Implementation of K-Means and DBSCAN algorithms: A Bibliometric Review

Fahmi Reza Ferdiansyah [1] and Rikky Wisnu Nugraha [2] and Rudy Sofian [3] and Heri Purwanto [4] and Didin Saepudin [5] and Edi Andriansyah [6]

[1,3] Department of Informatics Engineering Institut Digital Ekonomi LPKIA, Bandung, Indonesia
[2] Department of Information System, Universitas Widyatama, Bandung, Indonesia
[4,5,6] Department of Information System, Universitas Sangga Buana, Bandung, Indonesia

Abstract. K-Means and DBSCAN algorithms belong to Supervised Learning, they are one of the popular clustering algorithms used in Machine Learning/Data mining that do not need to be labeled. Both algorithms are always compared with other algorithms to find superior clusters. Bibliometric was used as a research methodology with stages using the Prisma framework, namely identification, screening, eligibility, included. The focus of this research is to find scientific articles related to the K-Means and DBSCAN algorithms. Units of analysis collected through Scopus. All articles were downloaded from 2014 to 2024, resulting in 170 scientific articles. The inspection was conducted in several stages, and the overall result was 104 articles. After careful consideration, the total number of articles considered eligible was 65. There are at least four major themes that discuss the use of K-Means and DBSCAN algorithms, namely the infrastructure, transportation, health, and education sectors. Of the four fields, health and transportation are most suitable for the implementation of the K-Means and DBSCAN algorithms. In addition, researchers use K-Means and/or DBSCAN algorithms to compare with other algorithms, the goal is to find the best clustering algorithm

Keywords: DBSCAN, K-Means, Scopus, BIBLIOMETRIC.

1 Introduction

Machine learning is divided into 3 types: Unsupervised Learning, Supervised Learning and Reinforcement Learning [1][2][3]. K-Means and DBSCAN algorithms belong to Supervised Learning, they are one of the popular clustering algorithms used in Machine Learning/Data mining [4][5][6] that do not need to be labeled [3][7]. Both algorithms are always compared with other algorithms to find superior clusters [8][9][10][11].

K-Means clustering algorithm is popularly used in data mining due to its simplicity [12], flexibility, and efficiency [12][13][14] in handling large datasets [15][16]. However, K-Means has limitations, such as the need for user-defined cluster numbers [29], sensitivity to initial cluster centers [17], and susceptibility to outliers [15][16].

V. Mardiansyah and B. A. Prasetyo (eds.), *Proceedings of the Widyatama International Conference on Engineering 2024 (WICOENG 2024)*, Advances in Engineering Research 252,
https://doi.org/10.2991/978-94-6463-618-5_21

Although K-Means has a longer history, the two algorithms have been empirically compared in terms of performance, time, cost, and effort, with DBSCAN showing superiority in identifying clusters [18][19]. In addition, this algorithm works well on data sets that contain outliers and noise, thus improving its efficiency [20][21][22]. However, the high computational complexity of DBSCAN can be a drawback when dealing with large/complex data sets [13] [23].

So in this study the aim is to answer the following research questions:

RQ1: What do most researchers discuss about clustering algorithms?

RQ2: What are the most dominant areas covered by researchers during 2014 - 2024, specifically K-Means and DBSCAN algorithms?

2 Research Method

Bibliometric is used as a research methodology with stages using the Prisma framework (www.prisma-statement.org), namely identification, screening, eligibility,included. To see the development of K-Means and DBSCAN algorithms, especially in terms of the use of these algorithms. Bibliometric analysis used with the help of VosViewer Application.

The focus of this research is to find scientific articles related to the utilization of K-Means and DBSCAN algorithms. The unit of analysis was collected through Scopus, which is a journal database that meets the standards and has an international reputation. Scopus as a database is owned and managed by a scientific publisher called Elsevier. The search was conducted with the scopus website using the keywords: “K Means” OR “K-Means” AND “DBSCAN” all articles downloaded from 2014 to 2024. 170 scientific articles were produced. The screening was conducted in several stages, including: 1) checking for duplication (9 articles), 2) checking the appropriateness of the title, subject area and theme 3) criteria for widely recognized journals and publishers. Therefore, the overall screening result was 104 articles. After careful consideration of the abstracts, 57 articles were excluded due to lack of relevance to the theme being studied. The total number of articles considered eligible was 65. The final step was to read the entire text of the articles with emphasis on the abstract, background, findings, and conclusions, where there was a relationship with the K-Means and DBSCAN algorithms

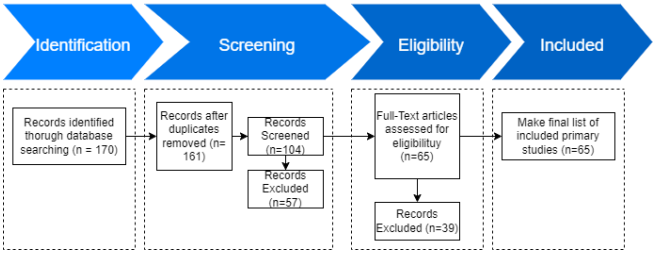

Fig. 1. Prisma Framework

3 Result and Disscussion

What most researchers discuss about clustering algorithms (RQ1). Researchers discussing clustering algorithms mainly focus on their diverse applications, challenges, and continuous development. Clustering is widely used in various domains, including marketing, healthcare [2][24][25][26], traffic [6][27][28] and data analysis [29]. The field is characterized by a large selection of algorithms, which causes confusion but also provides many algorithms for data analysis [30]. Most researchers discuss the latest algorithms and compare with the most popular algorithms such as K-Means and DBSCAN [1][5][12][14][31] .

A. Most Dominant Discussed Fields in Articles

The most dominant areas discussed by researchers during 2014 - 2024, specifically the K-Means and DBSCAN algorithms (RQ2). From 65 articles analyzed through Vosviewer software, there are at least three major themes that discuss the use of K-Means and DBSCAN algorithms, namely the infrastructure, transportation, health, and education sectors.

Table 1. Infrastructure Sector

Researchers	(Iwasaki et al. 2019)(Du and Siegel 2023)(Song, Yin, and Zhao 2023)(Sitanggang et al. 2024)(Mahata et al. 2021)(Y. Wang et al. 2020)(K. Wang et al. 2022)(J. Li et al. 2023)(Kossieris, Asgarimehr, and Wickert 2023)(Zuo et al. 2023)(Tan and Yiu 2024)
Finding	K-Means and DBSCAN are effective algorithms in various applications. In the security field, the combination of the two can reduce the warning of systems running improperly and improve the security measurement of the system. For air quality monitoring, low-cost sensors use K-Means and DBSCAN for data analysis, although these sensors are often inaccurate. In chemical processes, the combined method of K-Means, DBSCAN, and Bi-LSTM-MLP effectively detects abnormal conditions. In smog modeling, ST-DBSCAN and K-Means are used despite the limitations of real-time data. The use of K-Means and DBSCAN also helped identify land use from satellite images. In urban flood risk assessment in Fuzhou, these methods proved to be more effective than previous methods. In addition, K-Means, agglomerative, and DBSCAN were used to detect water levels using GNSS-R data. In S&C rail damage detection, K-Means and agglomerative are more suitable than DBSCAN.

Table 2. Transportation Sector

Researchers	(Farahnakian et al. 2023)(Sitanggang et al. 2024)(Joldasbayev et al. 2023; P. Li et al. 2022)(Hu and Jin 2022)(Yang et al. 2020) (Zuo et al. 2023)(Puthige et al. 2021)(Cebecauer et al. 2023)(Jian, Li, and Yu 2021)
Finding	K-Means and DBSCAN clustering methods are used to improve efficiency and safety in transportation systems. K-Means effectively detects dark ships and dangerous ship movements in maritime, while DBSCAN is used for advanced analysis. The combination of DBSCAN and K-Means (DK-Means) in China optimized demand-responsive transportation between urban and rural areas, reducing operational costs and travel time. The algorithm also analyzed online taxi booking data to identify traffic trends, weekday and weekend difference patterns, and optimal locations of taxi drivers. Another study used clustering to detect damage to railway infrastructure and analyze the safest routes in New York City based on crime statistics, helping users choose routes with lower crime risk. Overall, K-Means and DBSCAN are effective in improving the efficiency, safety, and quality of transportation services.

Table 3. Health Sector

Researchers	(Hao et al. 2021) (Gupta et al. 2022)(Taseef Hasan Farook · Tashreque Mohammed Haq · Lameesa Ramees · James Dudley 2023) (Navato and Mueller 2021) (Pourahmad et al. 2021) (Iqbal et al. 2022) (Kossakov et al. 2024)(Khan et al. 2019)(Xia et al. 2020)
Finding	Several studies have explored and compared various data clustering algorithms such as Kd-means, DBSCAN, GMM, and hierarchical methods for various applications, including diagnosis of small intestinal bacterial overgrowth (SIBO), COVID-19 contact tracing, muscle analysis during chewing activity, wastewater characterization, clustering of ovarian cancer patients, wearable-based stress monitoring, WHO TB data collection, document clustering, and racket sports gesture recognition. DBSCAN is often mentioned as an effective and easy-to-implement algorithm, but K-Means has also shown good performance in many studies. Better diagnosis and analysis criteria can be achieved by choosing an algorithm that suits the type of data and application context.

Table 4. Education Sector

Researchers	(Othman, Faye, and Hussaan 2022) (Gao et al. 2022) (Mohamed Nafuri et al. 2022)(Cahapin et al. 2023) (Valles-Coral et al. 2022) (Kalpana, Kirubakaran, and Tamije Selvy 2019)
Finding	The study compared various clustering algorithms to analyze student performance. K-Means, although simple and efficient, is less than optimal in determining cluster centers. Therefore, a combination of K-Means and DBSCAN (DB-Kmeans) is proposed to improve accuracy, especially at edge points. Experiments show that the KMoB model with K-Means has the best performance compared to and DBSCAN, with important factors such as GPA, number of activities, employment status, and drop-out status. This model achieved an accuracy of up to 99.92%. The silhouette method found

two main clusters: social science and management in determining majors, while DBSCAN produced a single cluster. DBSCAN proved to be the most robust with the best validity results. The CURE algorithm, using a hybrid SVD model for document representation, outperformed K-Means and DBSCAN in clustering. Overall, combining and improving clustering algorithms improved the accuracy and robustness of data analysis.

B. Bibliometric Analysis

After identification, screening, eligibility, and inclusion of the specified requirements, the 65 selected articles were subjected to bibliometric analysis using the Vosviewer application version 1.6.20 released in 2009-2023. The software was developed by Nees Jan van Eck and Ludo Waltman at the Center for Science and Technology Studies. The mapping is done based on the keywords that often appear, both in the title of the article and in the abstract. The visualization can be seen in Figure 2 as follows:

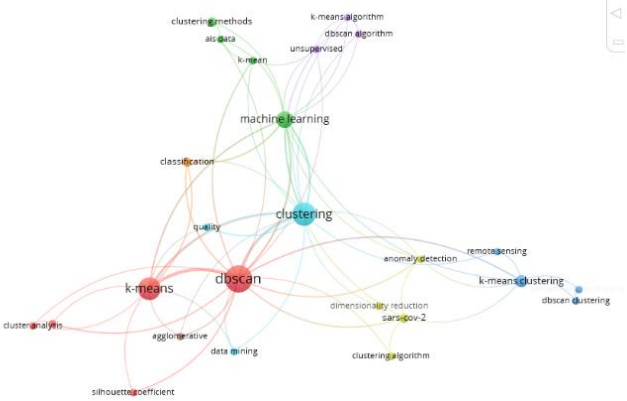

Fig. 2. Network visualization

The mapping above was formed based on the minimum requirement of keyword occurrence in both the title and abstract of the article twice, and obtained 219 keywords with 26 keywords that meet the specified threshold, which are spread into 8 clusters, which are presented in table 5 below.

Table 5. Keyword mapping cluster

Cluster	Cluster Term & Occurrences
Cluster 1	cluster analysis (2); dbscan (25);K-Means (17); outlier (2); silhouette coefficient (2)
Cluster 2	ais data(2); cluster methods(3); K-Means(17) ;machine learning(9)
Cluster 3	dbscan clustering(2); demand response transit(2); K-Means clustering (5); remote sensing(2)
Cluster 4	anomaly detection(2); clustering algorithm(2); dimensionality reduction(2); sars-cov-2(3)
Cluster 5	dbscan algorithm(2); K-Means algorithm(2); unsupervised(2)

Cluster 6	clustering(18); data mining(2); quality(2)
Cluster 7	classification(3)
Cluster 8	agglomerative (2)

While the level of keyword density can be described in the form of density visualization, where keywords that often appear are yellow, the stronger the color indicates that the keyword is widely researched and is a trending topic among researchers. While the color is green and from the outside it looks like the keyword still rarely appears in research so the opportunity for it to become future research material is still very open.

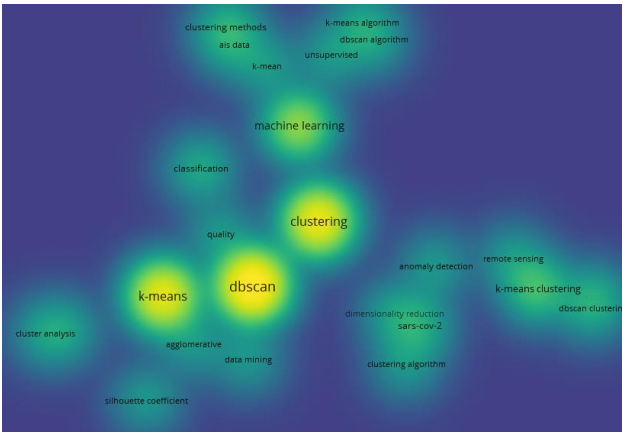

Fig. 3. Density visualization

4 CONCLUSION

The bibliometric analysis approach reveals the utilization of K-Means and/or DBSCAN algorithms in at least four areas of concern to researchers, namely education, infrastructure, transportation and health. Of the four fields, health and transportation are most suitable for the implementation of the K-Means and DBSCAN algorithms. In addition, researchers use K-Means and/or DBSCAN algorithms to compare with other algorithms, the goal is to find the best clustering algorithm.

Acknowledgments. We would like to express our sincere gratitude to Universitas Widyatama for organizing the 1st Widyatama International Conference on Engineering 2024. Our heartfelt appreciation is also extended to the lecturer for their dedication and support in making this conference a success.

References

1. X. Z. 6 Ruhao Liu 1 , Lei Zhang 1, 2,*, Xinrui Wang 3, Xuejuan Zhang 2, Xingzhou Liu 4, Xin He 5 and D. X. 7 and Z. C. 2, “Application and Comparison of Machine Learning Methods for Mud Shale Petrographic Identificatio,” 2023.

2. M. A. Valles-Coral *et al.*, "Density-Based Unsupervised Learning Algorithm to Categorize College Students into Dropout Risk Levels," *Data*, vol. 7, no. 11, 2022, doi: 10.3390/data7110165.
3. T. Iqbal, A. Elahi, W. Wijns, and A. Shahzad, "Exploring Unsupervised Machine Learning Classification Methods for Physiological Stress Detection," *Front. Med. Technol.*, vol. 4, no. March, pp. 1–12, 2022, doi: 10.3389/fmedt.2022.782756.
4. S. Iwasaki, T. Kakuta, Y. Sekiguchi, Y. Konishi, T. Ohtori, and N. Komoda, "A clustering-based judgment method of false positive alerts," *Multi Conf. Comput. Sci. Inf. Syst. MCCSIS 2019 - Proc. Int. Conf. Big Data Anal. Data Min. Comput. Intell. 2019 Theory Pract. Mod. Comput. 2019*, pp. 174–180, 2019, doi: 10.33965/tpmc2019_201907l022.
5. B. Wei, J. Zhang, C. Hu, and Z. Wen, "A Clustering Visualization Method for Density Partitioning of Trajectory Big Data Based on Multi-Level Time Encoding," *Appl. Sci.*, vol. 13, no. 19, 2023, doi: 10.3390/app131910714.
6. F. Farahnakian *et al.*, "A Comprehensive Study of Clustering-Based Techniques for Detecting Abnormal Vessel Behavior," *Remote Sens.*, vol. 15, no. 6, pp. 1–34, 2023, doi: 10.3390/rs15061477.
7. A. F. Mohamed Nafuri, N. S. Sani, N. F. A. Zainudin, A. H. A. Rahman, and M. Aliff, "Clustering Analysis for Classifying Student Academic Performance in Higher Education," *Appl. Sci.*, vol. 12, no. 19, 2022, doi: 10.3390/app12199467.
8. K. C. Alvarez-Uribe, E. Gañan-Cardenas, and D. Perez-Montoya, "Hybrid Clustering Strategy for Micro-hubs Location in Newspaper Distribution," *Oper. Supply Chain Manag.*, vol. 16, no. 4, pp. 473–487, 2023, doi: 10.31387/oscm0550405.
9. A. Margaris, I. Filippas, and K. Tsagkaris, "Hybrid Network–Spatial Clustering for Optimizing 5G Mobile Networks," *Appl. Sci.*, vol. 12, no. 3, 2022, doi: 10.3390/app12031203.
10. P. Kalpana, R. Kirubakaran, and P. Tamije Selvy, "Hybrid SVD model for document representation," *Int. J. Eng. Adv. Technol.*, vol. 8, no. 6 Special Issue 3, pp. 1147–1150, 2019, doi: 10.35940/ijeat.F1191.0986S319.
11. E. M. Cherrat, R. Alaoui, and H. Bouzahir, "Improving of fingerprint segmentation images based on K-Means and DBSCAN clustering," *Int. J. Electr. Comput. Eng.*, vol. 9, no. 4, pp. 2425–2432, 2019, doi: 10.11591/ijece.v9i4.pp2425-2432.
12. X. Gao, X. Ding, T. Han, and Y. Kang, "Analysis of influencing factors on excellent teachers' professional growth based on DB-Kmeans method," *EURASIP J. Adv. Signal Process.*, vol. 2022, no. 1, 2022, doi: 10.1186/s13634-022-00948-2.
13. Y. Zhang, "Application of nonlinear clustering optimization algorithm in web data mining of cloud computing," *Nonlinear Eng.*, vol. 12, no. 1, 2023, doi: 10.1515/nleng-2022-0239.
14. E. S. Othman, I. Faye, and A. M. Hussaan, "Automatic Clustering of Students by Level of Situational Interest Based on Their EEG Features," *Appl. Sci.*, vol. 12, no. 1, 2022, doi: 10.3390/app12010389.
15. B. Chong, "K-Means clustering algorithm: a brief review," *Acad. J. Comput. Inf. Sci.*, vol. 4, no. 5, pp. 37–40, 2021, doi: 10.25236/ajcis.2021.040506.
16. S. E. Y. Nayini, S. Geravand, and A. Maroosi, "A novel threshold-based clustering method to solve K-Means weaknesses," *2017 Int. Conf. Energy, Commun. Data Anal.*

Soft Comput. ICECDS 2017, pp. 47–52, 2018, doi: 10.1109/ICECDS.2017.8389496.
17. X. Shi, W. Wang, and C. Zhang, "An empirical comparison of latest data clustering algorithms with state-of-the-art," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 5, no. 2, pp. 410–415, 2017, doi: 10.11591/ijeecs.v5.i2.pp410-415.
18. H. K. Kanagala and J. R. Krishnaiah, "A COMPARATIVE STUDY OF K-MEANS, DBSCAN AND OPTICS," *2016 Int. Conf. Comput. Commun. Informatics*, pp. 1–6, 2016.
19. T. Tugay Bilgin and Y. Çamurcu, "DBSCAN, OPTICS ve K-Means Kümeleme Algoritmalarının Uygulamalı Karşılaştırılması," *Politek. Derg. J. Polytech. Cilt*, vol. 8, no. 2, pp. 139–145, 2005.
20. H. E. G. Lopes and M. de S. Gosling, "Cluster Analysis in Practice: Dealing with Outliers in Managerial Research," *Rev. Adm. Contemp.*, vol. 25, no. 1, pp. 1–19, 2021.
21. A. P. Navato and A. V. Mueller, "Enabling Automatic Detection of Anomalies in Wastewater: A Highly Simplified Approach to Defining 'Normal' in Complex Chemical Mixtures," *Front. Water*, vol. 3, no. December, pp. 1–8, 2021, doi: 10.3389/frwa.2021.734361.
22. H. T. Nguyen, E. H. Lee, C. H. Bae, and S. Lee, "Multiple object detection based on clustering and deep learning methods," *Sensors (Switzerland)*, vol. 20, no. 16, pp. 1–14, 2020, doi: 10.3390/s20164424.
23. M. A. Ahmed, H. Baharin, and P. N. E. Nohuddin, "Analysis of K-Means, DBSCAN and OPTICS Cluster algorithms on Al-Quran verses," *Int. J. Adv. Comput. Sci. Appl.*, vol. 11, no. 8, pp. 248–254, 2020, doi: 10.14569/IJACSA.2020.0110832.
24. S. Pourahmad, S. Foroozani, M. Nourelahi, A. Hosseini, and M. Razmkhah, "Evaluation of Twenty Genes in Prognosis of Patients with Ovarian Cancer Using Four Different Clustering Methods," *Asian Pacific J. Cancer Prev.*, vol. 22, no. 6, pp. 1781–1787, 2021, doi: 10.31557/APJCP.2021.22.6.1781.
25. M. Kossakov, A. Mukasheva, G. Balbayev, S. Seidazimov, D. Mukammejanova, and M. Sydybayeva, "Quantitative Comparison of Machine Learning Clustering Methods for Tuberculosis Data Analysis †," *Eng. Proc.*, vol. 60, no. 1, 2024, doi: 10.3390/engproc2024060020.
26. M. Gupta, R. Kumar, S. K. Chawla, S. Mishra, and S. Dhiman, "Clustering based Contact Tracing Analysis and Prediction of SARS-CoV-2 Infections," *EAI Endorsed Trans. Scalable Inf. Syst.*, vol. 9, no. 35, pp. 1–9, 2022, doi: 10.4108/eai.3-11-2021.171756.
27. I. S. Sitanggang, Apriliantono, L. Syaufina, M. A. Agmalaro, and H. Khotimah, "Haze Trajectory Simulation System from Forest and Land Fires in Indonesia Using HYSPLIT," *Environ. Ecol. Res.*, vol. 12, no. 1, pp. 54–65, 2024, doi: 10.13189/eer.2024.120106.
28. D. Hu and W. Jin, "Flex-route demand response transit scheduling based on station optimization," *Shenzhen Daxue Xuebao (Ligong Ban)/Journal Shenzhen Univ. Sci. Eng.*, vol. 39, no. 2, pp. 209–215, 2022, doi: 10.3724/SP.J.1249.2022.02209.
29. H. Yin, A. Aryani, S. Petrie, A. Nambissan, A. Astudillo, and S. Cao, "A Rapid Review of Clustering Algorithms," pp. 1–25, 2024, [Online]. Available: http://arxiv.org/abs/2401.07389
30. J. Yu, H. Huang, and S. Tian, "Survey of Clustering Algorithms," *Lect. Notes Comput.*

Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 3138, no. 3, pp. 957–965, 2004, doi: 10.1007/978-3-540-27868-9_105.

31. J. P. B. Andrade, J. E. B. Maia, and G. A. L. de Campos, "Centralized algorithms based on clustering with self-tuning of parameters for cooperative target observation," *Rev. Inform. Teor. e Apl.*, vol. 28, no. 2, pp. 39–49, 2021, doi: 10.22456/2175-2745.107154.
32. B. Du and J. A. Siegel, "Estimating Indoor Pollutant Loss Using Mass Balances and Unsupervised Clustering to Recognize Decays," *Environ. Sci. Technol.*, vol. 57, no. 27, pp. 10030–10038, 2023, doi: 10.1021/acs.est.3c00756.
33. X. Song, F. Yin, and D. Zhao, "Fused Data-Driven Approach for Early Warning Method of Abnormal Conditions in Chemical Process," *Processes*, vol. 11, no. 8, 2023, doi: 10.3390/pr11082435.
34. K. Mahata, R. Das, S. Das, and A. Sarkar, "Land Use Land Cover map segmentation using Remote Sensing: A Case study of Ajoy river watershed, India," *J. Intell. Syst.*, vol. 30, no. 1, pp. 273–286, 2021, doi: 10.1515/jisys-2019-0155.
35. Y. Wang, J. Tan, Z. Liu, and A. Ditta, "Lithium-ion battery screening by K-Means with DBSCAN for denoising," *Comput. Mater. Contin.*, vol. 65, no. 3, pp. 2111–2122, 2020, doi: 10.32604/cmc.2020.011098.
36. K. Wang, R. Yang, C. Liu, T. Samarasinghalage, and Y. Zang, "Extracting Electricity Patterns from High-dimensional Data: A comparison of K-Means and DBSCAN algorithms," *IOP Conf. Ser. Earth Environ. Sci.*, vol. 1101, no. 2, 2022, doi: 10.1088/1755-1315/1101/2/022007.
37. J. Li, A. Zheng, W. Guo, N. Bandyopadhyay, Y. Zhang, and Q. Wang, "Urban flood risk assessment based on DBSCAN and K-Means clustering algorithm," *Geomatics, Nat. Hazards Risk*, vol. 14, no. 1, p., 2023, doi: 10.1080/19475705.2023.2250527.
38. S. Kossieris, M. Asgarimehr, and J. Wickert, "Unsupervised Machine Learning for GNSS Reflectometry Inland Water Body Detection," *Remote Sens.*, vol. 15, no. 12, 2023, doi: 10.3390/rs15123206.
39. Y. Zuo, J. Lundberg, P. Chandran, and M. Rantatalo, "Squat Detection and Estimation for Railway Switches and Crossings Utilising Unsupervised Machine Learning," *Appl. Sci.*, vol. 13, no. 9, 2023, doi: 10.3390/app13095376.
40. X. Tan and S. M. Yiu, "Self-Adaptive Incremental PCA-Based DBSCAN of Acoustic Features for Anomalous Sound Detection," *SN Comput. Sci.*, vol. 5, no. 5, 2024, doi: 10.1007/s42979-024-02844-y.
41. P. Li, L. Jiang, S. Zhang, and X. Jiang, "Demand Response Transit Scheduling Research Based on Urban and Rural Transportation Station Optimization," *Sustain.*, vol. 14, no. 20, 2022, doi: 10.3390/su142013328.
42. S. Joldasbayev, S. Sapakova, A. Zhaksylyk, B. Kulambayev, R. Armankyzy, and A. Bolysbek, "Development of an Intelligent Service Delivery System to Increase Efficiency of Software Defined Networks," *Int. J. Adv. Comput. Sci. Appl.*, vol. 14, no. 12, pp. 644–656, 2023, doi: 10.14569/IJACSA.2023.0141267.
43. L. Yang, G. Jia, F. Wei, W. Chang, and S. Zhou, *The Trend Analysis Method of Urban Taxi Order Based on Driving Track Data*, vol. 12192 LNCS. Springer International Publishing, 2020. doi: 10.1007/978-3-030-49788-0_52.
44. I. Puthige *et al.*, "Safest route detection via danger index calculation and K-Means clustering," *Comput. Mater. Contin.*, vol. 69, no. 2, pp. 2761–2777, 2021, doi:

10.32604/cmc.2021.018128.

45. M. Cebecauer, E. Jenelius, D. Gundlegård, and W. Burghout, "Revealing representative day-types in transport networks using traffic data clustering," *J. Intell. Transp. Syst. Technol. Planning, Oper.*, vol. 0, no. 0, pp. 1–24, 2023, doi: 10.1080/15472450.2023.2205020.
46. S. Jian, D. Li, and Y. Yu, "Research on Taxi Operation Characteristics by Improved DBSCAN Density Clustering Algorithm and K-Means Clustering Algorithm," *J. Phys. Conf. Ser.*, vol. 1952, no. 4, 2021, doi: 10.1088/1742-6596/1952/4/042103.
47. R. Hao, L. Zhang, J. Liu, Y. Liu, J. Yi, and X. Liu, "A promising approach: Artificial intelligence applied to small intestinal bacterial overgrowth (sibo) diagnosis using cluster analysis," *Diagnostics*, vol. 11, no. 8, 2021, doi: 10.3390/diagnostics11081445.
48. Taseef Hasan Farook · Tashreque Mohammed Haq · Lameesa Ramees · James Dudley, "Deep learning and predictive modelling for generating normalised muscle function parameters from signal images of mandibular electromyography." 2023.
49. T. H. F. Khan, N. N. Alleema, N. Yadav, S. Mishra, and A. Shahi, "Text document clustering using K-Means and dbscan by using machine learning," *Int. J. Eng. Adv. Technol.*, vol. 9, no. 1, pp. 6327–6330, 2019, doi: 10.35940/ijeat.A2040.109119.
50. K. Xia, H. Wang, M. Xu, Z. Li, S. He, and Y. Tang, "Racquet sports recognition using a hybrid clustering model learned from integrated wearable sensor," *Sensors (Switzerland)*, vol. 20, no. 6, 2020, doi: 10.3390/s20061638.
51. E. L. Cahapin, B. A. Malabag, C. S. Santiago, J. L. Reyes, G. S. Legaspi, and K. L. Adrales, "Clustering of students admission data using K-Means, hierarchical, and DBSCAN algorithms," *Bull. Electr. Eng. Informatics*, vol. 12, no. 6, pp. 3647–3656, 2023, doi: 10.11591/eei.v12i6.4849.