UDC 004.942.048:004.85 **doi:** https://doi.org/10.20998/2522-9052.2024.3.10

Oleksandr Laktionov[1], Oleksandr Shefer[1], Iryna Laktionova[1], Vasyl Halai[1], Andrii Podorozhniak[2]

[1] National University "Yuri Kondratyuk Poltava Polytechnic", Poltava, Ukraine
[2] National Technical University "Kharkiv Polytechnic Institute", Kharkiv, Ukraine

IMPLEMENTATION OF UNSUPERVISED LEARNING MODELS FOR ANALYZING THE STATE'S SECURITY LEVEL

Abstract. Objective. Enhancing the effectiveness of preliminary analysis of the state's security level through the implementation of clustering models. **Methodology.** The process of creating unsupervised learning models and their peculiarities in tasks of analyzing the state's security level has been investigated. Techniques for creating the basic k-means model and its improvement through the use of Pearson correlation as a distance metric have been considered. Determining cluster centers was performed by both the basic method and the Cochran's maps method. The optimal quality indicator, according to the results of clustering, was considered to be the model demonstrating the minimum value of the Davies-Bouldin index. **Results.** An improved unsupervised learning model based on the k-means algorithm for analyzing the state's security level has been developed. The model is characterized by two clusters, with centroids determined as 1.112 and 1.009. **Scientific novelty.** The proposed model for clustering the state's security level differs from existing ones by using as input estimates derived from a comprehensive indicator based on the principles of interaction and emergent properties. This allows obtaining advantages of the clustering model in terms of the Davies-Bouldin index. The existing clustering model demonstrates a value of 0.4765, while the proposed one achieves 0.2166. **Practical significance.** The proposals serve as a useful additional tool for preliminary analysis of the state's security level during air alerts and extend the functionality of the previously researched forecasting technology.

Keywords: clustering model; linear scaling; factor interaction.

Introduction

The advancement of artificial intelligence technologies, on one hand, simplifies the analysis of large volumes of data [1, 2], while on the other hand, it entails the development of tools aimed at ensuring the security of state technological information [3]. This process is oriented not only towards the development of artificial intelligence methods but also towards considering fundamental aspects of information security [4, 5].

Diagnosing the state's security level is currently a top priority, where it is necessary to ensure speed, reliability, integrity, etc. [6, 7]. Therefore, the creation of new tools, both software and hardware, is being prioritized. In this process, a variety of tools are used, including optimization methods [8]. However, existing solutions are highly non-universal and require a series of additional adjustments, which are associated with the specific characteristics of the models being studied.

Therefore, there is a need to develop new and improve existing approaches to grouping objects in order to identify the security level of the state as a complex system.

The aim of the research is to enhance the efficiency of preliminary analysis of the state's security level by implementing clustering models.

The research object is a comprehensive indicator for solving the task of analyzing the state's security level.

The research subject is unsupervised learning models and methods of analyzing the state's security level.

The research tasks:

1. Improve the comprehensive indicator - the state security index by considering the principles of interaction and emergent properties.

2. Create an unsupervised learning model for analyzing the state's security level.

3. Conduct experimental verification of the proposed approaches.

A review of related scientific publications

Existing research is focused on creating new clustering methods and their software implementation, where there is no straightforward process for developing an effective approach. Therefore, each study is verified based on a variety of datasets from different directions, allowing the investigation of the characteristics of each approach.

In the scientific research by [9, 10], a technique for improving the k-means algorithm is proposed, which eliminates unnecessary distance calculations and increases the speed of the algorithm's operation in a multidimensional space. The distinctive feature of the proposed approach is the high quality of the model, which equals the classical clustering algorithm but is more precise than competitors.

One of the problems of clustering is adhering to the condition of balance of estimates within clusters. Unbalanced initial estimates complicate the process of creating state security level clustering algorithms. Therefore, in the work by [11, 12], active efforts were made to solve the problem of balancing estimates by creating adaptive algorithms based on insufficient samples. The idea of the algorithm [12] is as follows: the distance between data points in each cluster and cluster centroids is computed from two perspectives, and data are selected based on these distances. This principle allows achieving high performance, which was tested on 45 datasets.

Unlike the work [12], the study in [13] examines the creation of a comprehensive model for multi-objective optimization based on AdaBoost and K-means clustering. Thus, the proposed solution ensures high-quality model performance. However, the solutions proposed in [13] do not consider the aspects of a large number of clusters as done in [14].

The research work [14] studies the principles of considering a large number of clusters, particularly based on the kernel of a high-density cluster. This algorithm enables more precise determination of cluster centroids.

In addition to object grouping tasks, the development of security level prediction models using feedback principles is also studied [15]. This principle is interesting and worth focusing on in future research. However, the research in [15] is limited to technical systems and does not involve clustering of state security levels based on large datasets for model construction.

Efficient processing of large datasets is discussed in [16], where the creation of deep learning models using unsupervised structures, including variational autoencoders and Markov models, is investigated. By dividing the system into component parts, an approach to determining the optimal number of models using unsupervised learning is achieved [17]. This issue was also explored in [18], where an array of clustering models based on artificial input estimates was constructed.

Apart from research on the k-means algorithm, where determining centroids is a component, studies of Gaussian mixture models and fuzzy clustering algorithms are conducted [19]. Despite significant advantages of this approach, the authors note its drawbacks, where the final model result is sensitive to initial conditions and the number of clusters. Proposed improvements to the clustering algorithm are discussed in [20], continuing the ideas from [19]. Authors suggest using constrained k-means regularization graph models, solving the problem of determining cluster centroids and increasing model accuracy.

The accuracy of clustering algorithms is also improved by utilizing kernel density estimation [21]. This approach increased solution accuracy by 66.05% compared to analogues. However, one of the problems in the clustering process is the existence of unbalanced estimates. Therefore, to achieve even better performance metrics of clustering algorithms, [22] proposes using feature weights to regulate the influence of unbalanced features in different clusters. However, the question of unifying these ideas across different datasets, particularly for studying state security issues, remains unclear.

A solution to the problem of unifying clustering algorithms is provided in [23], where hybrid models are built, consisting of two stages. Firstly, initial data are divided into smaller clusters, then merged into real clusters. The idea of dividing data into subgroups and then merging them into clusters is innovative and reduces algorithm execution time.

Preprocessing of input data also affects the final result of clustering algorithm performance. In [24], an approach to preprocessing membership degree matrices by filling missing values expertly is proposed. As the results of the research in [24] showed, this technique is effective. On the other hand, expert judgments are a weak point as they take into account the influence of the human factor on the final result.

To eliminate the influence of the human factor on the final result of the model, the issue of creating semi-supervised fuzzy clustering algorithms based on Medoids for relational data with multiple representations is discussed in [25, 26]. In fact, the work in [26] is a continuation of the study in [23], as the problem of increasing clustering accuracy is solved by building hybrid models.

Another effective way to increase the accuracy of clustering algorithms is to use self-organizing neural network classes, such as Kohonen self-organizing maps [27]. The effectiveness of this approach has been experimentally confirmed. The advantage of this approach is the selection of optimal hyperparameters.

Existing studies [9–27] offer unique techniques for implementing each approach – this is a fact. On one hand, new methods for creating clustering algorithms are proposed, on the other hand, components of existing algorithms are improved, including data preprocessing, introduction of weight coefficients, different methods of determining cluster centroids. However, issues regarding determining the state's security level during air alarms are not fully studied. Therefore, the question of developing new proposals for clustering models of state security levels remains open.

Research methodology

The process of developing a classifier for the state's security level involved using the raw data from a previous study [28], where four rows of ratings ranging from [1, 5] were combined into a single score - the comprehensive indicator of the state's security. The volume of the initial sample under investigation reached 605 assessments of the state's security level. It was crucial to study the behavior of these ratings in conjunction with clustering algorithms, the basics of which are outlined in [29, 30].

According to the model's lifecycle, questions regarding the improvement of its mathematical operations for comparative analysis were studied. Achieving adequacy and objectivity in model construction is essential for a comparative analysis of the proposed and existing approaches; otherwise, the result will not meet expectations.

The mentioned methods were used to combine ratings into a comprehensive indicator and utilize its assessments as input to the clustering algorithm, where the analysis of ratings was carried out using histograms. Before constructing the clustering algorithms, the optimal number of clusters was determined using the elbow method.

The research methodology involved studying the basic k-means clustering algorithm with Euclidean distance metric and its modifications. The Pearson correlation coefficient was also used as a distance metric.

Additionally, other methods of improving the clustering algorithm were considered, including determining cluster centers using Kohonen maps. When choosing the optimal parameters for the Kohonen map, two conditions were applied.

Furthermore, the described techniques for building unsupervised learning models were combined into a single algorithm.

In the final result, the quality of clustering was compared using the Davies-Bouldin criterion. Special attention was paid to the algorithm's sensitivity, where unfiltered and filtered ratings within a certain interval were used. In case of unbalanced clusters, Random Under Sampler methods were used to balance ratings in clusters.

The software implementation was carried out using the Python programming language, with the sklearn [31] library used to build the clusterer. The MiniSom [32] library was used to build Kohonen maps, and the quality of algorithms was measured using the davies_bouldin_score metric from sklearn. Cluster balancing was performed using the imblearn library. The graphical interpretation of research results was implemented using the matplotlib library.

Formal statement of the research task

Given are security ratings of countries determined by the comprehensive indicator KSi. Construct a clustering algorithm that will group the complex ratings into k clusters, where the effectiveness of the model will be assessed based on the minimum value of the Davies-Bouldin index.

Furthermore, there is a constraint that each complex rating should belong to only one cluster.

Experimental research

According to the life cycle of state security research tools, the updating of the comprehensive indicator was carried out by using operations of summation and multiplication of the variables under investigation.

The security index of the *KS* region can be expressed by model (1):

$$KS = x_1 + x_2 + x_3 + x_4 + (x_1 \cdot x_2 \cdot x_3 \cdot x_4), \qquad (1)$$

where x_1 represents the type of unmanned aerial vehicle or missile, x_2 represents the number of launched missiles (from 1 to n), x_3 represents the number of missiles shot down (from 0 to k), x_4 represents the method of launching the object.

It is worth noting that the determined assessment of the comprehensive indicator is subjected to normalization using the linear scaling tool in the range [1, 5], which was extensively studied in [17]. The assessments of the comprehensive indicator of state security level were used to build a clustering model, which allowed determining the level of security, particularly high – cluster 1, low – cluster 0.

Table 1 presents the initial investigated sample of comprehensive indicators of the state security level.

Table 1 – **Initial investigated sample of comprehensive indicators of the state security level (existing KS and updated KS)**

№	***KS* existing**	***KS* updated**
1	1.858	1.132
2	1.030	1.001
3	1.065	1.003
…	…	…
605	1.168	1.01
SUM	843.34	655.45
RMSE	0.496	0.277

Based on the analysis of the initial table with assessments of comprehensive indicators, we observe a preference for the updated approach in terms of the standard deviation, which is 0.277, compared to the existing approach - 0.496. These two rows of comprehensive security indicators are used to build clustering models according to the research methodology. It is worth noting that reproducing the histogram of the distribution of the entire initial sample will reveal aspects of the uneven distribution, as shown in Fig. 1.

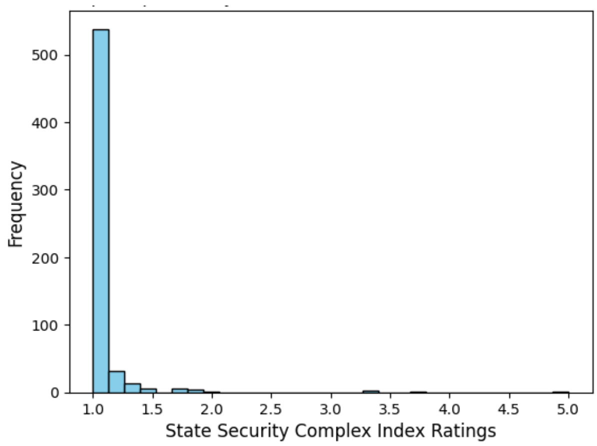

Fig. 1. Histogram of the distribution of assessments of the comprehensive indicator of the state security level without filtering

As seen from the histogram, the majority of assessments of the comprehensive indicator are located in the range from 1.0 to 1.25, with a minimum from 1.25 to 5.0. Since the clustering task involves grouping objects and does not require normal distribution, we will use the indicated assessments without filtering and with filtering to create a model of state security level clustering, as shown in Table 2.

Table 2 – **Results of building a basic clustering algorithm based on assessments determined by existing and updated methods of comprehensive evaluation of the state security level**

№	Clustering Algorithm	Number of Clusters	Existing/ Proposed Davies-Bouldin Index	Cluster Centers Existing/ Proposed
1	K-means	2	0.4765/0.2166	1.978; 1.124 / 1.061; 3.731

Based on the research results, the clustering model built on the basis of assessments from the updated comprehensive indicator of the state security level demonstrated superiority.

This is evidenced by the Davies-Bouldin Index values, which are 0.2166 compared to 0.4765 for the existing model.

However, in terms of the number of elements in the clusters, we obtained unbalanced subgroups. This is associated with the peculiarities of the input assessments of the comprehensive indicators.

Let's consider the process of finding cluster centers using Cochran's maps with the constraints specified in the research methodology, as presented in Table 3.

Table 3 – **Results of the investigation into the process of determining cluster centers using Cochran's maps based on assessments from the existing method**

Investigated Parameter	First Research Condition					Second Research Condition	
Number of Epochs	60	80	100	300	350	100	200
Sigma	0.95	1.8	0.95	0.95	0.95	1.85	1.75
Learning_rate	0.95	1.7	0.95	0.95	0.95	1.95	1.75
Optimal_width, Optimal_height	10x10	12x9	7x8	8x9	10x12	17x15	12x11
Fluctuations	0.033	0.0097	0.019	0.011	0.0109	0.0034	0.0047
Davies-Bouldin Index	0.4765	0.478	0.4765	0.4765	0.4786	0.4765	0.4765
Cluster Centers	1.978; 1.124	1.983; 1.126	1.124; 1.978	1.978; 1.124	1.126; 1.983	1.124; 1.978	1.978; 1.124

The results of building clustering models based on assessments of comprehensive indicators, where initially the search for cluster centers took place, followed by the construction of k-means clusters, showed different quality indicators. For instance, the Davies-Bouldin Index varies from 0.4765 to 0.4786 with different values of hyperparameters and parameters of Cochran's maps and epochs.

When increasing the number of epochs from 60 to 80, the Davies-Bouldin Index value increases. When using 100 epochs, the quality indicator minimizes to 0.4765, similar to that at 60 epochs. This is due to the selection of other hyperparameters and parameters of Cochran's maps. However, when using 300 and 350 epochs, we observe quality indicators at the levels of 0.4765 and 0.4786, respectively.

According to the results of the second research condition, the clustering quality does not improve either. This indicates the achievement of optimal values that the specified algorithm may demonstrate on the investigated dataset. Let's consider the peculiarities of constructing Cochran's maps for searching centroids based on assessments from the updated comprehensive indicator of the state security level, as presented in Table 4.

Based on the research results of the clustering model based on assessments from the updated comprehensive indicator of the state security level, we obtain model quality of 0.2166 for different numbers of epochs, specifically 60 and 80 epochs, respectively. As evident from the table, the cluster centers do not update, so investigations with a different number of epochs were not conducted.

To study the sensitivity of clustering models, let's consider the process of building a clusterer using filtered assessments of the comprehensive indicator from 1.0 to 1.2, as shown in Fig. 2.

The filtered investigated sample of comprehensive indicators had ratings ranging from 1.0 to 1.2. This allowed obtaining cluster centers of 1.112 and 1.009, as shown in Table 5.

Table 4 – **Results of investigating the process of determining cluster centers using Cochran's maps based on assessments from the updated comprehensive indicator of the state security level**

Investigated Parameter	First Research Condition	
Number of Epochs	60	80
Sigma	1.95	1.85
Learning_rate	1.65	1.85
Optimal_width, Optimal_height	12x8	12x13
Fluctuations	0.0027	0.0011
Davies-Bouldin Index	0.2166	0.2166
Cluster Centers	1.061; 3.731	3.731; 1.061

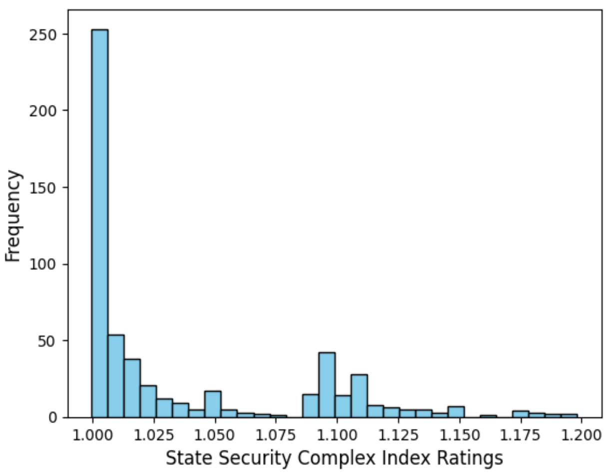

Fig. 2. Histogram of the distribution of assessments of the comprehensive indicator of the state security level with filtering from 1.0 to 1.2, N=565 assessments

As a result of the investigation of the filtered sample, a quality indicator of 0.2935 was obtained, with 414 and 151 ratings in each cluster, which is considered optimal for this set of assessments. It is noteworthy that the quality indicator value has increased compared to the previous study. This is due to the characteristics of the filtered assessments of the comprehensive indicator of the state security level and indicates the sensitivity of the method. Thus, differentiation of the investigated sample is possible based on the criterion of values that are outliers and do not carry value.

Table 5 – **Results of building a clustering algorithm based on assessments determined by the updated method of comprehensive evaluation of the state security level**

№	Clustering Algorithm	Number of Clusters	Davies-Bouldin Index	Cluster Centers
1	K-means	2	0.2935	1.112; 1.009

The clusters have 414 and 151 ratings of comprehensive indicators, indicating signs of uneven grouping or imbalance.

The problem of balancing clusters was addressed according to the research methodology using methods from the imblearn library, particularly the Random Under Sampler.

The use of the model allowed reducing the volume of the zero cluster and increasing the volume of the first cluster. As a result, we obtained 151 ratings in each cluster, with the larger cluster having ratings canceled out, as shown in Table 6.

Table 6 – **Results of building a clustering algorithm based on assessments determined by the updated method of comprehensive evaluation of the state security level using the Random Under Sampler model**

№	Clustering Algorithm	Number of Clusters	Davies-Bouldin Index	Cluster Centers
1	K-means	2	0.289	1.112; 1.009

The results of balancing the clusters showed the unchanged centroids of 1.112 and 1.009 and a decrease in the Davies-Bouldin Index, which is logical and explained by the decrease in the sample size, as depicted in Fig. 3.

As seen from the histogram, the assessments of the first cluster range from 1.0 to 1.06, while those of the second cluster range from 1.06 to 1.2. This allows decision-making regarding the state security level.

One limitation of the existing study is the use of only two clusters, where increasing their number requires forming a larger investigated sample.

On the other hand, the procedure for constructing the proposed clustering model is universal and will find its place in other research directions.

The comprehensive indicator constructed based on linear scaling principles is capable of normalizing assessments from different scales to the specified ranges.

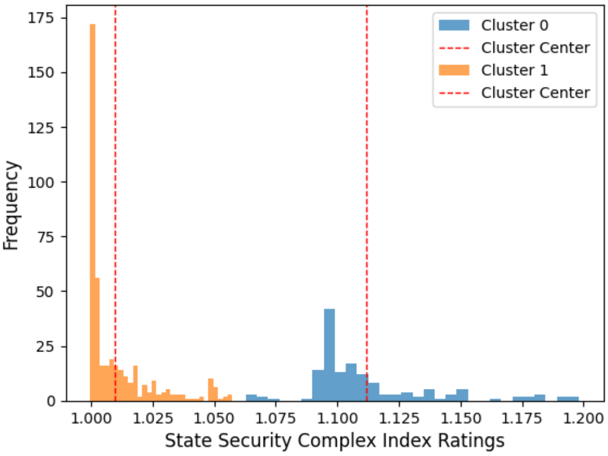

Fig. 3. Histogram of the distribution of assessments of the comprehensive indicator of the state security level (zero and first clusters)

Conclusions

1. Improving the comprehensive indicator of the state security level is achieved by adding the multiplication operation to the existing model, allowing for the consideration of the interaction of all components of the investigated object. The proposed solution demonstrates a lower value of the standard deviation compared to the existing approach (0.277 versus 0.496, respectively).

2. The task of developing an unsupervised learning model is addressed by using the comprehensive indicator, whose assessments are inputted into the k-means clustering algorithm. This approach is further enhanced by various methods for finding cluster centers and metrics for determining distances between clusters, including the Pearson correlation coefficient.

3. Experimental verification of the improved unsupervised learning model confirmed the effectiveness of the proposed solutions, as indicated by the Davies-Bouldin Index. The existing clustering model demonstrates a value of 0.4765, while the proposed one achieves 0.2166.

REFERENCES

1. Ikotun, A. M., Ezugwu, A. E., Abualigah, L., Abuhaija, B. and Heming, J. (2022), “K-means Clustering Algorithms: A Comprehensive Review, Variants Analysis, and Advances in the Era of Big Data”, *Information Sciences*, doi: https://doi.org/10.1016/j.ins.2022.11.139

2. Krasnobayev V., Yanko A. and Hlushko A. (2023), “Information Security of the National Economy Based on an Effective Data Control Method”, *Journal of International Commerce, Economics and Policy*, vol. 14, no. 03, 2350021, doi: https://doi.org/10.1142/s1793993323500217

3. Kuchuk, G., Kovalenko, A., Komari, I.E., Svyrydov, A. and Kharchenko, V. (2019), “Improving Big Data Centers Energy Efficiency: Traffic Based Model and Method”, In: Kharchenko V., Kondratenko Y., Kacprzyk J. (eds) Green IT Engineering:

Social, Business and Industrial Applications. *Studies in Systems, Decision and Control*, vol 171, Springer, Cham, doi: https://doi.org/10.1007/978-3-030-00253-4_8

4. Petrovska, I. and Kuchuk, H. (2023), "Adaptive resource allocation method for data processing and security in cloud environment", *Advanced Information Systems*, vol. *7* (3), pp. 67–73, doi: https://doi.org/10.20998/2522-9052.2023.3.10
5. Kuchuk, G., Nechausov, S. and Kharchenko, V. (2015), "Two-stage optimization of resource allocation for hybrid cloud data store", *International Conference on Information and Digital Technologies*, *Zilina*, pp. 266–271, doi: http://dx.doi.org/10.1109/DT.2015.7222982
6. Onyshchenko S., Yanko A. and Hlushko A. (2023), "Improving the efficiency of diagnosing errors in computer devices for processing economic data functioning in the class of residuals", *Eastern-European Journal of Enterprise Technologies*, vol. 5, no. 4 125), pp. 63–73, doi: https://doi.org/10.15587/1729-4061.2023.289185
7. Dotsenko, N., Chumachenko, I., Galkin, A., Kuchuk, H. and Chumachenko, D. (2023), "Modeling the Transformation of Configuration Management Processes in a Multi-Project Environment", *Sustainability (Switzerland)*, vol. 15 (19), 14308, doi: https://doi.org/10.3390/su151914308
8. Shtompel, M., Prykhodko, S., Shefer, O., Halai, V., Zakharchenko, R., and Topikha, B. (2020), "Performance analysis of the bioinspired method for optimizing irregular codes with a low density of parity checks", *Eastern-European Journal of Enterprise Technologies*, vol. 6, no. 9 (108), pp. 34–41, doi: https://doi.org/10.15587/1729-4061.2020.216762
9. Kuchuk, H., Mozhaiev, O., Kuchuk, N., Tiulieniev, S., Mozhaiev, M., Gnusov, Y., Tsuranov, M., Bykova, T., Klivets, S., and Kuleshov, A. (2024), "Devising a method for the virtual clustering of the Internet of Things edge environment", *Eastern-European Journal of Enterprise Technologies*, vol. 1, no. 9 (127), pp. 60–71, doi: https://doi.org/10.15587/1729-4061.2024.298431
10. Zhang, H., Li, J., Zhang, J. and Dong, Y. (2024), "Speeding up k-means clustering in high dimensions by pruning unnecessary distance computations", *Knowledge-Based Systems*, vol. 284, 111262, doi: https://doi.org/10.1016/j.knosys.2023.111262
11. Kuchuk, N., Mozhaiev, O., Semenov, S., Haichenko, A., Kuchuk, H., Tiulieniev, S., Mozhaiev, M., Davydov, V., Brusakova, O. and Gnusov, Y. (2023), "Devising a method for balancing the load on a territorially distributed foggy environment", *Eastern-European Journal of Enterprise Technologies*, vol. 1, no. 4 (121), pp. 48–55, doi: https://doi.org/10.15587/1729-4061.2023.274177
12. Zhou, Q. and Sun, B. (2023), "Adaptive K-means clustering based under-sampling methods to solve the class imbalance problem", *Data and Information Management*, 100064, doi: https://doi.org/10.1016/j.dim.2023.100064
13. Palakonda, V., Kang, J.-M. and Jung, H. (2023), "An Effective Ensemble Framework for Many-Objective Optimization based on AdaBoost and K-means Clustering", *Expert Systems with Applications*, 120278, doi: https://doi.org/10.1016/j.eswa.2023.120278
14. Kumar, Abhimanyu, Kumar, Abhishek, Mallipeddi, R. and Lee, D.-G. (2024), "High-density cluster core-based k-means clustering with an unknown number of clusters:, A*Applied Soft Computing*, vol. 155, 111419, doi: https://doi.org/10.1016/j.asoc.2024.111419
15. Li, G., Teng, Yi., Ding, S. and Hou X. (2024), "Complex mechanical system safety prediction based on multidimensional indexes: An MBSA-PCA-BPNN method", *Engineering Failure Analysis*, vol. 159, 108130, doi: https://doi.org/10.1016/j.engfailanal.2024.108130
16. Coraça, E. M., Ferreira, J. V. and Nóbrega, E. G. O. (2023), "An unsupervised structural health monitoring framework based on Variational Autoencoders and Hidden Markov Models", *Reliability Engineering & System Safety*, vol. 231, 109025, doi: https://doi.org/10.1016/j.ress.2022.109025
17. Fathi, M. and Bolandi, H. (2024), "Unsupervised optimal model bank for multiple model control systems: Genetic-based automatic clustering approach", *Heliyon*, e25986. doi: https://doi.org/10.1016/j.heliyon.2024.e25986
18. Rodriguez, M. Z., Comin, C. H., Casanova, D., Bruno, O. M., Amancio, D. R., Costa, L. da F. and Rodrigues, F. A. (2019), "Clustering algorithms: A comparative approach", *PLOS ONE* vol. 14, no. 1, e0210236, doi: https://doi.org/10.1371/journal.pone.0210236
19. Pietrzykowski, M. (2019), "Mini-models based on soft clustering methods", *Procedia Computer Science*, vol. 159, pp. 2512–2521, doi: https://doi.org/10.1016/j.procs.2019.09.426
20. Yang, X., Zhao, W., Xu, Yu., Wang, C.-D., Li, B. and Nie F. (2019), "Sparse K-means clustering algorithm with anchor graph regularization", *Information Sciences*, vol. 667, 120504. doi: https://doi.org/10.1016/j.ins.2024.120504
21. Pu, Yu., Yao, W., Li, X. and Alhudhaif A. (2024), "An adaptive highly improving the accuracy of clustering algorithm based on kernel density estimation", *Information Sciences*, vol. 663, 120187, doi: https://doi.org/10.1016/j.ins.2024.120187
22. Yu, H., Xu, X., Li, H., Wu, Yu. and Lei B. (2024), "Semi-Supervised Possibilistic c-Means Clustering Algorithm Based on Feature Weights for Imbalanced Data", *Knowledge-Based Systems*, vol. 286, 28 February 2024, 111388, doi: https://doi.org/10.1016/j.knosys.2024.111388
23. Akhter, M. M. and Mohanty, S. K. (2023), "A fast O(NlgN) time hybrid clustering algorithm using the circumference proximity based merging technique for diversified datasets", *Engineering Applications of Artificial Intelligence*, vol. 125, 106737, doi: https://doi.org/10.1016/j.engappai.2023.106737
24. Xu, S., Hao, Z., Zhu, Yu., Wang, Z., Xiao, Yu. and Liu B. (2023), "Semi-Supervised Fuzzy Clustering Algorithm Based on Prior Membership Degree Matrix with Expert Preference", *Expert Systems with Applications*, 121812, doi: https://doi.org/10.1016/j.eswa.2023.121812
25. Gorokhovatskyi, O. and Yakovleva, O. (2024), "Medoids as a packing of orb image descriptors", *Advanced Information Systems*, vol. 8, no. 2, pp. 5–11, doi: https://doi.org/10.20998/2522-9052.2024.2.01
26. Branco, D. P. P. and de A.T. De Carvalho F. (2023), "Medoid based semi-supervised fuzzy clustering algorithms for multi-view relational data", *Fuzzy Sets and Systems*, 108630, doi: https://doi.org/10.1016/j.fss.2023.108630

27. Vishwakarma, G.K., Paul, C., Hadi, A.S. and Elsawah A.M. (2023), "An automated robust algorithm for clustering multivariate data", *Journal of Computational and Applied Mathematics*, vol. 429, 115219, September 2023, doi: https://doi.org/10.1016/j.cam.2023.115219
28. Shefer, O., Laktionov, O., Pents, V., Hlushko, A. and Kuchuk, N. (2024), "Practical principles of integrating artificial intelligence into the technology of regional security predicting, *Advanced Information Systems*, vol. 8, no. 1, pp. 86–93, doi: https://doi.org/10.20998/2522-9052.2024.1.11
29. Khoroshun, G., Ryazantsev, O. and Cherpitskyi, M. "(2023), "Clustering and anomalies of USA stock market volatility index data", *Advanced Information Systems*, vol. 7, no. 2, pp. 9–15, doi: https://doi.org/10.20998/2522-9052.2023.2.02
30. Campbell, A. (2020), *Python Guide: Clear Introduction to Python Programming and Machine Learning*, Independently Published, 278 p., available at: https://books.google.com.ua/books/about/Python_Guide.html?id=7NHYzQEACAAJ&redir_esc=y
31. (2024), "Scikit-learn: machine learning in Python – scikit-learn 1.4.1 documentation", *Scikit-learn: machine learning in Python – scikit-learn 0.16.1 documentation*, available at: https://scikit-learn.org/stable/
32. (2024), *MiniSom. PyPI*, available at: https://pypi.org/project/MiniSom/

Received (Надійшла) 12.05.2024
Accepted for publication (Прийнята до друку) 31.07.2024

ВІДОМОСТІ ПРО АВТОРІВ/ ABOUT THE AUTHORS

Лактіонов Олександр Ігорович – кандидат технічних наук, доцент кафедри автоматики, електроніки та телекомунікацій, Національний університет «Полтавська політехніка імені Юрія Кондратюка», Полтава, Україна;
Oleksandr Laktionov – Candidate of Technical Sciences, Associate Professor of the Department of automation, electronics and telecommunications, National University "Yuri Kondratyuk Poltava Polytechnic", Poltava, Ukraine;
e-mail: itm.olaktionov@nupp.edu.ua; ORCID Author ID: https://orcid.org/0000-0002-5230-524X;
Scopus ID: https://www.scopus.com/authid/detail.uri?authorId=57210360300.

Шефер Олександр Віталійович – доктор технічних наук, професор, завідувач кафедри автоматики, електроніки та телекомунікацій, Національний університет «Полтавська політехніка імені Юрія Кондратюка», Полтава, Україна;
Oleksandr Shefer – Doctor of Technical Sciences, Professor, Head of the Department of Automation, Electronics and Telecommunications, National University "Yuri Kondratyuk Poltava Polytechnic", Poltava, Ukraine;
e-mail: itm.ovshefer@nupp.edu.ua; ORCID Author ID: https://orcid.org/0000-0002-3415-349X;
Scopus ID: https://www.scopus.com/authid/detail.uri?authorId=57210203269.

Лактіонова Ірина Сергіївна– викладач кафедри загального мовознавства та іноземних мов, Національний університет «Полтавська політехніка імені Юрія Кондратюка», Полтава, Україна;
Iryna Laktionova – Lecturer of the Department of General Linguistics and Foreign Languages, National University "Yuri Kondratyuk Poltava Polytechnic", Poltava, Ukraine;
e-mail: laktionova.iryna@nupp.edu.ua; ORCID Author ID: https://orcid.org/0009-0005-6340-8761.

Галай Василь Миколайович – кандидат технічних наук, доцент кафедри автоматики, електроніки та телекомунікацій, Національний університет «Полтавська політехніка імені Юрія Кондратюка», Полтава, Україна;
Vasyl Halai – Candidate of Technical Sciences, Associate Professor of the Department of automation, electronics and telecommunications, National University "Yuri Kondratyuk Poltava Polytechnic", Poltava, Ukraine;
e-mail: itm.galayvm@nupp.edu.ua; ORCID Author ID: https://orcid.org/0000-0002-1205-7923;
Scopus ID: https://www.scopus.com/authid/detail.uri?authorId=57221716022.

Подорожняк Андрій Олексійович – кандидат технічних наук, доцент, професор кафедри комп'ютерної інженерії та **програмування**, Національний технічний університет "Харківський політехнічний інститут", Харків, Україна;
Andrii Podorozhniak – Candidate of Technical Sciences, Associate Professor, Professor of Computer Engineering and Programming Department, National Technical University "Kharkiv Polytechnic Institute", Kharkiv, Ukraine;
e-mail: andrii.podorozhniak@khpi.edu.ua; ORCID Author ID: https://orcid.org/0000-0002-6688-8407;
Scopus ID: https://www.scopus.com/authid/detail.uri?authorId=57202229410.

Впровадження моделей навчання без учителя для аналізу рівня безпеки держави

О. І. Лактіонов, О. В. Шефер, І. С. Лактіонова, В. М. Галай, А. О. Подорожняк

Анотація. Мета. Підвищення ефективності попереднього аналізу рівня безпеки держави за рахунок впровадження моделей кластеризації. **Методика.** Досліджено процес створення моделей навчання без учителя та особливості їх використання у задачах аналізу рівня безпеки держави. Розглянуто техніки створення базової моделі k-means та її удосконалення за рахунок використання кореляції Пірсона як метрики визначення відстані. Визначення центрів кластерів проводилося базовим методом й методом карт Кохрена. Оптимальним показником якості, за результатами кластеризації вважалася модель, яка демонструє мінімальне значення індексу Девіса-Болдіна. **Результати.** Розроблено удосконалену модель навчання без учителя за алгоритмом k-means для аналізу рівня безпеки держави. Модель характеризується двома кластерами, центроїди котрих визначені як 1,112; 1,009. **Наукова новизна.** Запропонована модель кластеризації рівня безпеки держави відрізняється від існуючих використанням у якості вхідних оцінок визначних на основі комплексного показника, котрий побудований на основі принципів взаємодії та емерджентності. Це дозволяє отримати переваги кластеризаційної моделі за ознакою індексу Девіса-Болдіна. Існуюча модель кластеризації демонструє значення 0,4765, а пропонована 0,2166. **Практична значимість.** Пропозиції є корисними як додатковий інструмент попереднього аналізу рівня безпеки держави під час повітряних тривог та розширюють функціонал запропонованої у попередньому дослідженні технології прогнозування.

Ключові слова: модель кластеризації; лінійне масштабування; взаємодія факторів.