Data mining implementation: a survey

Wisnu Widayat[1], Priati Assiroj[1], Sohirin[1], Isidorus Anung Prabadhi[1], Pasha Adelia Kautsar[2]
[1]Immigration Polytechnic, Ministry of Law and Human Rights of the Republic of Indonesia, Depok, Indonesia
[2]Immigration Technology Management Study Program, Immigration Polytechnic, Jakarta, Indonesia

ABSTRACT

In the current era, the relentless advancement of information technology necessitates efficient information acquisition, which relies on proper data processing. To address the challenges in data organization, data mining emerges as a pivotal solution. This study aims to delve into various methodologies for data grouping. Employing a survey approach, the research scrutinizes journals published from 2020 to 2024. The findings illuminate prevalent techniques, algorithms, and software tools utilized in similar research domains. Notably, the study reveals that the predominant approach entails clustering via K-Means leveraging RapidMiner. This insight underscores the significance of employing robust methodologies and tools to streamline data processing and analysis in the contemporary information landscape. By elucidating the prevalent techniques and tools, this study contributes to enhancing understanding and fostering advancements in data mining practices, thereby facilitating more efficient data utilization and decision-making processes.
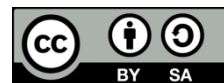
Corresponding Author:

Wisnu Widayat
Immigration Polytechnic, Ministry of Law and Human Rights of the Republic of Indonesia
Depok, West Java, Indonesia
Email: wisnu@poltekim.ac.id

1. INTRODUCTION

Information technology which is now increasingly developing is experiencing very rapid progress. This has an impact on the increasing need for accurate and useful information. Good data management is needed for these needs. Data collected in large and complex quantities requires proper management to be processed into useful information. One solution to the difficulties in grouping data is using data mining techniques. Data mining techniques can help manage data by identifying patterns and trends in data, predicting future data values, finding relationships between data, and developing decision-making models. Using data mining techniques can help organizations increase efficiency and effectiveness in data management and achieve mutual success. Ary Suryo said in his research that data can be developed with strategy [1].

Data mining involves a continuous and collaborative approach to discovering novel models and trends that are both applicable and comprehensible, particularly in large datasets. With the help of special tools, they can identify hidden patterns. Useful analysis results are the first step for further research, even when integrated with other decision support tool [2]. Various approaches can be used in data mining to analyze data and obtain useful information. Descriptive approaches, such as calculating statistics and finding trends, are used to understand data and their characteristics. The diagnostic approach helps explain why something happens in the data by identifying cause-and-effect relationships. A predictive approach helps devise more effective and efficient strategies by considering future possibilities. We can choose the most effective and efficient approach by understanding our needs and goals. The aim of this research is to examine frequently used approaches, techniques and algorithms used to implement data mining.

2. METHOD

A structural literature review regarding the application of data mining implemented through use preferred reporting items for systematic review and meta-analysis (PRISMA) method [3]. The PRISMA method is a guide that assists researchers in conducting a comprehensive and transparent systematic review. The systematic review process that is divided covers five stages, namely determine eligibility requirements, determine information references, determine literature, collect data, and determine data types [4].

a. Stage 1: defining article eEligibility criteria

Determined by inclusion criteria (IC), namely:
- IC1: articles must be based on original research that has been carried out and written in English.
- IC2: articles published in 2020 and 2024.
- IC3: this article aims to analyze the methods of other researchers to implement data mining.

b. Stage 2: define information resources
- Literature can be found through online databases in various academic collections research, including Elsevier (SCOPUS), ScienceDirect, Google Scholar, SpringOpen, and ProQuest.
- A search was also carried out on the articles that qualify for IC to find other studies related.

c. Stage 3: literature selection
- The keyword selection is "implementation of data mining"
- To peruse and select titles, summaries, and articles, keywords derived from search results that align with the requirements for this study will be reevaluated by implementing stages 3 through 4.

d. Stage 4: data collection

The data was gathered through the creation of a manual data extraction form. A total of 125.527 were evaluated using the keyword "data mining implementation" from all possible sources and criteria. Out of these, 193 articles met the initial requirements. Following additional screening, 61 articles were chosen for this study. Table 1 displays the collected data.

Table 1. Data collection

Source	Studies found (based on title and keywords) Implementation of data mining	Candidate	Selected
Elsevier (SCOPUS)	119	36	7
ScienceDirect	4.125	8	3
Google Scholar	251	80	38
SpringOpen	2.814	25	5
ProQuest	118.218	44	8
Total	125.527	193	61

e. Stage 5: data item selection

Data is obtained from selected articles containing methods or approaches to implement data mining.

3. RESULTS AND DISCUSSION

This research examines the various methods used by researchers to implement data mining. In addition, this research also analyzes the trend of "Selected Studies" literature based on the author's name, year of publication, and type of literature. Table 2 in APPENDIX shows the sources of publications.

Table 3 in APPENDIX highlights the various ways previous researchers have implemented data mining. This diverse research technique includes classification, prediction, relationships between data, and data grouping (clustering). Analysis based on Table 3 in APPENDIX in Figure 1 shows that clustering is the first most used technique in data mining implementations, followed by classification techniques, which are also widely used. These techniques have many advantages as they simplify complex data, make predictions, and increase efficiency. Analysis based on Table 3 in APPENDIX above shows the algorithms used in previous studies consist of K-Means, Linear regression, Apriori, BP Neural Network, C4.5, K-medoids, Logistic Regression, Binary, Naïve Bayes, density-based spatial clustering of applications with noise (DBSCAN), Hierarchical clustering, ARIMA-ARCH, decision tree (DT), Random Forest, Support vector machine (SVM), K-Nearest Neighbor, Fuzzy C-means, Auto Enconder, classification and regression trees (CART), Quick unbiased efficient statistical tree (QUEST), Pelican Optimizer (POA), Standar Differential Evolution, Improved Differential Evolution, and Isolation Forest. Figure 2 above shows the most widely used algorithm for data mining implementation research in this study is K-Means.

Analysis based on Table 4 in APPENDIX shows the percentage in previous studies. The results shown in Figure 3 show that 13% focus on managing healthcare and medical, 31% education, 50% focus on bussiness and finance, 4% focus on environmental and meteorological, and 2% focus on managing sports and athletics data. The result show that data mining management has potential in various categories, but its application in

business and finance is more dominant. The analysis based on Table 3 in APPENDIX above shows the software used in previous studies: RapidMiner, SPSS, Weka, Orange, Machine Learning, R, Python, Matlab, IBM Cloud Storage, and JMP. The results are shown in Figure 4, the most widely used software for data mining implementation research in this study is RapidMiner.
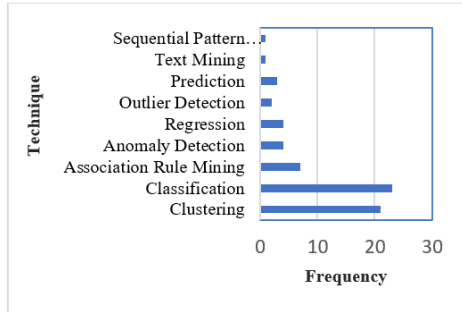

Figure 1. The most data mining techniques used for research in 2020 – 2024
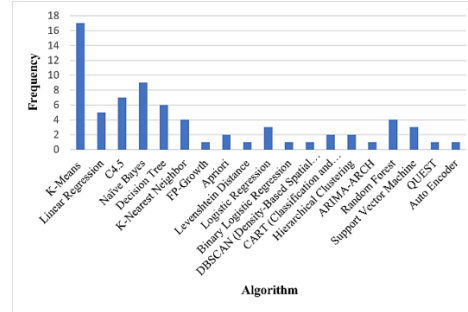

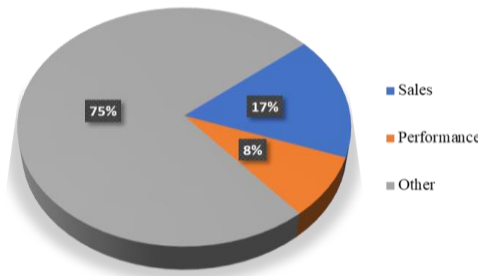Figure 2. The most algorithms used in data mining implementation in 2020 – 2024

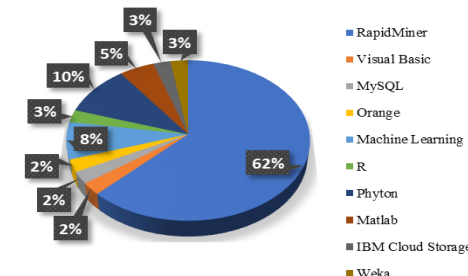Figure 3. Implementation of data mining

Figure 4. The most software tools used

4. CONCLUSION

Recent observations suggest that previous studies and research on data mining implementation have used various techniques, algorithms, and software. Our findings provide conclusive evidence that this phenomenon suggests clustering and classification techniques using the k-means algorithm and RapidMiner software are the most common. The simplicity, efficiency, and flexibility of the k-means algorithm make it suitable for analyzing complex and intricate data. Data mining will increasingly aid organizations and individuals in achieving their objectives as technology advances. However, it's worth noting that this research is limited to 2020 to 2024. Moreover, combining specific algorithms can yield more detailed and practical analyses.

APPENDIX

Table 2. Publication sources

No	Title	Year	Type
1	Implementation…[5]	2020	Journal
2	Analyzing…[6]	2020	Journal
3	Mining…[7]	2020	Journal
4	Research on…[8]	2020	Journal
5	Implementation…[9]	2020	Journal
6	Implementation of…[10]	2020	Journal
7	Data mining approach…[11]	2020	Journal
8	Implementation…[12]	2021	Journal
9	Implementation…[13]	2021	Journal
10	Implementation of data mining…[14]	2021	Journal
11	Implementation of the…[15]	2021	Journal
12	Implementation of data mining…[16]	2021	Journal
13	Design…[17]	2021	Journal

Table 2. Publication sources *(Continued)*

No	Title	Year	Type
14	Analysis…[18]	2021	Journal
15	Implementation of…[19]	2021	Journal
16	Implementation of data…[20]	2022	Journal
17	Implementation of data mining…[21]	2022	Journal
18	Implementation of data mining…[22]	2022	Journal
19	Implementation of data mining…[23]	2022	Journal
20	Implementation of data mining…[24]	2022	Journal
21	Impact…[25]	2022	Journal
22	Implementation…[26]	2022	Journal
23	Implementation of…[27]	2022	Journal
24	Implementation…[28]	2022	Journal
25	Implementation of…[1]	2022	Journal
26	The implementation…[29]	2022	Journal
27	Implementation…[30]	2022	Journal
28	Design…[31]	2022	Journal
29	Implementation of…[32]	2022	Journal
30	Development …[33]	2022	Journal
31	Architecture…[34]	2022	Journal
32	Applying…[35]	2022	Journal
33	Implementation of data mining…[36]	2023	Journal
34	The Application…[37]	2023	Journal
35	K-means…[38]	2023	Journal
36	Implementation of data mining…[39]	2023	Journal
37	Using…[40]	2023	Journal
38	Implementation of machine…[41]	2023	Journal
39	Implementation of...[42]	2023	Journal
40	Mining Data…[43]	2023	Journal
41	Implementation of data…[44]	2023	Journal
42	An application…[45]	2023	Journal
43	Implementation…[46]	2024	Journal
44	Predicting…[47]	2024	Journal
45	Efficiency…[48]	2024	Journal
46	Application of…[49]	2024	Journal
47	Identification…[50]	2024	Journal
48	Leveraging…[51]	2024	Journal

Table 3. Technique, algorithms, tools, and implementation

Author	Year	Technique	Algorithm	Tools	Implementation
Rahayu *et al.* [5]	2020	Clustering	K-means	RapidMiner	Management of rabies case data in Palembang
Ariyanto *et al.* [6]	2020	Classification	Linear regression	RapidMiner	Management data in the form of student online test scores
Bagui and Stanleey [7]	2020	Association rule mining	Apriori	-	management of frequent itemsets from streaming transaction data in the presence of drift concept.
Liang and Li [8]	2020	Classification, regression	BP neural network	MATLAB, SPSS	Management for performance evaluation of government administration personnel
Saifullah *et al.* [9]	2020	Clustering	K-means	RapidMiner	Management data on the percentage of illiterates aged 15+ by province
Alfianzah *et al.* [10]	2020	Association rule mining	Apriori	RapidMiner	Management sales data of Lakoe Dessert Pondok Kacang products
Bajdor and Pawełoszek [11]	2020	Clustering	K-means	Orange	Evaluation of the concept of sustainable entrepreneurship
Supriyono *et al.* [12]	2021	Association rule mining	Apriori	-	Management of agricultural product sales data
Sinaga *et al.* [13]	2021	Classification	C4.5	RapidMiner	Management PDAM Tirta Lihou customer satisfaction data
Jaja *et al.* [14]	2021	Classification	K-means	RapidMiner	Management WFH and WFO agent performance data
Kustandi *et al.* [15]	2021	Clustering	K-medoids	RapidMiner	Determine the average length of school
Muttaqien *et al.* [16]	2021	Classification	C4.5	RapidMiner	Predictions customer loyalty of PT Pegadaian (Persero)
Sharma and Sharma [17]	2021	Classification, Regression, Association rule mining, clustering, outlier detection	Logistic regression, linear regression	RapidMiner	Management data in the form of rainfall predictions

Table 3. Technique, algorithms, tools, and implementation *(Continued)*

Author	Year	Technique	Algorithm	Tools	Implementation
Sembiring and Tambunan [18]	2021	Classification	NB	RapidMiner	Management student academic performance data
Wang and Li [19]	2021	Clustering	DBSCAN	-	Management health and medical data
Wahyudi and Arroufu [20]	2022	Regression	Linear regression, CART	RapidMiner	Predicted delivery time
Wahyudi and Silfia [21]	2022	Clustering	K-means	RapidMiner	Sales data management in S&R Babystore
Mawaddah et al. [22]	2022	Clustering	K-means	RapidMiner	Management of oil palm yield data
Ulfah and Irtawaty [23]	2022	Clustering	K-means	RapidMiner	Data management in the form of books
Anggriawan and Gunawan [24]	2022	Clustering	K-means	R	Bike sales data management
Taylan et al. [25]	2022	Clustering	Hierarchical clustering, CART	Matlab, JMP	Management of economic data related to the COVID-19 pandemic in G20 countries
Tata et al. [26]	2022	Clustering	K-means	MATLAB	Management of BPJS employment old age Insurance claim data by region
Hendri et al. [27]	2022	Classification	C4.5	RapidMiner	Prediction of student scholarship recipients at universities
Kalas and Deshpande [28]	2022	Clustering, Association rule mining	Hierarchical clustering	-	Data management in the form of stock and shipping
Bimantoro et al. [1]	2022	Clustering	K-means	RapidMiner	Management of late payment data of shopping center tenants
Monika [29]	2022	Regression	ARIMA-ARCH	R	Data management in the form of rainfall predictions in the city of Bandung
Ige and Sikiru [30]	2022	Classification, anomaly detection	DT, RF	IBM cloud storage	Secure cloud computing environment data management
Zhang and Xie [31]	2022	Classification, clustering, association rule mining, prediction	K-means	Machine learning	Evaluation of online users learning behavior
Li [32]	2022	Clustering	K-means	-	Management of daily user consumption data in the community mall and damage repair data
Jiang [33]	2022	Clustering	K-means	-	Data management of development and implementation of STEM education
Gebremeskel et al. [34]	2022	Classification	DT	Weka, RapidMiner	Management of patient data for healthcare safety
Safdari et al. [35]	2022	Classification	SVM, RF, KNN, NB, LR, and DT	Phyton	Visualization of knowledge extraction in patient safety care
Putra [36]	2023	Classification	C4.5	RapidMiner	Management student study period data
Meng et al. [37]	2023	Clustering, regression	K-means, LR, BP neural network	-	Predicting the total box office of Chinese Kungfu films
Okereke et al. [38]	2023	Clustering	K-means, K-medoids, fuzzy C-means	-	Management data of energy consumption from smart meters
Pratama et al. [39]	2023	Classification	KNN, NB	-	Management visitor satisfaction level data at Boombara Waterpark Rantauprapat
Mydyti et al. [40]	2023	Classification, clustering, regression, association rule mining, sequential analysis	Decision tree, RF, NB, LR, KNN	Weka, Rapid Miner, Orange, R	Optimize the company's strategic goals and the relationship between the company and customers
Mahesh et al. [41]	2023	Classification, clustering	RF, auto encoder	Machine learning	Management data IDS

Table 3. Technique, algorithms, tools, and implementation *(Continued)*

Author	Year	Technique	Algorithm	Tools	Implementation
Siregar *et al.* [42]	2023	Classification, custering, Association rule mining, sequential pattern mining	CART	RapidMiner	Management data on prospective students of Harapan University Medan based on average report card, CAT test score, and interview score
Al Zahra *et al.* [43]	2023	Association rule mining	Apriori	-	Management cosmetic sales data
Putriyana and Nurdiawan [44]	2023	Classification	NB	RapidMiner	Management student graduation data
Bakator *et al.* [45]	2023	Classification, regression, prediction	Logistic regression, SVM, linear regression, Binary, DT, QUEST	Phyton	Analysis of ICT implementation and application within the company
Hasanah *et al.* [46]	2024	Clustering	K-means	MATLAB	Management of teacher performance assessment data
Zheng and Li [47]	2024	Classification	DT, BP neural network, KNN, CART, SVM	-	Prediction of student academic performance
Xie [48]	2024	Classification	DT, pelican optimizer (POA)	MATLAB	Analysis and prediction of student academic performance
Zhou and Li [49]	2024	Assosiation rule mining	Standard differential evolution, improved differential evolution	-	Management data of education in higher education
Chen [50]	2024	Clustering	K-means	Phython	Management of athlete data, including profiles, training load, physiology, and environment
Sarlis *et al.* [51]	2024	Assosiation rule mining, anomaly detection	Apriori, DBSCAN, Isolation Forest	-	Manages NBA player performance data, injury records, and salary details

Table 4. Implementation of data mining

	Implementation of data mining	Quantity
Healthcare and Medical	Management of rabies case data in Palembang Management health and medical data Management of BPJS Employment Old Age Insurance claim data by region Management of patient data for healthcare safety Visualization of knowledge extraction in Patient Safety Care Management of athlete data, including profiles, training load, physiology, and environment	6
Education	Management data in the form of student online test scores Management of frequent itemsets from streaming transaction data in the presence of drift concept Determine the average length of school Management student academic performance data Prediction of student scholarship recipients at universities Evaluation of online users learning behavior Data management of development and implementation of STEM education Management student study period data Management data on prospective students of Harapan University Medan based on average report card, CAT test score, and interview score Management cosmetic sales data Management student graduation data Management of teacher performance assessment data Prediction of student academic performance Analysis and prediction of student academic performance Management data of education in higher education Management data on the percentage of illiterates aged 15+ by province Management sales data of Lakoe Dessert Pondok Kacang products Evaluation of the concept of sustainable entrepreneurship Management of agricultural product sales data Management PDAM Tirta Lihou customer satisfaction data Management WFH and WFO agent performance data Predictions customer loyalty of PT Pegadaian (Persero) Predicted delivery time Sales data management in S&R Babystore	15

Table 4. Implementation of data mining *(Continued)*

	Implementation of data mining	Quantity
Bussiness and Finance	Management of oil palm yield data	24
	Data management in the form of books	
	Bike sales data management	
	Management of economic data related to the COVID-19 pandemic in G20 countries	
	Data management in the form of stock and shipping	
	Management of late payment data of shopping center tenants	
	Secure cloud computing environment data management	
	Management of daily user consumption data in the community mall and damage repair data	
	Predicting the total box office of Chinese Kungfu films	
	Management data of energy consumption from smart meters	
	Management visitor satisfaction level data at Boombara Waterpark Rantauprapat	
	Optimize the company's strategic goals and the relationship between the company and customers	
	Management data IDS	
	Analysis of ICT implementation and application within the company	
Environmental and Meteorological	Management for performance evaluation of government administration personnel	2
	Management data in the form of rainfall predictions	
	Data management in the form of rainfall predictions in the city of Bandung	
Sports and Athletics	Manages NBA player performance data, injury records, and salary details	1
	Total	48

REFERENCES

[1] A. S. Bimantoro, R. Athallah, N. I. Al-azhar, and N. Legowo, "Implementation of data mining to determine payment delays for mall shopping center tenants using K-Means clustering method," *Budapest International Research and Critics Institute-Journal (BIRCI-Journal)*, vol. 5, no. 3, pp. 20614–20626, 2022.

[2] X. Shu and Y. Ye, "Knowledge discovery: methods from data mining and machine learning," *Social Science Research*, vol. 110, p. 102817, Feb. 2023, doi: 10.1016/j.ssresearch.2022.102817.

[3] D. Moher, A. Liberati, J. Tetzlaff, and D. G. Altman, "Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement," *PLoS Medicine*, vol. 6, no. 7, p. e1000097, Jul. 2009, doi: 10.1371/journal.pmed.1000097.

[4] P. Assiroj, H. L. H. S. Warnars, E. Abdurachman, A. I. Kistijantoro, and A. Doucet, "The implementation of memetic algorithm on image: a survey," *Journal of Mathematical and Computational Science*, 2021, doi: 10.28919/jmcs/5961.

[5] K. Rahayu, L. Novianti, and M. Kusnandar, "Implementation data mining with K-Means algorithm for clustering distribution rabies case area in Palembang City," *Journal of Physics: Conference Series*, vol. 1500, no. 1, 2020, doi: 10.1088/1742-6596/1500/1/012121.

[6] Y. Ariyanto, B. Harijanto, and A. N. Asri, "Analyzing student's learning interests in the implementation of blended learning using data mining," *International Journal of Online and Biomedical Engineering (iJOE)*, vol. 16, no. 11, pp. 153–160, Oct. 2020, doi: 10.3991/ijoe.v16i11.16453.

[7] S. Bagui and P. Stanley, "Mining frequent itemsets from streaming transaction data using genetic algorithms," *Journal of Big Data*, vol. 7, no. 1, p. 54, Dec. 2020, doi: 10.1186/s40537-020-00330-9.

[8] W. Liang and T. Li, "Research on human performance evaluation model based on neural network and data mining algorithm," *EURASIP Journal on Wireless Communications and Networking*, vol. 2020, no. 1, p. 174, 2020, doi: 10.1186/s13638-020-01776-4.

[9] Saifullah, N. Hidayati, and Solikhun, "Implementation of data mining in grouping percentage of blind letters sge 15+ by province using K-Means algorithm," *Journal of Physics: Conference Series*, vol. 1641, no. 1, 2020, doi: 10.1088/1742-6596/1641/1/012081.

[10] R. Alfianzah, R. I. Handayani, and M. Murniyati, "Implementation of apriori algorithm data mining for increase sales," *SinkrOn*, vol. 5, no. 1, p. 17, Oct. 2020, doi: 10.33395/sinkron.v5i1.10587.

[11] P. Bajdor and I. Pawełoszek, "Data mining approach in evaluation of sustainable entrepreneurship," *Procedia Computer Science*, vol. 176, pp. 2725–2735, 2020, doi: 10.1016/j.procs.2020.09.284.

[12] Supriyono, K. F. Ferine, D. Puspitasari, Rulinawaty, and E. Timotius, "Implementation of data mining with apriori techniques to determine the pattern of purchasing of agricultural equipment (case study: XYZ store)," *Journal of Physics: Conference Series*, vol. 1933, no. 1, p. 012029, Jun. 2021, doi: 10.1088/1742-6596/1933/1/012029.

[13] T. H. Sinaga, A. Wanto, I. Gunawan, S. Sumarno, and Z. M. Nasution, "Implementation of data mining using C4.5 algorithm on customer satisfaction in Tirta Lihou PDAM," *Journal of Computer Networks, Architecture, and High-Performance Computing*, vol. 3, no. 1, pp. 9–20, Jan. 2021, doi: 10.47709/cnahpc.v3i1.923.

[14] J. Jaja, N. Priatna, and T. S. Ardan, "Implementation of data mining technique for performance of WFH and WFO agents using the k-means method case study study of PT. Infomedia Telkom consumer profiling services," *Budapest International Research in Exact Sciences (BirEx) Journal*, vol. 3, no. 2, pp. 117–125, Apr. 2021, doi: 10.33258/birex.v3i2.1810.

[15] C. Kustandi, G. W. Bhawika, E. Susanto, V. N. Van Harling, and A. D. Ekawati, "Implementation of the k-medoids method in the old school expectations in indonesia by utilizing educational data mining," *ARPN Journal of Engineering and Applied Sciences*, vol. 16, no. 17, pp. 1811–1816, 2021.

[16] R. Muttaqien, M. G. Pradana, and A. Pramuntadi, "Implementation of data mining using C4.5 algorithm for predicting customer loyalty of PT. Pegadaian (Persero) Pati area office," *International Journal of Computer and Information System (IJCIS)*, vol. 2, no. 3, pp. 64–68, 2021, doi: 10.29040/ijcis.v2i3.36.

[17] D. Sharma and D. P. Sharma, "Design and Implementation of rainfall prediction model using supervised machine learning data mining techniques," *Indian Journal of Data Mining*, vol. 1, no. 2, pp. 20–26, Nov. 2021, doi: 10.54105/ijdm.B1615.111221.

[18] M. T. Sembiring and R. H. Tambunan, "Analysis of graduation prediction on time based on student academic performance using the Naïve Bayes algorithm with data mining implementation (Case study: Department of Industrial Engineering USU)," *IOP Conference Series: Materials Science and Engineering*, vol. 1122, no. 1, 2021, doi: 10.1088/1757-899X/1122/1/012069.

[19] P. Wang and J. Li, "Implementation of real-time medical and health data mining system based on machine learning," *Journal of Healthcare Engineering*, vol. 2021, pp. 1–5, Nov. 2021, doi: 10.1155/2021/7011205.

[20] T. Wahyudi and D. S. Arroufu, "Implementation of data mining prediction delivery time using linear regression algorithm," *Journal of Applied Engineering and Technological Science*, vol. 4, no. 1, pp. 84–92, 2022, doi: 10.37385/jaets.v4i1.918.

[21]   T. Wahyudi and T. Silfia, "Implementation of data mining using k-means clustering method to determine sales strategy in s&r baby store," *Journal of Applied Engineering and Technological Science*, vol. 4, no. 1, 2022, doi: 10.37385/jaets.v4i1.913.

[22]   W. J. Mawaddah, I. Gunawan, and I. P. Sari, "Implementation of data mining algorithm for clustering of palm oil harvested data," *JOMLAI: Journal of Machine Learning and Artificial Intelligence*, vol. 1, no. 1, 2022, doi: 10.55123/jomlai.v1i1.163.

[23]   M. Ulfah and A. S. Irtawaty, "Implementation of data mining clustering using the k-medoids method in grouping library books," in *Proceeding International Applied Business and Engineering Conference*, 2022, pp. 188–194.

[24]   I. Anggriawan and W. Gunawan, "Implementation of data mining using k-means algorithm for bicycle sales prediction," *ILKOM Jurnal Ilmiah*, vol. 14, no. 3, pp. 284–293, Dec. 2022, doi: 10.33096/ilkom.v14i3.1291.284-293.

[25]   O. Taylan, A. S. Alkabaa, and M. T. Yılmaz, "Impact of COVID-19 on G20 countries: analysis of economic recession using data mining approaches," *Financial Innovation*, vol. 8, no. 1, p. 81, Sep. 2022, doi: 10.1186/s40854-022-00385-y.

[26]   T. Tata, R. Saragih, and S. Syahputra, "Implementation of data mining grouping of old age guarantee (Jht) based on region in pandemic period," *Journal of Information Systems and Technology Research*, vol. 1, no. 2, 2022, doi: 10.55537/jistr.v1i2.136.

[27]   H. Hendri, H. Andrianof, and A. P. Gusman, "Implementation of C4.5 algorithm in data mining for predicting scholarship recipient students in University," *International Journal of Dynamics in Engineering and Sciences (IJDES) Kopertis Wilayah X*, vol. 7, no. 2, 2022, doi: 10.22216/jod.v17i2.

[28]   M. Kalas and N. D. Deshpande, "Implementation of production planning and scheduling: a data mining approach," *International Journal of Science and Research (IJSR)*, vol. 11, no. 6, pp. 1120–1126, Oct. 2022, doi: 10.21275/SR22603081404.

[29]   P. Monika, B. N. Ruchjana, and A. S. Abdullah, "The implementation of the ARIMA-ARCH model using data mining for forecasting rainfall in Bandung city," *International Journal of Data and Network Science*, vol. 6, no. 4, pp. 1309–1318, 2022, doi: 10.5267/j.ijdns.2022.6.004.

[30]   T. Ige and A. Sikiru, "Implementation of data mining on a secure cloud computing over a Web API using supervised machine learning algorithm," in *Lecture Notes in Networks and Systems, Springer Science and Business Media Deutschland GmbH*, 2022, pp. 203–210. doi: 10.1007/978-3-031-09076-9_20.

[31]   B. Zhang and J. Xie, "Design and implementation of an online learning behavior evaluation system based on data mining," *Advances in Multimedia*, vol. 2022, pp. 1–10, Jul. 2022, doi: 10.1155/2022/4259913.

[32]   L. Li, "Implementation of a community data processing system based on data mining," *Journal of Robotics*, vol. 2022, pp. 1–6, Feb. 2022, doi: 10.1155/2022/8286007.

[33]   L. Jiang, "Development and implementation path of kindergarten stem educational activities based on data mining," *Computational Intelligence and Neuroscience*, vol. 2022, pp. 1–10, Jan. 2022, doi: 10.1155/2022/2700674.

[34]   G. Belay Gebremeskel, B. Hailu, and B. Biazen, "Architecture and optimization of data mining modeling for visualization of knowledge extraction: patient safety care," *Journal of King Saud University - Computer and Information Sciences*, vol. 34, no. 2, pp. 468–479, Feb. 2022, doi: 10.1016/j.jksuci.2019.12.001.

[35]   R. Safdari, A. Deghatipour, M. Gholamzadeh, and K. Maghooli, "Applying data mining techniques to classify patients with suspected hepatitis C virus infection," *Intelligent Medicine*, vol. 2, no. 4, pp. 193–198, Nov. 2022, doi: 10.1016/j.imed.2021.12.003.

[36]   A. R. Putra, L. Septiani, and W. Angga, "The implementation of data mining techniques for predicting student study period using the C4.5 algorithm," *Indonesian Journal of Electrical Engineering and Renewable Energy (IJEERE)*, vol. 3, no. 2, pp. 96–100, 2023, doi: 10.57152/ijeere.v3i1.

[37]   G. Meng, F. De Costa, and D. Kasimon, "The application of data mining in predicting the total box office of chinese kungfu movies," *RISTI - Revista Iberica de Sistemas e Tecnologias de Informacao*, vol. 2023, no. Special Issue E62, pp. 450–456, 2023.

[38]   G. E. Okereke, M. C. Bali, C. N. Okwueze, E. C. Ukekwe, S. C. Echezona, and C. I. Ugwu, "K-Means clustering of electricity consumers using time-domain features from smart meter data," *Journal of Electrical Systems and Information Technology*, vol. 10, no. 1, p. 2, Jan. 2023, doi: 10.1186/s43067-023-00068-3.

[39]   H. A. Pratama, J. Yanris, M. Nirmala, and S. Hasibuan, "Implementation of data mining for data classification of visitor satisfaction levels," *Jurnal dan Penelitian Teknik Informatika*, vol. 8, no. 3, pp. 1832–1851, 2023, doi: 10.33395/sinkron.v8i3.12674

[40]   H. Mydyti, A. Kadriu, and M. Pejic Bach, "Using data mining to improve decision-making: case study of a recommendation system development," *Organizacija*, vol. 56, no. 2, pp. 138–154, 2023, doi: 10.2478/orga-2023-0010.

[41]   Mahesh T R, V Vivek, and Vinoth Kumar, "Implementation of machine learning-based data mining techniques for IDS," *International Journal of Information Technology, Research and Applications*, vol. 2, no. 1, 2023, doi: 10.59461/ijitra.v2i1.23.

[42]   A. S. R. Siregar, Y. S. Siregar, and M. Khairani, "Implementation of the data mining cart algorithm in the characteristic pattern of new student admissions," *Journal of Computer Networks, Architecture and High Performance Computing*, vol. 5, no. 1, pp. 263–275, Feb. 2023, doi: 10.47709/cnahpc.v5i1.1975.

[43]   N. Aurora Al Zahra, D. Saripurna, and M. S. Zulfansyuri Siambaton, "Mining data implementation analyses sale of cosmetical products using apriori algoritms on Balqis Toko," *Computer Science, Data Science And Arificial Intelligence*, vol. 1, no. 1, 2023.

[44]   P. Putriyana and O. Nurdiawan, "Implementation of data mining to predict graduation of SMK Al Huda Kedungwungu students using the Naïve Bayes classifier algorithm," *Experimental Student Experiences*, vol. 1, no. 2, 2023, doi: 10.58330/ese.v1i4.202.

[45]   M. Bakator, D. Cockalo, M. Kavalić, E. Terek Stojanović, and V. Gluvakov, "An application of statistical methods in data mining techniques to predict ICT implementation of enterprises," *Applied Sciences (Switzerland)*, vol. 13, no. 6, 2023, doi: 10.3390/app13064055.

[46]   N. Hasanah, Z. ' Aini, I. Lubis, T. R. Pasaribu, S. Informasi, and S. Kaputama Binjai, "Implementation of data mining teacher performance assessment using the K-means clustering method in student learning styles in the 4.0 Era 1," *Journal of Engineering, Technology and Computing (JETCom)*, vol. 3, no. 1, 2024.

[47]   X. Zheng and C. Li, "Predicting Students' Academic performance through machine learning classifiers: a study employing the naive bayes classifier (NBC)," *International Journal of Advanced Computer Science and Applications*, vol. 15, no. 1, pp. 994–1008, 2024, doi: 10.14569/IJACSA.2024.0150199.

[48]   Y. Xie, "Efficiency of hybrid decision tree algorithms in evaluating the academic performance of students," *International Journal of Advanced Computer Science and Applications*, vol. 15, no. 2, pp. 480–491, 2024, doi: 10.14569/IJACSA.2024.0150251.

[49]   J. Zhou and X. Li, "Application of big data in college student education management based on data warehouse technology and integrated learning," *International Journal of e-Collaboration*, vol. 20, no. 1, pp. 1–20, Jul. 2024, doi: 10.4018/IJeC.346368.

[50]   Q. Chen, "Identification of Potential injury risk factors and prediction model construction of athletes using data mining algorithm," *Journal of Electrical Systems*, vol. 20, no. 6s, pp. 2048–2058, 2024, doi: 10.52783/jes.3120.

[51]   V. Sarlis, G. Papageorgiou, and C. Tjortjis, "Leveraging sports analytics and association rule mining to uncover recovery and economic impacts in NBA Basketball," *Data*, vol. 9, no. 7, p. 83, Jun. 2024, doi: 10.3390/data9070083.

BIOGRAPHIES OF AUTHORS

Wisnu Widayat received his Bachelor of Laws degree from Mahendradatta University in 1995 and Master of Science degree from Muhammadiyah University Jakarta in 2010. He currently works at the Immigration Polytechnic, Ministry of Law and Human Rights, Republic of Indonesia. He can be contacted at email: wisnu@poltekim.ac.id

Priati Assiroj receives her Ph.D. in computer science from Binus Unviersity in May 2022. She is a lecturer in Politeknik Imigrasi, Ministry of Law and Human Rights, Republic of Indonesia. In 2019 She received an international research grant from Bina Nusantara International research endowment. Her research interests are data mining, high-performance computing, AI, machine learning and evolutionary algorithm. She can be contacted at email: priati.assiroj@poltekim.ac.id and tie.assiroj@gmail.com.

Sohirin received his Bachelor of Laws from Mahasaraswati University Matarampada in 1995 and Master of Humanities from IBLAM College of Law in 2001. He currently works at the Immigration Polytechnic, Ministry of Law and Human Rights, Republic of Indonesia. He can be contacted via email: sohirin67@yahoo.com.

Isidorus Anung Prabadhi currently, he is worked at the Directorate General of Immigration since 1999 and then in 2012 at the Immigration Academy which in 2017 became the Immigration Polytechnic, graduated from the Master of Human Resources at Mercu Buana University, Jakarta. He can be contacted at email: anung53@gmail.com.

Pasha Adelia Kautsar Currently, her is a Diploma IV student of the Immigration Technology Management Study Program at the Immigration Polytechnic. She can be contacted at email: adeliakautsarr@gmail.com.