; 6-е изд. , испр. и доп. - Москва : ГЭОТАР-Медиа, 2021

[8] Башмаков Ю.Э., Федосеев А.Н., Денисова В.Е. Бета-адреноблокаторы и хроническая болезнь почек: литературный обзор. Качественная клиническая практика. 2023;(4):68-77

[9] Карпов Ю.А. Блокаторы кальциевых каналов при артериальной гипертонии и коморбидных состояниях. ЖУРНАЛ -Атмосфера. Новости кардиологии. №1/2020

[10] Китаева Е.Ю., Шпрах В.В. Антиагреганты в лечении и вторичной профилактике ишемического инсульта (обзор литературы). Журнал: ACTA BIOMEDICA SCIENTIFICA Том 2 № 2 , 2017

[11] Ўзбекистон Республикаси Давлат фармакопеяси. Биринчи нашр. Биринчи жилд. Биринчи ?исм. Тошкент-2021

[12] Nishanov A. Ruzibaev O. Tran N. Modification of decision rules 'ball Apolonia' the problem of classification // 2016 International Conference on Information Science and Communications Technologies, ICISCT 2016 (2016), doi: 10.1109/ICISCT.2016.7777382.

[13] Nishanov A., Saidrasulov Sh., Babadjanov E. Analysis of Methodology Of Rating Evaluation Of Digital Economy And E-Government Development In Uzbekistan // International Journal Of Early Childhood Special Education, Volume:14 Issue:2 2022, p. 2447-2452, doi: 10.9756/INT-JECSE/V14I2.230.

[14] Nishanov A., Ruzibaev O., Chedjou J.C., Kyamakya K., Abhiram, Kolli, De Silva, Djurayev G., Khasanova M. Algorithm for the selection of informative symptoms in the classification of medical data // Developments Of Artificial Intelligence Technologies In Computation And Robotics, Volume:12, 2020, p. 647-658, doi:10.1142/9789811223334_0078

[15] Nishanov A.Kh, Turakulov A.Kh, Turakhanov Kh.V., Reshaiushchee pravilo dlia klassifikatsii patologii zritel'noi sistemy [A decisive rule in classifying diseases of the visual system]. Med Tekh. 1999 Jul-Aug;(4) 16-8. Russian. PMID: 10464756.

[16] Jain A.K. (2010). Data clustering: 50 years beyond K-means. Pattern recognition letters, 31(8), 651-666. DOI: 10.1016/j.patrec.2009.09.011

[17] Johnson S.C. (1967). Hierarchical clustering schemes. Psychometrika, 32(3), 241-254. DOI: 10.1007/BF02289588

[18] Ester, M., Kriegel, H. P., Sander, J., & Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. KDD, 96(34), 226-231.

[19] Dempster A.P., Laird N.M., & Rubin D.B. (1977). Maximum likelihood from incomplete data via the EM algorithm. Journal of the royal statistical society. Series B (methodological), 39(1), 1-38. DOI: 10.1111/j.2517-6161.1977.tb01600.

[20] Ng A.Y., Jordan M.I., & Weiss, Y. (2002). On spectral clustering: Analysis and an algorithm. Advances in neural information processing systems, 2, 849-856.

Авторлар туралы мәлімет:

Нишанов Ахрам Хасанович – доктор технических наук, Мұхаммед әл-Хорезми атындағы Ташкент ақпараттық технологиялар университетінің программалық инженерия факультетінің профессоры, (Ташкент, Өзбекстан, электрондық пошта: nishanov_ akram@mail.ru);

Турсунов Алишер Тулкунович – Ташкент фармацевтикалық институтының физика, математика және ақпараттық технологиялар кафедрасының ага оқытушысы (Ташкент, Өзбекстан, электрондық пошта: tursunovr484za@gmail.com);

Олламберганов Файзулла Фарход угли – Мұхаммед әл-Хорезми атындағы Ташкент ақпараттық технологиялар университетінің жүйелік және қолданбалы программалау кафедрасының докторанты (Ташкент, Өзбекстан, электрондық пошта: fayzulla0804@gmail.com);

Рашидова Дильфуза Элмуродовна – Мұхаммед әл-Хорезми атындағы Ташкент ақпараттық технологиялар университетінің жүйелік және қолданбалы программалау кафедрасының докторанты (Ташкент, Өзбекстан, электрондық пошта: dilfuzarashidova23@gmail.com).

Информация об авторах:

Нишанов Ахрам Хасанович (ответственный автор) – доктор технических наук, профессор факультета программной инженерии Ташкентского университета информационных технологий имени Мухаммада Ал-Хоразмий (Ташкент, Узбекистан, электронная почта: nishanov_ akram@mail.ru);

Турсунов Алишер Тулкунович – Старший преподаватель кафедры физики, математики и информационных технологий Ташкентского фармацевтического института (Ташкент, Узбекистан, электронная почта: tursunovr484za@gmail.com);

Олламберганов Файзулла Фарход угли – докторант кафедры системного и прикладного программирования Ташкентского университета информационных технологий имени Мухаммада Ал-Хоразмий (Ташкент, Узбекистан, электронная почта: fayzulla0804@gmail.com);

Рашидова Дильфуза Элмуродовна – докторант кафедры системного и прикладного программирования Ташкентского университета информационных технологий имени Мухаммада Ал-Хоразмий (Ташкент, Узбекистан, электронная почта: dilfuzarashidova23@gmail.com).

Information about authors:

Nishanov Akhram Khasanovich (corresponding author) – DSc, professor of the Faculty of Software engineering of Tashkent University of Information Technologies named after Muhammad Al-Khwarizmi (Tashkent, Uzbekistan, e-mail: nishanov_ akram@mail.ru);

Tursunov Alisher Tulkunovich – Senior lecturer of the Department of Physics, Mathematics and Information Technologies of Tashkent Pharmaceutical Institute (Tashkent, Uzbekistan, e-mail: tursunovr484za@gmail.com);

Ollamberganov Fayzulla Farxod o'g'li – PhD student Department of System and applied programming of Tashkent University of Information Technologies named after Muhammad Al-Khwarizmi (Tashkent, Uzbekistan, e-mail: fayzulla0804@gmail.com);

Rashidova Dilfuza Elmurodovna – PhD student Department of System and applied programming of Tashkent University of Information Technologies named after Muhammad Al-Khwarizmi (Tashkent, Uzbekistan, e-mail: dilfuzarashidova23@gmail.com).

Received: September 13, 2024
Accepted: February 23, 2025