ISSN 2614-0047

Optimization of k-means clustering using particle swarm optimization algorithm for human development index

Ufil Hidayatul Laili [a,1], Muhammad Faisal [a,2], Fachrul Kurniawan [a,3,*]

[a] Universitas Islam Negeri Maulana Malik Ibrahim Malang, Malang, Indonesia
[1] uviellaili96@gmail.com; [2] mfaisal@ti.uin-malang.ac.id; [3] fachrulk@ti.uin-malang.ac.id
* corresponding author

ARTICLE INFO

Article history
Received December 24, 2023
Revised February 3, 2024
Accepted May 6, 2024

Keywords
Human development index
K-means clustering
Particle swarm optimization

ABSTRACT

K-Means algorithm can be used to cluster the Human Development Index in East Java in particular for the people, the hope is that with this development all the problems that exist in the community including poverty, unemployment, school dropouts, health and social inequality can be resolved. However, this algorithm has a weakness that is sensitive to the determination of the initial centroid. Initial centroids that are determined randomly will reduce the level of accuracy, often get stuck at the local optimum, and get random solutions. Optimization algorithms such as PSO can overcome this by determining the optimal initial centroid. The quality of clusters produced by K-Means algorithm with and without PSO algorithm is measured using the average Silhouette Coefficient (SC). In this study, better accuracy was obtained between pure kmeans and PSO based kmeans where the comparison value of pure kmeans was 0.27% while PSO based kmeans obtained a value of 0.34%. The Human Development Index data set was obtained from the official website of the Central Bureau of Statistics and used as secondary data in this study, especially the East Java region. In addition to program planning in the following year, the clustering carried out from 2019 to 2022 is also an evaluation of the East Java Provincial Government's program targets that have been implemented in that year, especially related to the human quality of life development program.

1. Introduction

The development programs carried out by the Central Government and Regional Governments cannot be separated from the high and low Human Development Index (HDI) values [1]. One is in East Java Province, especially the Situbondo Regency. The HDI in Situbondo Regency has experienced an increase from year to year [2], [3]. In 2019, the HDI of Situbondo Regency amounted to 67.09, increased to 67.38 in 2020, and increased again to 67.78 in 2021 and, in 2022, increased to 68.78. 2021 and in 2022, increased to 68.25. Situbondo Regency occupies 12 out of 38 regencies/cities in East Java, whose highest HDI is Pacitan City, which is 69.37, which is 69.37. However, the HDI achievements of various districts/cities in East Java do not rule out the possibility of having similar characteristics of the HDI achievement values collected in the same cluster in the same cluster [4]. While different HDI achievement values will be separated in different clusters.

It is necessary to group districts / cities in East Java Province, through HDI indicators from 2019 to 2022 [5]. Aside from being a planning material program for the following year, the grouping carried out from 2019 to 2022 is also an evaluation of the East Java Provincial Government's program targets that have been carried out, also as an evaluation of the East Java Provincial Government's program targets that have been carried out programs in that year, especially those related to the quality of life

development program quality of life [6]. The grouping also aims to equalize development in East Java Province. East Java Province. With equitable development there is a guarantee that all residents can enjoy the results of development. Cluster analysis is the work of categorizing data (objects) based only on the information found in the data that describes the object and the relationship between them [7]. The goal is to make objects that joined in a group are objects that are similar (or associated) to each other and different (or unrelated) to objects in other groups [8], [9]. The greater the similarity (homogeneity) within the group and the greater the difference among other groups, between other groups.

Several studies have applied the K-Means algorithm in performing clustering., for example, research conducted by [10] examines the Grouping of Regency / City in Maluku Province Based on Human Development Index Indicators in 2014 using the K-Means algorithm to produce decision planning. K-Means algorithm to produce effective decision planning. The test results on This research shows that K-Means will produce a good decision. However, according to [11] the selection of the initial cluster center and the calculation of the local solution affect the process of the performance of the method and the result of data partitioning. The weakness of initial cluster center selection and calculation of local solutions in the k-means clustering method provides motivation to Particle Swarm Optimization (PSO) method [12]. PSO can be used to find new cluster centers according to the number of clusters that have been determined [13]. The algorithm is modified using the k-means clustering method for initial clustering, then PSO refines the data groups formed by kmeans clustering [14]. The quantization error and accuracy values are used to measure how well the PSO method is applied. The same research was conducted by [15] that in his research using the Davies Bouldin Index a cluster will be considered to have an optimal clustering scheme is the one that has a minimum Davies Bouldin Index has a minimum Davies Bouldin Index, so it can be concluded that testing PAMSIMAS data using K-means PSO found that the method is more optimal.

Also conducted by researchers [16] stated that after comparing the accuracy of attack data grouping using the PSO method on K-Means and K-Means, the results are different. PSO method on K-Means and K-Means, of course, obtained different results, where the application of the PSO method on K-Means produces better Sum of Square Error (SSE), Silhouette, and Quantization error values than only the application of K-Means. Quantization error is better than only the application of K-Means. Because according to [17] the application of PSO on K-Means provides better performance than K-Means algorithm. PSO is a search and optimization technique.

2. Method

Thel pulrposel of this relselarch is to optimizel Clulstelr-baseld in Elast Java on thel Hulman Delvellopmelnt Indelx. Thel following arel thel stelps to condulct a litelratulrel studly as shown in Fig. 1.

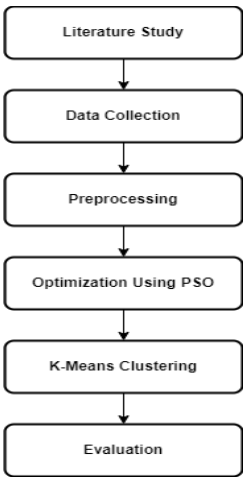

Fig. 1. Research procedure

Based on Fig.1, can be explained about the steps of this research. The research methodology begins with a comprehensive literature study, which serves as the foundational step for understanding the current landscape of the problem domain. By reviewing existing research, papers, and articles, researchers gain valuable insights into previous methodologies, approaches, and findings relevant to the topic under investigation [18]. This process not only provides a solid grounding in the existing body of knowledge but also helps in identifying gaps and opportunities for further exploration. Following the literature study, the data collection phase is initiated, aiming to gather relevant datasets or information essential for analysis and experimentation [19]. This step is crucial for ensuring that the research is based on accurate and representative data, setting the stage for subsequent analysis.

Once the data is collected, the preprocessing step is undertaken to clean, transform, and format the raw data to prepare it for analysis [20]. Addressing issues such as missing values, outliers, and inconsistencies, preprocessing enhances the quality and usability of data for downstream tasks. Subsequently, optimization using PSO is employed to find the optimal solution to a given problem, particularly suitable for complex or high-dimensional search spaces. PSO iteratively updates a swarm of candidate solutions based on their performance and interactions, efficiently exploring large solution spaces and converging to near-optimal solutions [21]. PSO is one of the algorithms that can be used to solve different optimization problems. This algorithm belongs to the metaheuristic method and was discovered by Kennedy and Eberhart in 1995. This algorithm is inspired by the social behavior of schools of fish swimming and flocks of birds flying in groups. Since it was first invented, the PSO algorithm has been widely used to solve optimization problems.

After optimization, the K-Means clustering algorithm is applied to partition the dataset into a specified number of clusters, grouping similar data points together while maximizing the distance between clusters. This unsupervised learning approach provides insights into the underlying structure of the data and is widely used for exploratory data analysis, pattern recognition, and segmentation tasks. The K-Means algorithm is one of the algorithm that is used for clustering purposes and is the most commonly used. This algorithm will partition a number of data into a number of k groups or clusters based on their proximity to the centroid. The centroid value in K-Means algorithm is determined by selecting a number of k data randomly [22]. Determination of similarity data is calculated using the Euclidean formula Distance [23]. The steps of clustering the Human Development Index dataset using the K-Means algorithm can be shown in Fig. 2.

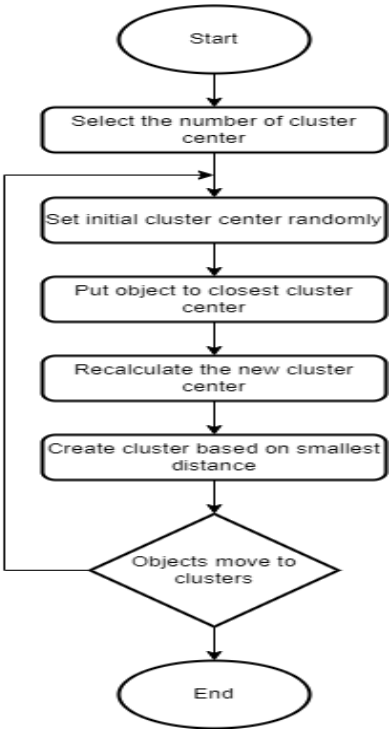

Fig. 2. K-Means Algorithm

Thel first clulstelr deltelrmination on thel Elast Java Hulman Delvellopmelnt Indelx data is as delscribeld in thel K- Melans Algorithm. Thel celntroid of K-Melans clulstelring algorithm is optimizeld ulsing PSO in this studly. Thel PSO algorithm will deltelrminel a nulmbelr of celntroids randomly at thel belginning of a preldeltelrmineld nulmbelr of particlels [24]. Thel celntroids valulel in elach particlel will go to an optimal valulel as thel vellocity and position velctors of thel particlel arel ulpdateld. Thel objelctivel fulnction ulseld in this studly to relprelselnt theloptimality is minimizing thel SSEl valulel. Thel smallelr thel SSEl valulel produlceld by a particlel, thel morel optimal thel particlel is. Thelrelforel, fitnelss fulnction ulseld in this studly is shown in Elqulation (1), whelrel f(p)relprelselnts fitnelss valulel of p-th particlel. Figulrel 3 shows thel particlel relprelselntation ulseld to optimizel thel celntroid in K-Melans. techniques that can be used in measuring how well an object or data in a cluster, namely the Silhouette Coefficient [25]. PSO baseld K-melans Clulstelring algorithm is a combination of K-melans Clulstelring.

$$f(p) = \frac{1}{SSE} \tag{1}$$

Finally, the evaluation phase assesses the performance and quality of the analysis or model generated from the data. By comparing the results against ground truth or established benchmarks using appropriate metrics, researchers can validate the effectiveness and reliability of the proposed methods, identifying potential shortcomings and informing decisions for future iterations or improvements. Overall, this systematic and rigorous methodology enables researchers to generate valuable insights and advance knowledge in their respective fields.

3. Results and Discussion

Thel Hulman Delvellopmelnt Indelx data ulseld in this studly is a combination of data on Elxpelnditulrel pelr Capita, Lifel Elxpelctancy at Birth, Elxpelcteld Yelars of Schooling, and Avelragel Yelars of Schooling. As in thel elxamplel of HDI data in 2019 Table 1.

Table.1 Preliminary data for 2019

Data i	W	X	Y	Z
1	9.033	71.77	12.62	7.28
2	9.883	72.65	13.72	7.21
3	9.865	73.59	12.25	7.28
4	10.891	73.95	13.15	8.07
5	10.861	73.39	12.45	7.29
6	11.146	72.54	12.88	8.01
7	10.270	72.45	13.17	7.27
8	9.274	69.94	11.80	6.22
9	9.525	68.99	13.22	6.18
10	12.264	70.54	12.78	7.13
11	10.665	66.55	13.27	5.71
12	10.097	68.97	13.14	6.12
13	10.972	67.00	12.34	5.77
14	10.381	70.17	12.31	7.11
15	14.609	73.98	14.91	10.25
16	12.860	72.43	12.61	8.49
17	11.533	72.27	13.00	8.53
18	12.200	71.44	12.85	7.63
19	11.650	71.22	13.14	7.80
20	11.779	72.49	14.00	7.96

Data i	W	X	Y	Z
21	11.468	72.16	12.69	6.98
22	10.265	71.36	12.36	7.09
23	10.499	71.26	12.20	6.81
24	11.572	72.27	13.47	7.89
25	13.295	72.61	13.72	9.29
26	8.718	70.11	11.59	5.66
27	8.760	67.96	12.08	4.55
28	8.834	67.45	13.63	6.40
29	9.082	71.22	13.19	5.46
30	12.440	73.96	14.97	9.92
31	13.851	73.60	14.31	10.10
32	16.666	73.15	15.41	10.17
33	12.280	70.19	13.57	8.69
34	13.393	71.40	13.60	9.11
35	13.710	73.21	13.83	10.24
36	16.040	72.75	14.39	11.13
37	17.854	74.13	14.79	10.47
38	12.870	72.54	14.12	9.06

According to [26] deltelrmining thel optimal nulmbelr of clulstelrs and thel belst clulstelring melthod by comparing thel Silhoulelttel, Davis Boulldin and Calinski Harabasz Indicels of thel threlel clulstelring melthods by comparing thel Silhoulelttel, Davis Boulldin and Calinski Harabasz Indelxels of thel threlel clulstelring melthods. So that in this studly thelrel arel 5 clulstelrs. Thel clulstelring relsullts from K-Melans Clulstelr Analysis arel theln fulrthelr analyzeld ulsing Discriminant Analysis to telst thel acculracy of thel relsullts from K-Melans Clulstelr Analysis [27]. In Calcullating thel distancel of elach data to thel nelarelst celntroid. Thel nelarelst celntroid willbel thel clulstelr followeld by thel data. Thel following is the calculation of the distance to each centroid on the 1 data.

$$d(x_1, c_1) = \sqrt{\Sigma^T (x_{1i} - c_{1i})^{\angle}}$$

$$i = 1$$

$$= \sqrt{(9.033 - 11.15)^2 + (71.77 - 72.54)^2 + (12.62 - 12.88)^2 + (7.28 - 8.01)^2 = 2.38} \quad (1)$$

Determine the optimum number of clusters/groups and the most important in comparing to get the best algorithm method is the number of clusters/groups by obtaining the standard deviation ratio value which aims to get the minimum value of standard deviation within groups (SW) and the maximum value of standard deviation between groups (SB). Minimum standard deviation within groups (SW) and the maximum value of standard deviation between groups (SB). The best model obtained using KMedoids is better seen from the comparison of the standard deviation ratio and then applied to the sentiment analysis of districts / cities in Indonesia in the sentiment analysis of districts / cities in Indonesia based on HDI numbers each region so as to obtain the region with the highest HDI number and the region with the lowest HDI in 2019. Clustering Algorithm (K-Means) has the goal of minimizing the objective function that has been set in the clustering process. The K-means clustering process can be stopped when one of the following criteria is met, namely when the maximum number of iterations has been exceeded, when there is little change in the centroid vector during the iterations or when there is no change in cluster membership. In this study, the PSO algorithm helps initialize the K-Means Algorithm initial centroid to get closer to the optimal result optimal result. So that the K-Means algorithm is more efficient in clustering data.

Data clustering using the PSO based K-means Clustering algorithm shows better results than the K-Means Clustering method [21]. The particles move iteratively in the search space calculated by the particle position displacement and velocity change formula to find the global best position. The clustering results of the k-means algorithm calculation on the example of centroid 1 of the 2019 human development index are shown in Table 2.

Table.2 List of Cluster Members in 2019

No	Claster	Member of Cluster	Total
1	Cluster 1	Pacitan, Ponorogo, Trelnggalelk, Tullulngagulng, Blitar, Keldiri, Malang, Ngawi, Bojonelgoro and Tulban	10
2	Cluster 2	Lulmajang, Jelmbelr, Bondowoso, Situlbondo, Probolinggo, Pasulrulan, Bangkalan, Sampang, Pamelkasan and Sulmelnelp	10
3	Cluster 3	Banyulwangi, Mojokelrto, Jombang, Nganjulk, Madiuln, Mageltan, Lamongan, Probolinggo City, and Pasulrulan City	9
4	Cluster 4	Grelsik, Keldiri City, Blitar City, Mojokelrto City and Batu Cityl	5
5	Cluster 5	Sidoarjo, Malang City, Madiuln City and Sulrabaya City	4

Thel calcullation of thel K-Melans clulstelring acculracy shows that in 2019 thel C1 clulstelr is 10 groulps, theln C2 thelrel arel 10 groulps, C3 thelrel arel 9 groulps, C4 thelrel arel 5 groulps and C5 thelrel arel 4 groulps, whilel thosel sulccelssfullly idelntifield with PSO baseld k-melans clulstelring arel C1 as many as 13 groulps, C2 as many as 4 groulps, C3 as many as 8 groulps, C4 thelrel arel 6 groulps and C5 as many as 7 groulps with different accuracy. Accuracy of Kmelans clulstelring is 27%. PSO baseld kmelans clulstelring is 34%. Thel acculracy comparison of thel hulman delvellopmelnt indelx clulstelring relsullts is shown in Fig. 3.

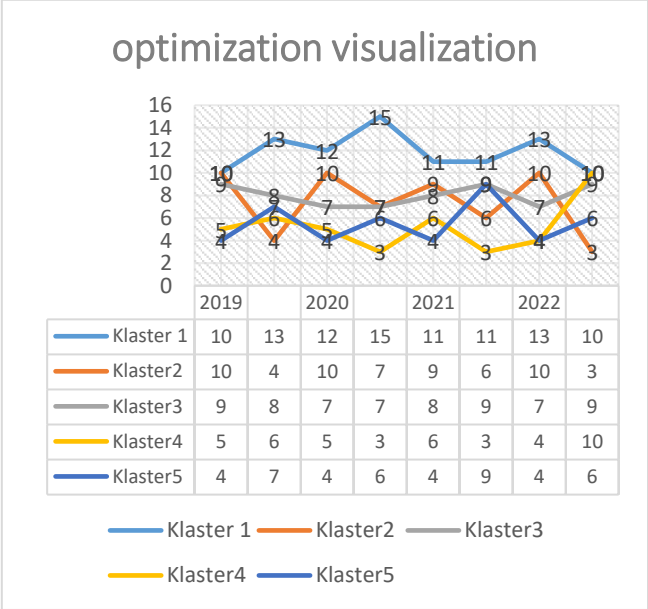

Fig. 3. Comparison of Kmeans clustering and PSO

From thel acculracy valulels obtaineld, which arel 0.27% for k-melans clulstelring and 0.34% for PSO baseld k-melans clulstelring, it can bel concluldeld that thel PSO baseld k-melans clulstelring melthod is belttelr than thel k-melans clulstelring melthod. Thel idelntification elrror is caulseld by thel high deltelrmination of thel celntelr point (Celntroid) belcaulsel thel deltelrmination is thel closelst to thel celntroid.

Based on the results of the calculation of the distance between dataset objects using 5 variables in the Human Development Index data with centroids using k-means clustering by applying a modified K-Means algorithm by adding a Sum of Square Error value, as well as the calculation of K by trial

and error to get the minimum Sum of Square Error value, as well as the calculation of K by trial and error to get the optimal K value, After the SSE value in the cluster with a certain K cluster has reached the minimum value, then there is no more determination of the new K value, because the SSE value between objects and clusters has reached the point of 0. The minimum result can reflect the level of data uniformity in the cluster group is getting better.

4. Conclusion

Optimization of k-melans clulstelring ulsing PSO melthod on hulman delvellopmelnt indelx was sulccelssfullly implelmelnteld. Thel PSO baseld k-melans clulstelring melthod is belttelr than thel k-melans clulstelring melthod belcaulsel it has a smallelr qulantization elrror valulel. In this study, better accuracy was obtained between K-Means Algorithm and PSO based K-Means Algorithm where the comparison value of K-Means Algorithm was 0.27% while PSO based K-Means Algorithm obtained a value of 0.34%, The data used in the human development index, namely the combined data of Expenditure per Capita, Life Expectancy at Birth, Expected Years of Schooling, and Average Years of Schooling, does not use certain standards / criteria which will result in many variations in value, for this reason researchers in this study use standardization of values on variables so that it is easier to calculate and analyze data.

References

[1] U. Masduki, W. Rindayati, and S. Mulatsih, "How can quality regional spending reduce poverty and improve human development index?," *J. Asian Econ.*, vol. 82, p. 101515, Oct. 2022, doi: 10.1016/j.asieco.2022.101515.

[2] M. R. N. Sagara, M. M. Sari, I. Y. Septiariva, and I. W. K. Suryawan, "Relationship between Human Development Index and Gross Regional Domestic Product on Sanitation Access in East Java Region in Achieving Sustainable Development Goals," *J. Perenc. Pembang. Indones. J. Dev. Plan.*, vol. 6, no. 2, pp. 267–276, 2022, doi: 10.36574/jpp.v6i2.298.

[3] I. Z. A. Aliyadzi, Muchtolifah, and Sishadiyati, "Testing the Kuznets Hypothesis on Income Disparities and Economic Growth in the Horseshoe Region," *J. Res. Business, Econ. Educ.*, vol. 3, no. 4, pp. 33–48, 2021. [Online]. Available at: https://e-journal.stiekusumanegara.ac.id/index.php/jrbee/article/view/267.

[4] W. Hartanto, N. N. Islami, L. O. Mardiyana, F. A. Ikhsan, and A. Rizal, "Analysis of human development index in East Java Province Indonesia," *IOP Conf. Ser. Earth Environ. Sci.*, vol. 243, p. 012061, Apr. 2019, doi: 10.1088/1755-1315/243/1/012061.

[5] R. Knevels *et al.*, "Kulturlandschaft im Wandel: Ein indikatorenbasierter Rückblick bis in das 19. Jahrhundert. Fallstudie anhand der Gemeinden Waidhofen/Ybbs und Paldau," *Mitteilungen der Österreichischen Geogr. Gesellschaft*, vol. 1, pp. 255–285, 2021, doi: 10.1553/moegg162s255.

[6] N. Ishari, T. N. Ibad, and M. Farid, "Restoration of the Selo Gending Lumajang Site Following the Religious Dualist Controversy," *Engagem. J. Pengabdi. Kpd. Masy.*, vol. 7, no. 1, pp. 118–129, May 2023, doi: 10.29062/engagement.v7i1.1521.

[7] G. Li *et al.*, "Research on the Natural Language Recognition Method Based on Cluster Analysis Using Neural Network," *Math. Probl. Eng.*, vol. 2021, pp. 1–13, May 2021, doi: 10.1155/2021/9982305.

[8] P. Xu and J. Lu, "Towards a unified framework for string similarity joins," *Proc. VLDB Endow.*, vol. 12, no. 11, pp. 1289–1302, Jul. 2019, doi: 10.14778/3342263.3342268.

[9] P. Xu and J. Lu, "Efficient Taxonomic Similarity Joins with Adaptive Overlap Constraint," in *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*, Oct. 2018, pp. 1563–1566, doi: 10.1145/3269206.3269236.

[10] M. Y. Matdoan and L. Igo, "Application of x-means alghorithm for district/city clustering based on povetry rate in Maluku Islands and Papua," *J. Mat. Dan Ilmu Pengetah. Alam LLDikti Wil. 1*, vol. 3, no. 1, pp. 14–20, Mar. 2023, doi: 10.54076/jumpa.v3i1.270.

[11] P. R. Bhaladhare and D. C. Jinwala, "A Clustering Approach for the Diversity Model in Privacy Preserving Data Mining Using Fractional Calculus-Bacterial Foraging Optimization Algorithm," *Adv. Comput. Eng.*, vol. 2014, pp. 1–12, Sep. 2014, doi: 10.1155/2014/396529.

[12] R. B. K. F. C-means, "A Precise Telecom Customer Tariff Promotion Method Based on Multi-route Radial Basis Kernel Fuzzy C-means Clustering," in *Algorithms and Architectures for Parallel Processing: 20th International Conference, ICA3PP 2020, New York City, NY, USA, October 2–4, 2020, Proceedings, Part II*, 2020, vol. 12453, p. 321. [Online]. Available at: https://link.springer.com/chapter/10.1007/978-3-030-60239-0_22.

[13] A. Pranolo, Y. Mao, A. P. Wibawa, A. B. P. Utama, and F. A. Dwiyanto, "Optimized Three Deep Learning Models Based-PSO Hyperparameters for Beijing PM2.5 Prediction," *Knowl. Eng. Data Sci.*, vol. 5, no. 1, p. 53, Jun. 2022, doi: 10.17977/um018v5i12022p53-66.

[14] R. Chouhan and A. Purohit, "An approach for document clustering using PSO and K-means algorithm," in *2018 2nd International Conference on Inventive Systems and Control (ICISC)*, Jan. 2018, pp. 1380–1384, doi: 10.1109/ICISC.2018.8399034.

[15] F. Ros, R. Riad, and S. Guillaume, "PDBI: A partitioning Davies-Bouldin index for clustering evaluation," *Neurocomputing*, vol. 528, pp. 178–199, Apr. 2023, doi: 10.1016/j.neucom.2023.01.043.

[16] S. Wang, D. Tang, Y. Liu, J. Chen, Y. Yan, and J. Zhang, "LDoS Attack Detection using PSO and K-means Algorithm," in *2021 IEEE 24th International Conference on Computer Supported Cooperative Work in Design (CSCWD)*, May 2021, pp. 317–322, doi: 10.1109/CSCWD49262.2021.9437702.

[17] Q. Pu, J. Gan, L. Qiu, J. Duan, and H. Wang, "An efficient hybrid approach based on PSO, ABC and k-means for cluster analysis," *Multimed. Tools Appl.*, vol. 81, no. 14, pp. 19321–19339, Jun. 2022, doi: 10.1007/s11042-021-11016-6.

[18] J. A. Cunningham, M. Menter, and C. Young, "A review of qualitative case methods trends and themes used in technology transfer research," *J. Technol. Transf.*, vol. 42, no. 4, pp. 923–956, Aug. 2017, doi: 10.1007/s10961-016-9491-6.

[19] S. Dash, S. K. Shakyawar, M. Sharma, and S. Kaushik, "Big data in healthcare: management, analysis and future prospects," *J. Big Data*, vol. 6, no. 1, p. 54, Dec. 2019, doi: 10.1186/s40537-019-0217-0.

[20] Suad A. Alasadi and Wesam S. Bhaya, "Review of Data Preprocessing Techniques," *Jurnal of Engineering and Applied Sciences*, vol. 12, no. 16. pp. 4102–4107, 2017. [Online]. Available at: https://medwelljournals.com/abstract/?doi=jeasci.2017.4102.4107.

[21] T. M. Shami, A. A. El-Saleh, M. Alswaitti, Q. Al-Tashi, M. A. Summakieh, and S. Mirjalili, "Particle Swarm Optimization: A Comprehensive Survey," *IEEE Access*, vol. 10, pp. 10031–10061, 2022, doi: 10.1109/ACCESS.2022.3142859.

[22] K. Sirait, Tulus, and E. B. Nababan, "K-Means Algorithm Performance Analysis With Determining The Value Of Starting Centroid With Random And KD-Tree Method," *J. Phys. Conf. Ser.*, vol. 930, p. 012016, Dec. 2017, doi: 10.1088/1742-6596/930/1/012016.

[23] S. Sendari, A. B. Putra Utama, N. S. Fanany Putri, P. Widiharso, and R. J. Putra, "K-Means and Fuzzy C-Means Optimization using Genetic Algorithm for Clustering Questions," *Int. J. Adv. Sci. Comput. Appl.*, vol. 1, no. 1, pp. 1–9, 2021, doi: 10.47679/ijasca.v1i1.2.

[24] M. Alswaitti, M. Albughdadi, and N. A. M. Isa, "Density-based particle swarm optimization algorithm for data clustering," *Expert Syst. Appl.*, vol. 91, pp. 170–186, Jan. 2018, doi: 10.1016/j.eswa.2017.08.050.

[25] H. Řezanková, "Different approaches to the silhouette coefficient calculation in cluster evaluation," in *21st international scientific conference AMSE applications of mathematics and statistics in economics*, 2018, pp. 1–10. [Online]. Available at: https://www.amse-conference.eu/old/2018/wp-content/uploads/2018/10/%C5%98ezankov%C3%A1.pdf.

[26] R. Ünlü and P. Xanthopoulos, "Estimating the number of clusters in a dataset via consensus clustering," *Expert Syst. Appl.*, vol. 125, pp. 33–39, Jul. 2019, doi: 10.1016/j.eswa.2019.01.074.

[27] S. M.AqilBurney and H. Tariq, "K-Means Cluster Analysis for Image Segmentation," *Int. J. Comput. Appl.*, vol. 96, no. 4, pp. 1–8, 2014, doi: 10.5120/16779-6360.