EPJ Web of Conferences **344**, 01057 (2025) https://doi.org/10.1051/epjconf/202534401057
AIPTEC 2025

Optimizing MSMEs clusters using type-2 Fuzzy K-Medoids method in Sampang Madura District

Yeni Kustiyahningsih[1*], *Eza* Rahmanita[1], *Muhammad* Iqbal Firmansyah[1], *Norjihan* Abdul Ghani[2], *Jaka* Purnama[3]

[1]Program study of Information System, Faculty of Engineering, Universitas Trunodjoyo, Madura, Indonesia
[2]Department of Information System Faculty of Computer Science and Information Technology, Universiti Malaya
[3]Program study of Industrial Engineering, Faculty of Engineering, Universitas 17 Agustus 1945 Surabaya, Indonesia

Abstract. MSMEs in Madura, Indonesia, play a crucial role in driving local economic growth, yet face significant challenges such as limited capital, disparity in market information, and low technology adoption. Data disparities among MSMEs, such as disproportionate differences in assets and production capacity, indicate uncertainty in data characteristics. The research proposes the use of Fuzzy K-Medoids Type-2, which is capable of handling data ambiguity and uncertainty through fuzzy membership representation. The study using CRISP-DM approach, starting from business understanding. The data used included 1,275 MSMEs with five main variables: number of employees, production capacity, turnover, assets, and business licenses. The evaluation was conducted by comparing the Silhouette Coefficient (SC) for K-Medoids and Partition Coefficient (PC) for Fuzzy K-Medoids Type-2. The results show that the optimal configuration is obtained with a value of k = 2 in Fuzzy K-Medoids Type-2, with a PC of 0.9548. The first cluster is categorized as small MSMEs with a low average variable, while the second cluster is categorized as medium MSMEs with a higher average variable. These findings provide a strong basis for developing a more targeted and data-driven MSME development strategy.

1 Introduction

Micro, Small, and Medium Enterprises (MSMEs) play a strategic role in national economic development. Under Law No. 11 of 2020 on Job Creation, which revises Law No. 20 of 2008, MSMEs are defined more specifically in terms of assets and turnover, providing a clear legal foundation for the sector's development [1]. Beyond serving as a backbone of the domestic economy, MSMEs contribute to job creation, foster innovation, and strengthen a people-centered economy [2]. In an increasingly competitive global landscape, the role of MSMEs is ever more crucial for enhancing national competitiveness. The COVID-19 pandemic, beginning in early 2020, had a profound impact on MSMEs. Many small businesses experienced drastic revenue declines due to mobility restrictions, market closures, and weakened consumer purchasing power [3]. In addition, limited raw materials, disrupted supply chains, and increasingly difficult access to capital have worsened the situation. Some MSMEs were forced to temporarily close their businesses, and some even went out of business permanently. However, MSMEs remain the backbone of the national economy amid this crisis [4]. In the post-pandemic era, digital transformation and innovation are seen as important solutions for MSMEs to adapt, increase competitiveness, and support economic resilience [5].

The MSME sector in Madura, particularly in Sampang District, contributes significantly to the local economy [6]. Data from Sampang District Industry and Trade Office shows that there are more than 3,000 active MSMEs in various sectors, such as fisheries, livestock, industry, trade, services, and food and beverages [1]. However, the challenges faced are quite complex, ranging from limited access to capital to difficulties in obtaining market information that supports business development [7]. Another challenge that arises is the high heterogeneity of MSME characteristics, such as the number of workers, production capacity, turnover, assets, and legal status. This condition makes the manual grouping process inefficient. Conventional clustering methods such as K-Medoids tend to be rigid because they only divide data into distinct groups, without taking into account the uncertainty and overlap between MSME characteristics. This causes the clustering results to be less representative of the actual conditions in the field. To overcome these limitations, this study proposes the use of Fuzzy K-Medoids Type-2, an algorithm that combines the resilience of K-Medoids to outliers with the flexibility of Fuzzy Type-2 logic in handling uncertainty [8][4][9]. Unlike conventional K-Medoids, which only use data points as cluster centers, Fuzzy K-Medoids Type-2 takes into account the degree of second-order fuzzy membership. Thus, this method is more adaptive to noise and variations in MSME data, resulting in more realistic clustering [10][11]. In

* Corresponding author: ykustiyahningsih@trunojoyo.ac.id

addition to algorithm selection, evaluating the number of clusters is also important to optimize the clustering results. In this study, the Partition Coefficient (PC) method was used to determine the optimal number of clusters. PC measures the clarity of fuzzy membership in each cluster, with values ranging from 0 to 1. Higher values indicate more compact and clear clusters [12][13]. Previous studies have proven the effectiveness of fuzzy methods in improving clustering accuracy. Madbouly et al. utilized Apache Spark-based Fuzzy K-Medoids to process large geospatial data and successfully reduced clustering error by 51.6% compared to K-Medoids and K-Means, while increasing accuracy by 3.5% [14][15][16]. Another study by Harsemadi et al. applied Fuzzy C-Means (FCM) with Fuzzy Partition Coefficient (FPC) for customer segmentation at the STIKOM Bali Business Incubator. The results yielded seven clusters with the highest FPC value of 0.793, emphasizing the importance of FPC in determining cluster quality [15][17].

This study aims to: (1) evaluate the effectiveness of the Fuzzy K-Medoids Type-2 method in improving the accuracy of MSME clustering in Madura District by comparing it with the K -Medoids method, (2) determine the optimal number of clusters using a comparison of the Silhouette Coefficient and Partition Coefficient (PC), and (3) analyze and interpret cluster labels based on the optimal number of clusters obtained. With these results, it is hoped that this study can contribute academically to the development of clustering methods, while also providing practical benefits for local governments and business actors in designing data-based MSME development strategies.

2 Research methods

The research methodology is the research steps that describe the initial steps to the end of the research. The research steps can be seen in Fig. 1.

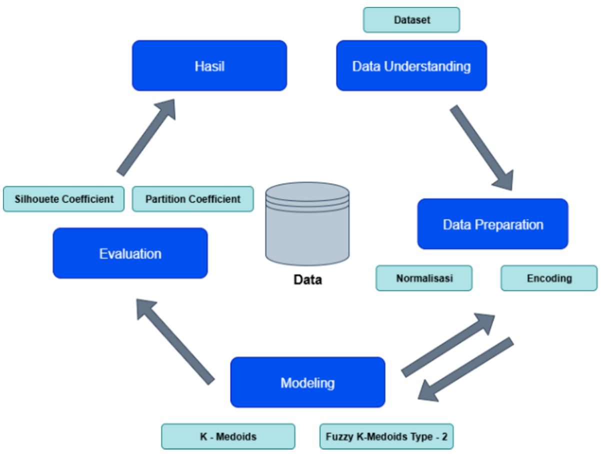

Fig. 1. Research stages.

Fig. 1. shows the research stages that follow the CRISP-DM flow, including Business Understanding, Data Understanding, Data Preparation, Modeling, and Evaluation. This process includes data normalization, encoding application, and analysis using the K-Medoids and Fuzzy K-Medoids Type-2 algorithms. Cluster quality evaluation is performed using the Silhouette Coefficient and Partition Coefficient, then implemented in the form of data visualization to facilitate analysis. Data Understanding: Data was collected from the Madura Cooperative, Industry, and Trade Office in 2023 with a total of 1,275 MSMEs.. Data Preparation: This stage is carried out to ensure that the data is ready for use in analysis. The two main steps are: Encoding: The categorical Permit variable (present/absent) was converted into a numerical variable using Label Encoding (present = 1, absent = 0). Data Normalization: Performed using the Min-Max Scaling method to standardize the variable scale in the range of 0–1, so that each variable has a balanced contribution in the clustering process. With a combination of encoding and normalization, the dataset becomes more structured and allows the algorithm to work optimally. Modeling: The algorithms used in the data clustering stage are K-Medoids and Fuzzy K-Medoids Type-2. The clustering process with K-Medoids is evaluated using the Silhouette Coefficient, while Fuzzy K-Medoids Type-2 is used to handle data uncertainty with evaluation using the Partition Coefficient.

The K-Medoids and Fuzzy K-Medoids Type-2 clustering stages include[14][4]:

1. Determine the number of clusters (k), for example k = 2
2. Calculate the initial distance of each data point to the medoid using Euclidean Distance:

$$d(x,y) = |x-y| = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \tag{1}$$

Where:
d = distance between x and y
x = cluster center data
y = data on the attribute
i = each data point
n = number of data points
x_i= data at the center of cluster i
y_i = data at each data point i

3. Form the initial cluster by grouping the data to the nearest medoid.
4. K-Medoids: Update the medoids by selecting the data with the smallest total distance, then repeat the calculation until convergence. Evaluation is performed using the Silhouette Coefficient.
5. Fuzzy K-Medoids Type-2: After the initial clusters are formed, calculate the fuzzy membership degree with the following steps:

- Calculate Type-1 fuzzy membership, which indicates the degree of membership of data in each cluster:

$$\mu_{ij}^{type-1} = \frac{1}{\sum_{k=1}^{c}\left(\frac{d_{ij}}{d_{ik}}\right)^{\frac{2}{m-1}}} \tag{2}$$

Explanation as follows:

μ_{ij}^{type-1} : Type-1 fuzzy membership degree of data i to cluster j.
d_{ij} : Euclidean distance between data i and medoid j
m : Fuzziness parameter (usually m=2).
c : Number of clusters.

- Add uncertainty (δ) to Type-2 fuzzy membership. From μ_ij, calculate the lower and upper bounds:

$$\mu_{ij}^{lower} = max\left(0, \mu_{ij}^{type-1} - \delta\right)$$
$$\mu_{ij}^{upper} = \min\left(1, \mu_{ij}^{type-1} + \delta\right) \quad (3)$$

Explanation as follows:
δ : Uncertainty adjustment parameter (e.g., 0.05 for ±5%).
μ_{ij}^{lower} : Lower Membership Function.
μ_{ij}^{upper} : Upper membership Function.

- Calculate the final membership value by combining the lower, middle, and upper limits (e.g., using the average method).
- Update the medoid by selecting the data with the smallest total distance based on Type-2 fuzzy membership.

6. Repeat the calculation until the total distance between iterations is very small or stable, indicating that the algorithm has converged.
7. Cluster Evaluation: Cluster evaluation in this study uses two methods according to the algorithm applied. For K-Medoids, the Silhouette Coefficient (SC) is used, which assesses cluster quality based on the proximity of objects within clusters (cohesion) and the separation between clusters (separation). The SC value is calculated using equation [11][15]:

$$s(i) = \frac{b(i) - a(i)}{max\ (a(i), b(i))} \quad (4)$$

Explanation:
$s(i)$: Silhouette value for data i, ranging from -1 to 1.
$a(i)$: average distance of data i to all data in the same cluster.
$b(i)$: average distance of data i to the nearest cluster.
The PC formula is defined as follows:

$$PC = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{k}\left(\frac{\mu_{ij}^{lower} + \mu_{ij}^{upper}}{2}\right)^2 \quad (5)$$

Explanation as follows:
m : Partition Coefficient.
n : Number of Data.
k : Number of Clusters.

μ_{ij}^{lower}, μ_{ij}^{upper} : Lower and upper limits of fuzzy Type-2 membership of data i to cluster j.
μ_{ij}^{middle} : Middle Membership Function calculated as $\frac{\mu_{ij}^{lower}+\mu_{ij}^{upper}}{2}$.

3 Results and discussion

3.1 MSME dataset

The dataset used in this study consists of 1,275 officially registered MSME entries. This dataset is used as the basis for data processing and analysis. A description of the batik MSME data variables can be seen in **Table 1.**

Table 1. Madura MSME Dataset 2023.

Business Name	Number of Employees	Production Capacity	Turnover	Assets	License
Abdur Rohman	2	570	12000000	5000000	Exists
Fish Farming Ahmad	3	100	5000000	1000000	Exists
LOBSTER Farming Farhan Abdillah	4	750	40000000	15000000	Exists
Fish Farming Pak Sawwir	3	555	50000000	15000000	Exists
…	…	…	…	…	…
PKL Juma'ati	2	100	2000000	5000000	Exists
Snacks Nurmala	1	300	3000000	5000000	Exists

3.2 Data preparation

The data preparation stage is carried out to ensure that the data used in the analysis is of good quality, structured, and ready to be processed by the algorithm. This process involves two main steps, namely data encoding and normalization. **Encoding :** At this stage, the Permit Letter variable, which is categorical and has two values, "Exists" and "Doesn't Exists" is encoded. To make it easier for the algorithm to process the data, these categorical values are converted into numerical values using the Label Encoding technique. **Normalization :**In this study, the normalization process was applied to several numerical columns, namely Number of Workers, Production Capacity, Turnover, Assets, Permit_Exists, and Permit_DoesNotExist. Normalization was performed using the Min-Max Scaling method, which converts the values in each column to a range between 0 and 1. The results of normalization can be seen in the following figure:

3.3 Modeling

The algorithms used in the data clustering stage are K-Medoids and Type-2 Fuzzy K-Medoids. In general, the K-Medoids process is carried out by determining the number of clusters k, calculating the distance of the data to the medoid using Euclidean Distance, grouping the data into clusters with the closest distance, then updating the medoid iteratively until the algorithm reaches convergence (stability). The evaluation of K-Medoids clustering results is done with the Silhouette Coefficient to assess the quality of cluster separation. Fuzzy K-Medoids Type-2 was developed from the K-Medoids algorithm by adding a degree of fuzzy membership. Each data point does not only belong to one cluster explicitly, but has a membership range (lower, middle, upper membership) that takes uncertainty into account. These membership values are aggregated and used in medoid updates at each iteration. The results are evaluated using the Partition Coefficient (PC) to measure the clarity of fuzzy membership and determine the optimal number of clusters. K-Medoids Euclidean Distance calculation results

Table 2. K-Medoids euclidean distance calculation.

Data Ke-	C1	C2	Proximity	*Cluster*
1	0	0,4322	0	1
2	0,2845	0,6232	0	2

- Results of the Euclidean Distance Fuzzy K-Medoids Type-2 calculation

Table 3. Euclidean distance Fuzzy K-Medoids Type-2 calculation.

Data Ke-	C1	C2	Proximity	*Cluster*
1	0	0,4322	0	1
2	0,2845	0,6232	0	1

This table shows the initial clustering results based on Euclidean distance. Data points 1 and 2 have the shortest distance to cluster center C1, so they are both assigned to Cluster 1. This table serves as the basis for initial cluster formation before fuzzy calculations are performed.

- Lower, Middle, and Upper membership values

Table 4. Lower, middle dan upper.

Data	*Lower*	*Middle*	*Upper*
Fish Farming	0,3458	0,4322	0,5186
Abdurohman Livestock	0,4986	0,6232	0,7478

This table illustrates Type-2 fuzzy membership that takes uncertainty into account with parameter α. The lower, middle, and upper values indicate the range of data membership degrees to the cluster. The higher the value, the greater the likelihood that the data belongs to the cluster.

- The results of the recalculation iterations by K-Medoids and Fuzzy K-Medoid Type-2 have the same values:

Table 5. Iteration process.

No	Iterasi	Result
1	Iterasi Ke-1	16,8662
2	Iterasi Ke-2	17,1815
3	Iterasi Ke-3	17,0671

This table shows the results of the total distance calculation (cost function) at each iteration. The values change at each iteration, and at the third iteration, the best result of 17.0671 was obtained. This indicates that the algorithm has reached a stable point (convergence), so the clustering process is complete.

3.4 Data evaluation

In the evaluation stage, an assessment was made of the clustering results that had been formed using the K-Medoids and Fuzzy K-Medoids Type-2 methods. The aim was to determine the extent to which the grouping performed was optimal and representative of the patterns found in the data.

- Silhouette coefficient

Table 6. Silhouette coefficient results.

Method	Parameter
Silhouette Coefficient	0,8075

Table 6. show a Silhouette Coefficient of 0.8075 for the final clustering configuration. This value indicates strong intra-cluster cohesion and clear inter-cluster separation, supporting the suitability of the selected model for the subsequent analysis. Accordingly, this configuration was retained as the final solution used in the study.

- Partition coefficient

Table 7. Partition coefficient result.

Method	Parameter
Partition Coefficient	0,9250

Table 7. also shows the results of evaluating the fuzzy clustering method using the partition coefficient (PC). The partition coefficient is used to measure the definitiveness of data division into clusters in the fuzzy method. The PC value ranges from 0 to 1. From the results obtained in the table.

3.5 Analysis of results

An analysis was conducted on the test results with various cluster number scenarios, starting from k = 2 to k = 10. The purpose of this trial was to evaluate the effect of changes in the number of clusters on the performance of two different evaluation methods, namely the Silhouette Coefficient (SC) in the K-Medoids algorithm and the Partition Coefficient (PC) in the Fuzzy K-Medoids Type-2 algorithm. This comparison is expected to provide an overview of the efficiency and stability of cluster formation based on each evaluation method.

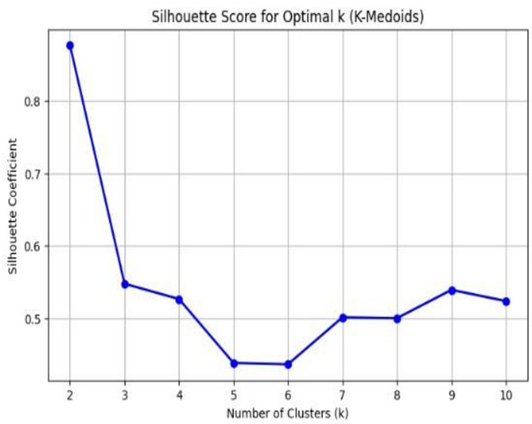

Fig. 2. Silhouette coefficient.

Fig. 2. shows a visualization of the graph, which displays changes in the Silhouette score for each k value, allowing the identification of the optimal number of clusters that produce the highest score. This optimal k value is used in the final clustering process.

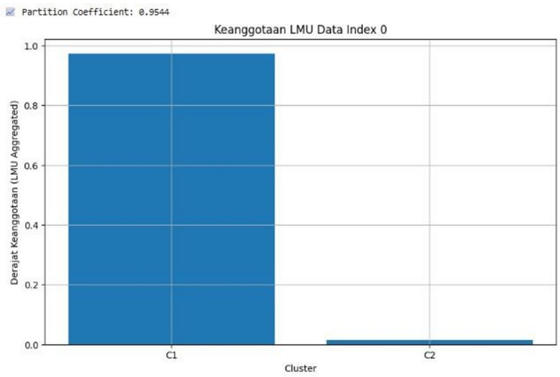

Fig. 3. Partition coefficient.

Fig. 3. shows a visualization that helps provide a more intuitive picture of the fuzzy membership distribution of the data for each cluster. These evaluation results support the analysis of the quality of fuzzy clustering performed using the Type-2 Fuzzy K-Medoids method.

Table 8. Test results for each cluster.

Number of Clusters	Silhouette Coefficient	Partition Coefficient
K = 2	0.8764	0.9548
K = 3	0.5478	0.8848
K = 4	0.5435	0.7357
K = 5	0.4390	0.6883
K = 6	0.5038	0.6510
K = 7	0.5018	0.6143
K = 8	0.5006	0.5555
K = 9	0.5059	0.4465
K = 10	0.5242	0.4464

Table 8. shows the test results for each number of clusters, with the Silhouette Coefficient and Partition Coefficient values obtained from each trial. From the table, it can be seen that the highest values for the Silhouette Coefficient of 0.8764 and the Partition Coefficient of 0.9548 were obtained when the number of clusters was k = 2. This indicates that the formation of two clusters is the most optimal configuration for the data used in this study, as it provides the clearest and most compact cluster separation compared to other clusters.

The analysis results show that:

1. Cluster 1 is dominated by MSMEs with an average number of workers of 1.38 people, an average production capacity of 290.18 units, and an average turnover of IDR 9,565,248.87. The average assets of MSMEs in this cluster are IDR 11,423,031.67, with all businesses in this cluster having obtained business licenses. Based on these indicators, Cluster 1 can be categorized as Small MSMEs.
2. Cluster 2 includes MSMEs with an average number of workers of 1.7 people, an average production capacity of 313.09 units, and an average turnover of IDR 14,555,000.00. The average assets in this cluster reached IDR 19,095,588.24, but all MSMEs in this cluster did not have a license. Based on these characteristics, Cluster 2 can be categorized as Medium MSMEs.

These results provide a clear picture of the characteristics of each cluster, which can be used as a basis for designing more targeted MSME guidance and development strategies in Madura District.

4 Conclusion

The results of the study show that the Fuzzy K-Medoids Type-2 method is superior to conventional K-Medoids in grouping MSMEs in Madura District, with a Partition Coefficient (PC) value of 0.9548 and a Silhouette Coefficient (SC) value of 0.8764 at the optimal number of clusters k = 2. The two clusters formed describe Small MSMEs with low turnover but with permits, and Medium MSMEs with higher production capacity but without permits. These findings confirm the effectiveness of Fuzzy K-Medoids Type-2 in handling heterogeneous MSME data and provide an important basis for more targeted development strategies.

References

1. D. Marcelina, A. Kurnia, T. Terttiaavini, Analisis Klaster Kinerja Usaha Kecil dan Menengah Menggunakan Algoritma K-Means Clustering. MALCOM Indones. J. Mach. Learn. Comput. Sci. **3**, 2, 293–301 (2023). https://doi.org/10.57152/malcom.v3i2.952
2. Y. Kustiyahningsih, J. Purnama, An integrated approach to determine mapping of SMEs during Covid-19 pandemic. (1945)
3. S. Rajak, K. Mathiyazhagan, V. Agarwal, K. Sivakumar, V. Kumar, A. Appolloni, Issues and analysis of critical success factors for the sustainable initiatives in the supply chain during COVID-19 pandemic outbreak in India: A case study. Res. Transp. Econ. **2021**, 101114 (2021). https://doi.org/10.1016/j.retrec.2021.101114
4. B. K. Khotimah, F. Agustina, O. R. Puspitarini, A. D. C. Cahyani, Y. Kustiyahningsih, D. R. Anamisa, Hyperparameters and Centroid Improvements in the K-Medoids Method for Grouping Processed Beef SMEs. Commun. Math. Biol. Neurosci. **2024**, 1–21 (2024). https://doi.org/10.28919/cmbn/8369
5. A. Ortiz, L. Raymond, Enabling innovation in the face of uncertainty through IT ambidexterity: A fuzzy set qualitative comparative analysis of industrial service SMEs. Int. J. Inf. Manage. **50**, 244–260 (2020). https://doi.org/10.1016/j.ijinfomgt.2019.05.007
6. M. Z. Kurniawan, M. B. S. Gitayuda, The Role of Financial Inclusion in the Development of MSMEs in Madura. EKONIKA J. Ekon. Univ. Kadiri **6**, 2, 154 (2021). https://doi.org/10.30737/ekonika.v6i2.1411

7. A. Hidayah, Implementing Data Clustering to Identify Capital Allocation for Small and Medium Sized Enterprises (SMEs). ASEAN Mark. J. **10**, 1, – (2021). https://doi.org/10.21002/amj.v10i1.10627
8. F. Fathurrahman, S. Harini, R. Kusumawati, Evaluasi Clustering K-Means dan K-Medoid pada Persebaran Covid-19 di Indonesia dengan Metode Davies-Bouldin Index (DBI). J. Mnemon. **6**, 2, 117–128 (2023). https://doi.org/10.36040/mnemonic.v6i2.6642
9. M. Shakoor, A k-medoids-based partitioned method for computational homogenization of heterogeneous materials. Comput. Struct. **316**, 107875 (2025). https://doi.org/10.1016/j.compstruc.2025.107875
10. N. Sureja, B. Chawda, A. Vasant, An improved K-medoids clustering approach based on the crow search algorithm. J. Comput. Math. Data Sci. **3**, 100034 (2022). https://doi.org/10.1016/j.jcmds.2022.100034
11. H. L. Ramadhania, L. Zakaria, Aplikasi Metode Silhouette Coefficient, Metode Elbow dan Metode Gap Statistic dalam Menentukan K Optimal pada Analisis K-Medoids. **4**, 1, 1–10 (2023).
12. N. I. Kalla, S. Annas, M. Fahmuddin, Metode Subtractive Fuzzy C-Means (SFCM) dalam Pengelompokan Kabupaten/Kota di Provinsi Sulawesi Selatan Berdasarkan Indikator Kemiskinan. VARIANSI J. Stat. Its Appl. Teach. Res. **4**, 2, 95–108 (2022). https://doi.org/10.35580/variansiunm25
13. R. Rustiyan, M. Mustakim, Penerapan Algoritma Fuzzy C Means untuk Analisis Permasalahan Simpanan Wajib Anggota Koperasi. J. Teknol. Inf. Ilmu Komput. **5**, 2, 171–176 (2018). https://doi.org/10.25126/jtiik.201852605
14. A. S. C. Martins, L. R. de Araujo, D. R. R. Penido, K-Medoids clustering applications for high-dimensionality multiphase probabilistic power flow. Int. J. Electr. Power Energy Syst. **157**, 109861 (2024). https://doi.org/10.1016/j.ijepes.2024.109861
15. R. Zheng, Y. Bao, L. Zhao, L. Xing, Prediction of steelmaking process variables using K-medoids and a time-aware LSTM network. Heliyon **10**, 12, e32901 (2024). https://doi.org/10.1016/j.heliyon.2024.e32901