The study addresses the task to improve the accuracy of clustering air raid danger levels by constructing a hybrid clustering algorithm.

A target air clustering algorithm has been developed, which involves using a modified distance metric and integrates air danger level assessments directly into the algorithm.

The reported features demonstrate superiority over existing algorithms based on the Silhouette and Davies-Bouldin indices. The proposed model yields a Silhouette index of 0.72306 compared to 0.3481 for the existing model, and a Davies-Bouldin index of 0.3389 compared to 1.209. Models such as Random Forest Classifier and Gradient Boosting Classifier, evaluated using the clusterizer, exhibit higher accuracy, specifically 0.87 and 0.87, respectively, compared to existing models with 0.48 and 0.49, respectively.

The distinctive feature of the clusterizer is the use of more accurate input assessments, determined by the principle of interaction and linear scaling. The proposed algorithm involves using a modified chi-square distance metric, which includes assessments of state security indices. A notable feature of the proposed approach is the more accurate determination of cluster centers using Kohonen self-organizing maps. This helps solve the task of analyzing and improving the accuracy of predicting air threat levels. The results are explained by the use of more accurate input assessments and a well-chosen distance metric between clusters in combination with Kohonen self-organizing maps.

In practice, the results could be used for analyzing air danger levels by a ground-based platform

Keywords: danger clustering, linear scaling, unsupervised learning, composite indicator, artificial intelligence, air danger

IDENTIFICATION OF AIR TARGETS USING A HYBRID CLUSTERING ALGORITHM

Oleksandr Laktionov
Corresponding author
PhD
Department of Automation,
Electronics and Telecommunications*
E-mail: itm.olaktionov@nupp.edu.ua
Alina Yanko
PhD, Associate Professor
Department of Computer and Information
Technologies and Systems*
Nazar Pedchenko
PhD
Department of Oil and
Gas Engineering and Technology*
*National University
"Yuri Kondratyuk Poltava Polytechnic"
Pershotravnevyi ave., 24, Poltava, Ukraine, 36011

1. Introduction

The number of new artificial intelligence models for air target clustering is increasing every day; the advantages of these tools are obvious – constant training and updating of the model [1]. Most projects are similar to each other, and this does not distinguish them from others. Therefore, currently preference is given to ideas that can be scaled to industrial sites [2]. This ensures the implementation of ideas on an industrial scale.

Research on the clustering of air targets should be carried out under today's conditions for effective analysis and forecasting of the security of the European space. Existing developments are complex and focused on ensuring infrastructure security both physically and software [3].

From a practical point of view, the design of new tools for alternative data processing systems [4] and artificial intelligence [5] meets the requirements for Europe's digital transformation [6]. Therefore, research on devising new approaches to air target clustering is relevant.

2. Literature review and problem statement

One of the approaches to solving the problem of air target clustering can be based on the use of ANFIS-type systems.

Such systems, on the basis of a set of input parameters (attributes), make it possible to identify the object (find the initial value). Thus, in work [7], the system is used to find the amount of signal attenuation in wireless networks. The data are subsequently used to make a decision on adapting the signal strength. Therefore, the use of this type of system for solving the clustering problem is possible in the presence of a set of input data with appropriate characteristics of classes of air targets. This is not fully resolved.

The performance of clustering algorithms remains an open question today, which is solved in various ways, in particular, by wireless hardware [8]. But work [8] lacks a solution for clustering the level of threats of air danger. Work [9] presents the process of image clustering, where the uniqueness of the method is the possibility of analyzing 2D and 3D images. The use of the proposed method, according to the purpose of the study, is limited only to image clustering, which affects performance. One of the ways to improve the performance of the DBScan algorithm is considered in [10]. The authors investigated the algorithm using textual data, taking into account the ideas of the noise/signal approach. Among several algorithms that failed, the most productive results were demonstrated by the DBScan algorithm. This is explained by its features. However, the issue of creating a clusterizer for grouping the level of threats of air objects remains unresolved.

There are well-known techniques for increasing productivity through the use of ideas of semi-controlled clustering, taking into account the security aspect [11]. The security component is provided by regularization, which limits the existence of false clustering results. But on the other hand, the proposed approach does not involve the double use of chi-square tools, which would make it possible to determine the difference between the input data.

Semi-supervised learning is not only used for aerial target detection. Thus, in work [12], semi-supervised learning is used to diagnose abnormal values of the network. Among the limitations of the proposed approach, worth noting is the work only with the existing data package, which makes it difficult to obtain results online.

One of the simple and easy-to-use methods of unsupervised learning is k-means, the evolution of which is reviewed in work [13]. Thus, in [14], a method of similarity clustering based on weighted and Euclidean distances is proposed. Proving the effectiveness of the proposed approach was carried out with the help of three datasets, which had different data volumes of 150, 178, and 306 objects, respectively. Both binary and multiple clustering were considered. According to the research results, the advantage of the proposed approach, which provides more accurate clustering of objects, has been proven. However, the issue of defining the differences and key advantages between the existing and the proposed method remained out of the authors' attention.

In contrast to [14], in [15] the k-means algorithm is improved through the elbow rule. This makes it possible to simplify the process of choosing the center of the cluster, which also affects the speed of operation. Experimental verification was carried out on synthetic data with a noise level of 0.2 and ten real sets of estimates. F-score, Adjusted Rand Index, Normalized Mutual Information, and others were used as criteria for determining clustering quality. The results of experimental verification of the proposed solutions confirmed their effectiveness on ten real datasets. However, data on the number of air alarms were not used as datasets.

The common limitations of works [14, 15] are that they were not used for computer vision tasks, in particular segmentation, as in [16]. According to the research methodology [16], clustering was considered as one of the tools for segmentation of the color gamut of the image. With a confidence level of 95 % at $p < 0.05$, the results showed the disadvantages of k-means compared to Fuzzy C-Means. Work [16] involves the use of images rather than numerical data for the analysis of air targets, which is one of its limitations.

In [17], a high-speed clustering algorithm was proposed using a single-threaded protocol that ensures information confidentiality. The uniqueness of the solution is the computing mechanism of the protocol, which is capable of clustering 100,000 records in two hours. According to research results [17], the authors proved that the algorithm is five orders of magnitude faster than existing analogues. Thus, companies can conduct joint research and not share commercial information. But as in the previous studies, the issue of working with the number of air alarms was not fully studied.

Large volumes of data create problems with grouping of relevant objects by means of k-means, which is emphasized in paper [18]. In this regard, an improvement of k-means using the Lévy flight path is proposed, which does not increase the complexity and performance of the algorithm. As the practical verification of the proposed solution showed, on ten data sets, the algorithm provides a uniform distribution of centroids. But

in accordance with the purpose of the study, the issue of using distance metrics, in particular chi-square, was not studied.

The problem of checking the difference in means between clusters of observations is studied in [19]. To solve the problem, the use of p-value is proposed, and the use of lemmas is considered. This idea improves the results of clustering digital data, in particular text and data used in medicine. However, the research does not fully reveal the issue of creating or improving the metric for determining the distance between clusters.

In contrast to work [19], paper [20] thoroughly studied the use of clustering for a large volume of data. This scaling of proposed ideas sets the work apart from others. The advantage of the approach is the ability to conduct research online. However, the proposed approach was not used to solve the problems of air target clustering.

Methods of analysis of large volumes of data were studied in works [21] similarly to [20]. The approach combines deep learning with a convolutional autoencoder to characterize the data. Experimental verification of the solution was tested on different types of data, where 4 or more clusters were studied. As the authors of work [21] note, the limitations of the study are the model's operating time, it works more slowly than dimensionality reduction methods.

A deep learning tool capable of working online is also studied in [22]. This is ensured by the use of convolutional and recurrent networks. The proposed solutions are oriented towards the use and the field of energy, where the experimental verification was carried out on the data on the consumption of electrical energy. The proposed approach can be the basis for creating a clusterizer of air alarm estimates, which is worth further study.

Work [23] is also characterized by the use of hybrid models combining deep learning tools and modified k-means algorithms. The proposed solution is characterized by low energy consumption and high productivity.

There are studies [24] in which clustering tasks are studied by means of fuzzy logic. From a practical point of view, this requires the use of powerful hardware resources. Similar research results were reported in [25], in which the interaction of the components of the Internet of Things was additionally studied. At the initial stages of research, it is difficult to implement the Internet of Things, so some works focus on studying the construction of recommendation systems [26]. These systems allow detecting anomalies [27]. But without a created clusterizer model, the study of anomalies is not expedient.

As can be seen from [7–27], these studies contain a number of tools for creating clusterizers for various tasks. The gaps are insufficient study of the use of modified ranging metrics and the integration of air threat level estimates directly into the algorithm. The proposed ideas also do not fully take into account the principles of interaction and linear scaling when devising appropriate methods. This is what forms the essence of the general problem of our research aimed at increasing the accuracy of clustering the assessments of danger level of an air raid alarm.

3. The aim and objectives of the study

The purpose of our study is to increase the accuracy of identification of air targets through the differentiation of objects based on a new clustering method. This will make it possible to analyze the air threat level more efficiently.

To achieve the goal, the following tasks were set:
– to propose a method for clustering air targets;
– to carry out experimental verification of the proposed solutions.

4. The study materials and methods

The object of our study is the process of developing an air alarm danger level clusterizer.

The main hypothesis of the study is to increase the accuracy of the clustering of estimates due to the use of more accurate estimates submitted to the input of the clusterizer. This can be achieved by using the principle of linear scaling and combining the series of estimates directly in the algorithm before the start of clustering (hybridization principle).

To design a clusterizer, we used the results from work [28], in particular, an experimental sample of estimates of the state security index in the range [1, 5], $N = 605$ estimates.

An experimental sample of estimates obtained by the proposed and existing approach was used to develop a clustering algorithm, which was built according to the hybrid principle. The optimal model of the clusterizer was chosen using the Davies-Bouldin index and silhouette estimation [29, 30]. The study of this stage provided for the following limitations of the algorithm, in particular, the number of iterations is 100, the number of clusters is 2. The effectiveness of the clusterizer is determined by comparative analysis of the proposed and existing approaches using a percentage ratio.

One of the initial values of the clustering algorithm were the cluster numbers and distance estimates from the actual value to the center of the cluster, which were input to the five classification algorithms. Random Forest Classifier, Support Vector Machine Classifier, k-Nearest Neighbors Classifier, Logistic Regression Classifier, Gradient Boosting Classifier were considered as classification algorithms. The objective classification task was to determine the accuracy of the classification algorithm based on the input clustered scores, where class balances were not taken into account from the beginning and hyperparameters were not used. Class balancing was studied at the stage of the selected optimal classification model, with the aim of studying its adequacy.

Algorithms for classification and clustering were built in a classical way using the Python programming language and the Sklearn library. Differentiation of estimates into bulk and test sets, in the classification task, was carried out in the classic way – train_test_split from the Sklearn library, where shuffle = True, test_size = 0.35. Accuracy, precision, recall, and fi-score were used as metrics for determining the quality of classification.

The study of class balances was carried out using the tools of the imblearn library, in particular, the Random Under der Sampler method. Learning curves were used to prove the adequacy of the constructed model. Graphical interpretation of scatter maps of scores is implemented by means of matplotlib.

The clusterizer model built is saved in a file with the extension .pkl, which is used to create a web page using Streamlit tools.

The formal statement of the research task is the construction of a clusterizer of state security level assessments based on the hybrid principle. We use Chi-square as both a distance metric and a distance metric from the cluster center to the actual estimate. The optimal clustering model is selected according to the criteria of the Davies-Bouldin Index, silhouette assessment under the conditions Optimal model$_{\text{DBI}} = \arg \ \min_m \text{DBI}_m$; Optimal model$_S = \arg \max_m S_m$; Optimal model$ = \text{model}_{\text{DBI}} \cap \text{models}$. The found optimal model is used to make decisions.

5. Results of research on the identification of air targets using a hybrid clustering algorithm

5. 1. The proposed method of air target clustering

According to the research methodology, the input estimates for building the clusterizer are determined from formula (1) given in paper [28]:

$$Ks = \left(\left((5-1) \left(\text{current assessment} \begin{pmatrix} x1,t + x2,t + x3,t + x4, \\ t + (x1,t \cdot x3,t) + (x2,t \cdot x4,t) \end{pmatrix} \right) \right) / Ki \max - 1 \right) + 1, \quad (1)$$

where $x1$ is the type of unmanned aerial vehicle or missile; $x2$ – the number of launched missiles, (from 1 to n); $x3$ – the number of downed missiles, (from 0 to k); $x4$ – technique for launching the object.

In addition to the array of values determined from formula (1), the research involves the use of a well-known approach [28], in which the specified methods are implemented in a clusterizer. Our studies on determining the structure of the clusterizer showed the achievement of an effective result through the combination of two series of estimates in the clustering algorithm immediately before clustering. This simplifies the clustering process, Steps 1–6:

Step 1. Determining the distance using Xi^2 [31]. The input values are the values of the proposed and existing indices, and the output value is the distance. The initial array of scores is used to determine the initial cluster centers.

Step 2. Determination of initial cluster centers randomly from the array of scores obtained in the previous step using Cochrane self-organizing maps [32].

Step 3. Calculation of the value from the center of the cluster to the actual estimate using Xi^2 [31].

Step 4. The values are recalculated and sorted until the grades are fully grouped.

Step 5. The silhouette index [29] and Davies-Bouldin [30] metrics are calculated, the results are recorded in Table 1.

Table 1

Results of the construction of the clustering algorithm

Name of the approach studied		
Number of ratings in a cluster	Silhouette index	Davies-Bouldin index
Cluster 0: $k1$; Cluster 1: $k2$;	S_m	DBI_m

Step 6. The results in Table 1 are used to provide recommendations and make decisions.

5. 2. Experimental verification of the proposed solutions

The results of clustering by the proposed and existing method are given in Table 2.

As can be seen from Table 2, the proposed method has an advantage over the existing one, which is confirmed by the

silhouette index values of 0.72306; 0.3481 and Davies-Bouldin 0.3389; 1.209, respectively. One of the ways to explain the obtained result is to more accurately determine the centers of clusters through the use of the proposed metric for determining the distance and the method for determining the centers of clusters. The graphic interpretation of the research results is shown in Fig. 1.

Table 2

Comparative analysis of clusters of the proposed and existing methods

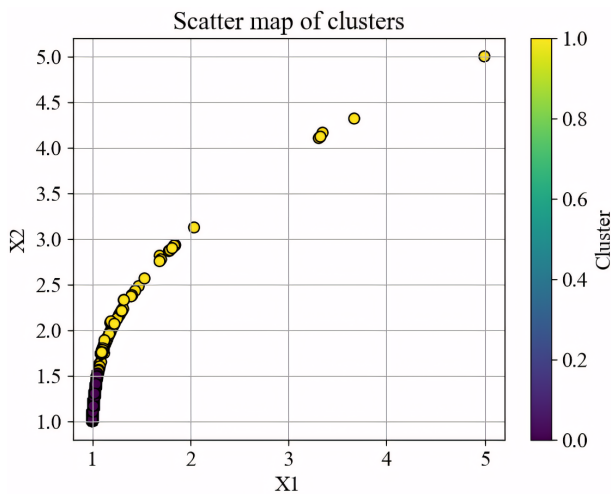| The proposed method (modified distance determination metric)		
Number of grades in a cluster	Silhouette index	Davies-Bouldin index
0:414; 1:190	0.72306	0.3389
Existing method 2 (the Euclidean distance metric)		
Number of ratings in a cluster	Silhouette index	Davies-Bouldin index
0:294; 1:310	0.3481	1.209

Fig. 1. Scatter map of clustering results by the proposed method

Graphical interpretation of the results of clustering of scores confirms the existence of two clusters. The values of the first cluster are in the range from 1.0 to 1.25, the second – from 1.251 to 5.0, where the number of assessments of each cluster is 190 and 414, respectively. In comparison with existing approach, in which the number of features of each cluster is approximately evenly distributed between 294 and 310 estimates. On the one hand, the proposed method demonstrated the formation of unbalanced clusters, on the other hand, the density of estimates made it possible to gain an advantage over existing approach according to the criteria of the silhouette and Davies-Bouldin indices.

The silhouette index of the proposed method is 0.72306, which reaches 100 %. The known approach yields 0.3481, $-x$. If we find the percentage of the number, we shall get the efficiency of the proposed method, which has increased by 51.85 %. From a practical point of view, this significantly affects the results of the analysis of arrays of air attack data and the accuracy of classifiers; Table 3.

According to the research methodology, the task of classifying clustered estimates of air alarms was studied using five algorithms. The optimal accuracy value was demonstrated by the Random Forest Classifier and Gradient Boosting Classifier algorithms. The accuracy of the specified algorithms on the proposed/existing data reached 0.87/0.48 and 0.87/0.49, respectively.

Algorithms Support Vector Machine Classifier, k-Nearest Neighbors Classifier, Logistic Regression Classifier showed less accurate classification results, where the accuracy (based on estimates by the proposed/existing approach) reached 0.7/0.48; 0.83/0.51; 0.68/0.5, respectively. The level of accuracy of the algorithms, which is logical, affected the number of correctly/falsely classified scores.

All used classification algorithms confirmed the fact that increasing the accuracy of clustering improves the accuracy of classification. This further confirms the advantage of clustering air attack level estimates by the proposed tools.

Practical use of the proposed method is possible with the help of a single-board computer, in particular raspberry pi, which will be placed in a specialized case for transportation. This will make it possible to analyze and predict the level of danger of an air alert. The proposed solution could be used as an addition to existing air alarm monitoring systems. To use the solution on Android or other means of communication, implementation was carried out using Streamlit tools, Fig. 2.

Table 3

Comparative analysis of classification results based on a number of estimates obtained by the proposed and existing methods

Indicator	Precision proposed/existing	Recall proposed/existing	F1-score proposed/existing	Support proposed/existing
Random Forest Classifier				
Cluster 0	0.92/0.46	0.9/0.41	0.91/0.43	145/103
Cluster 1	0.79/0.5	0.82/0.55	0.8/0.52	67/109
Accuracy	0.87/0.48			
macro avg	0.85/0.48	0.86/0.48	0.85/0.48	212
weighted avg	0.87/0.48	0.87/0.48	0.87/0.48	212
Gradient Boosting Classifier				
Cluster 0	0.9/0.48	0.9/0.54	0.9/0.51	145/103
Cluster 1	0.79/0.51	0.79/0.44	0.79/0.47	67/109
Accuracy	0.87/0.49			
macro avg	0.85/0.49	0.85/0.49	0.85/0.49	212
weighted avg	0.87/0.49	0.87/0.49	0.87/0.49	212

Data Clustering Using Saved Model

Upload Excel file with data for clustering

Drag and drop file here

> ☁ **Drag and drop file here**
> Limit 200MB per file • XLSX

📄 processed_data_2-1 experement.xlsx 89.9KB

Uploaded file: processed_data_2-1 experement.xlsx (89.9 KB)

Loaded data:

	model	launched	destroyed	carrier	Unnamed: 4	Unnamed: 5	launched_1	destroyed_1	model_1	carrier_1
0	4	5	4	4	17	None	1.1684	1.1622	4	4
1	1	3	3	1	8	None	1.0842	1.1081	1	1
2	1	7	5	1	14	None	1.2526	1.2162	1	1
3	1	3	3	1	8	None	1.0842	1.1081	1	1

Fig. 2. General view of the implemented clusterizer using Streamlit tools

Using the clusterizer requires downloading the source data in .xlsx format, as shown in Fig. 2. After that, the user can view the structure of the downloaded data and the clustering results. The premise of the loaded data is the preliminary setting of the column names of the table for diagnosing air targets.

6. Discussion of results of research on the identification of aerial targets by a hybrid clustering algorithm

Our research results regarding the development of the air target clusterizer are explained by the use of the security index (1), which is determined directly in the clustering algorithm.

The proposed method increases the accuracy of the initial clustering result by using the principles of linear scaling and interaction, which determine the security index (1). This ensures a decrease in the mean square deviation of the index, which was shown in [28]. By modifying the metric for determining the distance of the clusterizer by using the security index (1), an increase in the accuracy of object grouping is observed.

Therefore, the results of air target clustering were obtained (Table 2) according to the criteria of the silhouette and Davies-Bouldin indices; compared to existing results of the studies, they demonstrate an advantage.

The results shown in Fig. 1 are a graphical interpretation of the estimates of two clusters, which are placed on certain ranges. The estimates of the second cluster are located at a certain distance from each other, in particular, there are no estimates in the range [2.0, 3.25], which is explained by the peculiarities of the input dataset. The density of placement of estimates is greater than that of the existing solution, which shows the advantage of the proposed method using mathematical tools.

The effectiveness of the proposed method is also confirmed by using classifiers, where the input of the classifiers was provided with estimates determined by the proposed and existing approach. Therefore, classifiers created on the basis of estimates determined by the proposed solution (Table 3) show higher accuracy values. Thus, the proposed method resolves the issue.

In contrast to the results reported in [24], in which issues of clusterizer development were considered, our research result showed better quality indicators. This was made possible by using the clusterizer's distance metric, which takes into account the security index scores. On the other hand, the result is explained by the use of a more accurate method for determining the centers of clusters. This is possible thanks to Cochrane self-organization maps, where, unlike averaged values [23], a more stable search for the corresponding cluster centers is provided.

Among the limitations of the study worth noting is the possibility of data analysis only offline when the number of studied clusters is 2. From a practical point of view, this allows more effective forecasting of the air threat level. From a theoretical point of view, our research result will make it possible to build more effective clustering models taking into account specific requirements.

The disadvantages of the study are the use of only four types of air targets, which are combined into a single assessment of the safety index. This limits the number of clusters and the flexibility of decisions based on the results of using the clusterizer.

The development of this research involves studying the metric of determining the distance from the actual assessment to the center of the cluster, which is the main tool for improving the clustering algorithm.

7. Conclusions

1. The task to devise a method for air target clustering is solved by combining two series of assessments of the level of air danger directly in the algorithm before the start of clustering. In addition, the algorithm provides for the use of a modified chi-square metric, which includes estimates of

state security indices. This makes it possible to increase the efficiency of clustering of grades by 51.85 %.

2. The proposed clustering model has an advantage over existing ones according to the criteria of silhouette and Davies-Bouldin indices. Thus, the proposed model shows a silhouette index of 0.72306, and the existing one – 0.3481; the Davies-Bouldin index is 0.3389; 1.209, respectively.

Our clustering results affect the accuracy of the classification models. The accuracy of the Random Forest Classifier, Gradient Boosting Classifier models based on the proposed estimates is 0.87; 0.87, respectively, and existing – 0.48; 0.49, respectively. The results are explained by using the principles of linear scaling and hybridization by combining the series of estimates directly in the algorithm before the start of clustering.

Conflicts of interest

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

Funding

Data availability

The manuscript has associated data in the data warehouse.

Use of artificial intelligence

The authors confirm that they did not use artificial intelligence technologies when creating the current work.

References

1. O'Shaughnessy, D. (2024). Trends and developments in automatic speech recognition research. Computer Speech & Language, 83, 101538. https://doi.org/10.1016/j.csl.2023.101538

2. Seo, D., Kim, S., Oh, S., Kim, S.-H. (2022). K-Means Clustering-Based Safety System in Large-Scale Industrial Site Using Industrial Wireless Sensor Networks. Sensors, 22 (8), 2897. https://doi.org/10.3390/s22082897

3. Leal Piedrahita, E. A. (2019). Hierarchical Clustering for Anomalous Traffic Conditions Detection in Power Substations. Ciencia e Ingeniería Neogranadina, 30 (1), 75–88. https://doi.org/10.18359/rcin.4236

4. Yanko, A., Koshman, S., Krasnobayev, V. (2017). Algorithms of data processing in the residual classes system. 2017 4th International Scientific-Practical Conference Problems of Infocommunications. Science and Technology (PIC S&T), 117–121. https://doi.org/10.1109/infocommst.2017.8246363

5. Onyshchenko, S., Haitan, O., Yanko A., Zdorenko, Y., Rudenko, O. (2024). Method for detection of the modified DDoS cyber attacks on a web resource of an Information and Telecommunication Network based on the use of intelligent systems. Proceedings of the Modern Data Science Technologies Workshop (MoDaST 2024), 219–235. Available at: https://ceur-ws.org/Vol-3723/paper12.pdf

6. Krasnobayev, V., Yanko, A., Hlushko, A. (2023). Information Security of the National Economy Based on an Effective Data Control Method. Journal of International Commerce, Economics and Policy, 14 (03). https://doi.org/10.1142/s1793993323500217

7. Zdorenko, Y., Lavrut, O., Lavrut, T., Nastishin, Y. (2020). Method of Power Adaptation for Signals Emitted in a Wireless Network in Terms of Neuro-Fuzzy System. Wireless Personal Communications, 115 (1), 597–609. https://doi.org/10.1007/s11277-020-07588-5

8. Fachrizal, F., Zarlis, M., Sihombing, P., Suherman, S. (2024). Optimization of the LEACH algorithm in the selection of cluster heads based on residual energy in wireless sensor networks. Eastern-European Journal of Enterprise Technologies, 1 (9 (127)), 14–21. https://doi.org/10.15587/1729-4061.2024.298268

9. Szalontai, B., Debreczeny, M., Fintor, K., Bagyinka, Cs. (2020). SVD-clustering, a general image-analyzing method explained and demonstrated on model and Raman micro-spectroscopic maps. Scientific Reports, 10 (1). https://doi.org/10.1038/s41598-020-61206-9

10. Crețulescu, R. G., Morariu, D. I., Breazu, M., Volovici, D. (2019). DBSCAN Algorithm for Document Clustering. International Journal of Advanced Statistics and IT&C for Economics and Life Sciences, 9 (1), 58–66. https://doi.org/10.2478/ijasitels-2019-0007

11. Gan, H., Yang, Z., Zhou, R. (2023). Adaptive safety-aware semi-supervised clustering. Expert Systems with Applications, 212, 118751. https://doi.org/10.1016/j.eswa.2022.118751

12. Powell, B. A. (2022). Role-based lateral movement detection with unsupervised learning. Intelligent Systems with Applications, 16, 200106. https://doi.org/10.1016/j.iswa.2022.200106

13. Pérez-Ortega, J., Nely Almanza-Ortega, N., Vega-Villalobos, A., Pazos-Rangel, R., Zavala-Díaz, C., Martínez-Rebollar, A. (2020). The K-Means Algorithm Evolution. Introduction to Data Science and Machine Learning. https://doi.org/10.5772/intechopen.85447

14. Zhao, Y., Zhou, X. (2021). K-means Clustering Algorithm and Its Improvement Research. Journal of Physics: Conference Series, 1873 (1), 012074. https://doi.org/10.1088/1742-6596/1873/1/012074

15. Zhao, H. (2022). Design and Implementation of an Improved K-Means Clustering Algorithm. Mobile Information Systems, 2022, 1–10. https://doi.org/10.1155/2022/6041484

16. Wiharto, W., Suryani, E. (2020). The Comparison of Clustering Algorithms K-Means and Fuzzy C-Means for Segmentation Retinal Blood Vessels. Acta Informatica Medica, 28 (1), 42. https://doi.org/10.5455/aim.2020.28.42-47

17. Mohassel, P., Rosulek, M., Trieu, N. (2020). Practical Privacy-Preserving K-means Clustering. Proceedings on Privacy Enhancing Technologies, 2020 (4), 414–433. https://doi.org/10.2478/popets-2020-0080

18. Hu, H., Liu, J., Zhang, X., Fang, M. (2023). An Effective and Adaptable K-means Algorithm for Big Data Cluster Analysis. Pattern Recognition, 139, 109404. https://doi.org/10.1016/j.patcog.2023.109404

19. Chen, Y. T., Witten, D. M. (2023). Selective inference for k-means clustering. Journal of Machine Learning Research, 24 (152), 1–41. Available at: https://jmlr.org/papers/v24/22-0371.html

20. Mortensen, K. O., Zardbani, F., Haque, M. A., Agustsson, S. Y., Mottin, D., Hofmann, P., Karras, P. (2023). Marigold: Efficientk-Means Clustering in High Dimensions. Proceedings of the VLDB Endowment, 16 (7), 1740–1748. https://doi.org/10.14778/3587136.3587147

21. Hu, H., Li, Z., Li, X., Yu, M., Pan, X. (2021). ScCAEs: deep clustering of single-cell RNA-seq via convolutional autoencoder embedding and soft K-means. Briefings in Bioinformatics, 23 (1). https://doi.org/10.1093/bib/bbab321

22. Somu, N., Raman M R, G., Ramamritham, K. (2021). A deep learning framework for building energy consumption forecast. Renewable and Sustainable Energy Reviews, 137, 110591. https://doi.org/10.1016/j.rser.2020.110591

23. Bisen, D., Lilhore, U. K., Manoharan, P., Dahan, F., Mzoughi, O., Hajjej, F. et al. (2023). A Hybrid Deep Learning Model Using CNN and K-Mean Clustering for Energy Efficient Modelling in Mobile EdgeIoT. Electronics, 12 (6), 1384. https://doi.org/10.3390/electronics12061384

24. Kondruk, N. (2017). Clustering method based on fuzzy binary relation. Eastern-European Journal of Enterprise Technologies, 2 (4 (86)), 10–16. https://doi.org/10.15587/1729-4061.2017.94961

25. Mohammed, A. S., Balaji, B. S., Basha M. S., S. (2019). Fuzzy applied energy aware clustering based routing for iot networks. Advanced Information Systems, 3 (4), 140–145. https://doi.org/10.20998/2522-9052.2019.4.22

26. Krepych, S., Spivak, I. (2021). Improvement of svd algorithm to increase the efficiency of recommendation systems. Advanced Information Systems, 5 (4), 55–59. https://doi.org/10.20998/2522-9052.2021.4.08

27. Khoroshun, G., Ryazantsev, O., Cherpitskyi, M. (2023). Clustering and anomalies of USA stock market volatility index data. Advanced Information Systems, 7 (2), 9–15. https://doi.org/10.20998/2522-9052.2023.2.02

28. Shefer, O., Laktionov, O., Pents, V., Hlushko, A., Kuchuk, N. (2024). Practical principles of integrating artificial intelligence into the technology of regional security predicting. Advanced Information Systems, 8 (1), 86–93. https://doi.org/10.20998/2522-9052.2024.1.11

29. davies_bouldin_score. Scikit-learn. Available at: https://scikit-learn.org/stable/modules/generated/sklearn.metrics.davies_bouldin_score.html

30. silhouette_score. Scikit-learn. Available at: https://scikit-learn.org/stable/modules/generated/sklearn.metrics.silhouette_score.html

31. Laktionov, A. (2019). Application of index estimates for improving accuracy during selection of machine operators. Eastern-European Journal of Enterprise Technologies, 3 (1 (99)), 18–26. https://doi.org/10.15587/1729-4061.2019.165884

32. García-Tejedor, Á. J., Nogales, A. (2022). An open-source Python library for self-organizing-maps. Software Impacts, 12, 100280. https://doi.org/10.1016/j.simpa.2022.100280