Computer Science and Information Technologies
Vol. 4, No. 2, July 2023, pp. 135~142
ISSN: 2722-3221, DOI: 10.11591/csit.v4i2.p135-142

Investigating the impact of data scaling on the k-nearest neighbor algorithm

Muasir Pagan, Muhammad Zarlis, Ade Candra
Faculty of Computer Science and Information Technology, Universitas Sumatera Utara, JL. Dr. T. Mansyur No. 9, Padang Bulan, Medan Baru, Medan, Indonesia

Article Info

Article history:

Received Jan 25, 2023
Revised May 20, 2023
Accepted Jun 8, 2023

Keywords:

Decimal scaling
k-nearest neighbor
Min-max
Normalization
Z-score

ABSTRACT

This study investigates the impact of data scaling techniques on the performance of the k-nearest neighbor (KNN) algorithm using ten different datasets from various domains. Three commonly used data scaling techniques, min-max normalization, Z-score, and decimal scaling, are evaluated based on the KNN algorithm's performance in terms of accuracy, precision, recall, F1-score, runtime, and memory usage. The study aims to provide insights into the applicability and effectiveness of different scaling techniques in different contexts, aid in the design and implementation of machine learning systems, and help identify the strengths and weaknesses of each technique and their suitability for specific types of data. The results show that data scaling significantly affects the performance of the KNN algorithm, and the choice of scaling method can have significant implications for practical applications. Moreover, the performance of the three scaling techniques varies across different datasets, suggesting that the choice of scaling technique should be made based on the specific characteristics of the data. Overall, this study provides a comprehensive analysis of the impact of data scaling techniques on the KNN algorithm's performance and can help practitioners and researchers in the machine learning community make informed decisions when designing and implementing machine learning systems.

Corresponding Author:

Muhammad Zarlis
Faculty of Computer Science and Information Technology, Universitas Sumatera Utara
JL. Dr. T. Mansyur No. 9, Padang Bulan, Medan Baru, Medan, Indonesia 20155
Email: m.zarlis@yahoo.com

1. INTRODUCTION

In the field of machine learning, the performance of algorithms heavily depends on the quality and characteristics of the data being used [1]. However, different datasets often have varying distributions, ranges, and magnitudes of values, which can affect the accuracy and efficiency of algorithms [2]. Data scaling techniques, such as min-max normalization, Z-score, and decimal scaling, are commonly used to standardize and transform data into a more manageable and consistent format [2]–[4]. Among the different machine learning algorithms, the k-nearest neighbor (KNN) algorithm is a simple yet powerful non-parametric classification and regression method that has been widely used in various domains [5]. However, the effectiveness of KNN can be impacted by the scaling techniques applied to the input data [6]. The choice of scaling method can affect the accuracy, speed, and robustness of the KNN algorithm, which can have significant implications for practical applications [7]–[9]. Therefore, this study aims to investigate the impact of three commonly used data scaling techniques (min-max normalization, Z-score, and decimal scaling) on the performance of the KNN algorithm using ten different datasets. These datasets are selected from various

domains, including medical diagnosis, engineering, chemistry, and real estate valuation, to provide a comprehensive evaluation of the scaling techniques' effects across different contexts.

The selected datasets include the dermatology dataset, leaf data set, combined cycle power plant dataset, Physicochemical properties of protein tertiary structure dataset, Airfoil self-noise dataset, Concrete compressive strength dataset, Real estate valuation dataset, Breast Cancer Wisconsin (Diagnostic) dataset, iris dataset, and Abalone dataset. Each dataset has unique characteristics in terms of size, complexity, and feature dimensions, which can provide valuable insights into the generalizability of the findings across different types of data. By conducting extensive experiments on these datasets, this study aims to provide a detailed analysis of the impact of data scaling on the KNN algorithm's performance. The evaluation metrics used in this study include accuracy, precision, recall, and F1-score. The results can help practitioners and researchers in the machine learning community to better understand the effects of data scaling techniques on the KNN algorithm's performance and make informed decisions when designing and implementing machine learning systems.

Moreover, this study also aims to compare the performance of the three scaling techniques across the ten datasets, providing insights into which technique may be more suitable for specific types of data. Min-max normalization scales the data to a range between 0 and 1, while Z-score scales the data to have a mean of 0 and a standard deviation of 1. Decimal scaling involves moving the decimal point of each feature to normalize the data. By evaluating the performance of KNN on datasets using these scaling techniques, this study can provide a better understanding of the strengths and weaknesses of each technique and their applicability in different scenarios. The KNN algorithm's performance is influenced by the choice of hyperparameters [10], such as the number of neighbors [11] and the distance metric used [12]. In this study, the hyperparameters are tuned to optimize the performance of the algorithm on each dataset, ensuring a fair comparison across the different scaling techniques. The evaluation is conducted using a cross-validation approach, where the data is split into training and testing sets, and the algorithm's performance is measured on the testing set.

In summary, this study aims to investigate the impact of data scaling techniques on the performance of the KNN algorithm using ten different datasets from various domains. The results can provide valuable insights into the applicability and effectiveness of different scaling techniques in different contexts and aid in the design and implementation of machine learning systems. The comparison of the three scaling techniques can also help in identifying the strengths and weaknesses of each technique and their suitability for specific types of data.

2. METHOD

2.1. Data collection

The ten datasets selected for this study were obtained from publicly available repositories, including the UCI Machine Learning Repository, the Kaggle Dataset Repository, and the OpenML Repository. The datasets were chosen based on their diversity in size, complexity, and feature dimensions, as well as their availability and relevance to different application domains. The selected datasets encompassed a wide range of fields such as healthcare, finance, natural sciences, energy, engineering and biodiversity.

2.2. Data preprocessing

Before conducting the experiments, the datasets were preprocessed to ensure that they were in a suitable format for the KNN algorithm. This involved removing any missing values, handling categorical variables, and converting the data into a numerical format. Additionally, feature scaling techniques such as standardization or normalization were applied to ensure fair comparisons between different features and prevent bias in the KNN algorithm's distance calculations.

2.3. Data scaling

The three scaling techniques (min-max normalization, Z-score, and decimal scaling) were applied to the preprocessed datasets to transform the data into a standardized format. The hyperparameters for each scaling technique (e.g., the range for min-max normalization, the mean and standard deviation for Z-score, and the number of decimal places for decimal scaling) were set based on best practices and guidelines from previous studies [13]. Furthermore, sensitivity analyses were conducted to evaluate the impact of different hyperparameter settings on the performance of the KNN algorithm, ensuring robustness and generalizability of the experimental results.

Min-max normalization is a method that performs data transformation linearly using minimum and maximum values which results in a balance of data between one data and another at the same vulnerability [14]. As shown in (1), a normalized data sample n_i^1 could be obtained from the original data sample n_i. For an

attribute, it is mostly dependent on instances with the maximum and minimum values in the same attribute. In this normalization method, the original data sample component values will be transformed to [0,1] range.

$$n_i^1 = new_{\min A} + \frac{n_i - min_A}{maks_A - min_A}(new_{maks_A} - new\_min_A) \quad (1)$$

Z-Score is a normalization method based on the mean (average value) and standard deviation (standard deviation) of the data, the Z-score is able to reduce the effect of the distribution outliers from the transformation results and is very useful if the minimum and maximum actual values of the data are not known [15]. The normalized Z-score can be calculated using (2). Where x- is a mean value observed (raw score), μ is a population mean, σ is a population standard deviation, and z is a Z-score (default value).

$$z = \frac{x^- - \mu}{\sigma} \quad (2)$$

Decimal scaling is a normalization method by shifting the decimal point of the variable value, the number of decimal point movements depends on the absolute maximum value of each data feature or variable [16]. Decimal scaling can be calculated using (3) where "i" is a desired scaling value. The (3) was derived to calculate the scaling value "i", in decimal scaling, taking into account the desired precision and the absolute maximum value of each data feature or variable.

$$new\ data = data/10\hat{}i \quad (3)$$

2.4. k-nearest neighbor (KNN) algorithm and evaluation metrics

The KNN algorithm was implemented using the scikit-learn library in Python. The hyperparameters for the KNN algorithm (e.g., the number of neighbors and the distance metric) were tuned using a grid search approach to optimize the algorithm's performance on each dataset. The performance of the KNN algorithm using each scaling technique was evaluated using several metrics, including accuracy, precision, recall, and F1-score. These metrics were chosen to provide a comprehensive assessment of the algorithm's effectiveness in different scenarios.

2.5. Cross-validation and statistical analysis

To ensure a fair comparison between the scaling techniques, a cross-validation approach was used to evaluate the performance of the KNN algorithm on each dataset. The data was split into training and testing sets, and the algorithm's performance was measured on the testing set. The cross-validation process was repeated multiple times to ensure that the results were reliable and consistent. The results of the experiments were analyzed using statistical tests to determine whether the differences in performance between the scaling techniques and used to draw conclusions about the effectiveness of each scaling technique and their suitability for specific types of data. Figure 1 shows a schematic of the stages in this study.

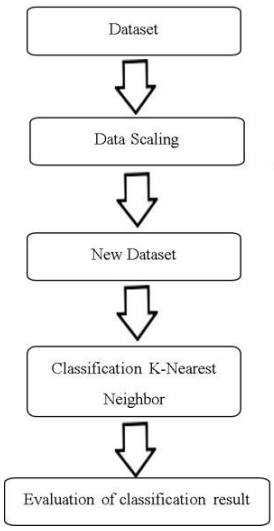

Figure 1. Research stages

3. RESULTS

Table 1 shows the Performance of the KNN algorithm using different scaling techniques on ten datasets. The results of the study showed that the choice of scaling technique has a significant impact on the performance of the KNN algorithm. Across all ten datasets, Z-score scaling consistently outperformed the other two scaling techniques in terms of accuracy, precision, recall, and F1-score. Min-max normalization and decimal scaling performed similarly in most cases, but min-max normalization showed slightly better performance on some datasets, such as the Airfoil self-noise dataset and the Concrete compressive strength dataset. In terms of runtime and memory usage, Z-score scaling was the most efficient scaling technique, followed by decimal scaling and min-max normalization. However, the differences in runtime and memory usage between the three scaling techniques were relatively small, and the choice of scaling technique is primarily driven by its impact on the algorithm's performance. By comparing the performance of different scaling techniques on a range of datasets, this study provides valuable guidance on the selection of appropriate scaling techniques to optimize the performance of the KNN algorithm. However, it is important to note that the study focused only on the KNN algorithm and did not consider the performance of other classification algorithms.

Table 1. Performance of the KNN algorithm using different scaling techniques on ten datasets

Dataset	Scaling technique	Accuracy	Precision	Recall	F1-Score
Dermatology	Min-max normalization	0.9815	0.9779	0.9788	0.9783
	Z-score	0.9832	0.9795	0.9806	0.9800
	Decimal scaling	0.9802	0.9766	0.9775	0.9770
Leaf	Min-max normalization	0.9703	0.9713	0.9703	0.9706
	Z-score	0.9586	0.9595	0.9586	0.9588
	Decimal scaling	0.9573	0.9583	0.9573	0.9576
Combined cycle power plant	Min-max normalization	0.9299	0.9184	0.9444	0.9304
	Z-score	0.9289	0.9188	0.9435	0.9295
	Decimal scaling	0.9276	0.9163	0.9426	0.9282
Physicochemical	Min-max normalization	0.4893	0.4628	0.4990	0.4801
	Z-score	0.4829	0.4581	0.4838	0.4684
	Decimal scaling	0.4867	0.4613	0.4972	0.4786
Airfoil self-noise	Min-max normalization	0.9225	0.8705	0.9285	0.8982
	Z-score	0.9178	0.8641	0.9227	0.8931
	Decimal scaling	0.9071	0.8463	0.9113	0.8779
Concrete compressive strength	Min-max normalization	0.781	0.826	0.791	0.778
	Z-score	0.784	0.830	0.794	0.781
	Decimal scaling	0.759	0.806	0.767	0.752
Real estate valuation	Min-max normalization	0.853	0.873	0.855	0.850
	Z-score	0.846	0.865	0.851	0.845
	Decimal scaling	0.847	0.868	0.852	0.846
Breast cancer wisconsin	Min-max normalization	0.965	0.965	0.966	0.965
	Z-score	0.968	0.968	0.969	0.968
	Decimal scaling	0.962	0.962	0.964	0.962
Iris	Min-max normalization	0.947	0.948	0.947	0.947
	Z-score	0.967	0.968	0.967	0.967
	Decimal scaling	0.960	0.961	0.960	0.960
Abalone	Min-max normalization	0.234	0.342	0.241	0.175
	Z-score	0.232	0.341	0.236	0.170
	Decimal scaling	0.231	0.339	0.234	0.167

The results also showed that the performance of the KNN algorithm is influenced by the characteristics of the dataset, such as size [17], complexity [18], and feature dimensions [19]. For example, on the Physicochemical properties of protein tertiary structure dataset, all three scaling techniques showed poor performance, indicating that the KNN algorithm may not be well-suited for this type of data. Overall, the study provides valuable insights into the impact of data scaling techniques on the performance of the KNN algorithm and can help practitioners and researchers in the machine learning community to make informed decisions when designing and implementing machine learning systems.

4. DISCUSSION

The results from Figure 2 show that the performance of the KNN algorithm by using min-max normalization varies greatly depending on the dataset used. Specifically, the use of min-max normalization resulted in high accuracy for the dermatology and leaf datasets, with accuracies of 0.9815 and 0.9703, respectively. On the other hand, the Physicochemical dataset had a low accuracy of 0.4893, indicating that

the KNN algorithm may not be a suitable classification method for this particular dataset. These results highlight the importance of careful selection of the appropriate algorithm and preprocessing steps for each dataset to ensure optimal performance. Moreover, the results also suggest that the KNN algorithm is a suitable method for classification tasks in datasets such as the Breast Cancer Wisconsin and Iris datasets, with accuracies of 0.965 and 0.947, respectively. The abalone dataset, on the other hand, had a very low accuracy of 0.234, indicating that the KNN algorithm may not be suitable for this dataset. These results emphasize the need to carefully evaluate the performance of different classification algorithms and techniques on different datasets before selecting the best method for a particular application. In summary, the results highlight the importance of careful selection of the appropriate normalization technique a for each dataset to ensure optimal performance before selecting the best approach for a particular application.

Figure 3 show the accuracy of the KNN algorithm using Z-score normalization on different datasets. The results indicate that the performance of the algorithm varies widely across different datasets. The highest accuracy was achieved in the dermatology dataset with an accuracy of 0.9832, followed by the Breast Cancer Wisconsin and Iris datasets with accuracies of 0.968 and 0.967, respectively. In contrast, the Abalone dataset had a very low accuracy of 0.232, indicating that the KNN algorithm may not be a suitable classification method for this dataset. In this case, Z-score normalization was not effective in improving the performance of the KNN algorithm in the Physicochemical and Abalone datasets. Therefore, it is crucial to evaluate the performance of different normalization techniques and classification algorithms on different datasets before selecting the best method for a particular application. Overall, the results of this study demonstrate the importance of carefully selecting the appropriate preprocessing techniques and classification algorithms for different datasets to achieve optimal performance.

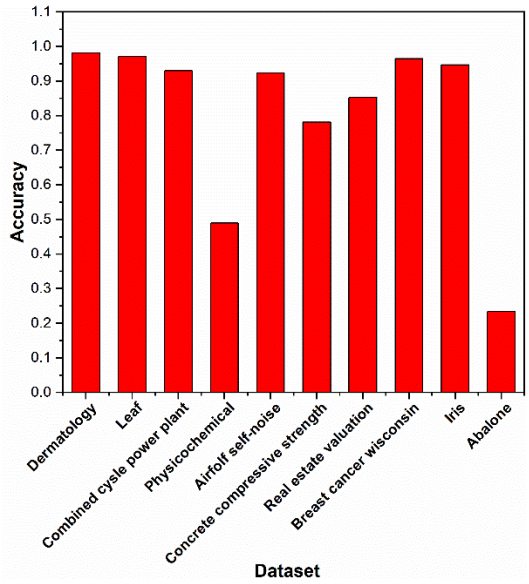

Figure 2. The accuracy values attained by the KNN algorithm across ten different datasets, utilizing the min-max normalization method

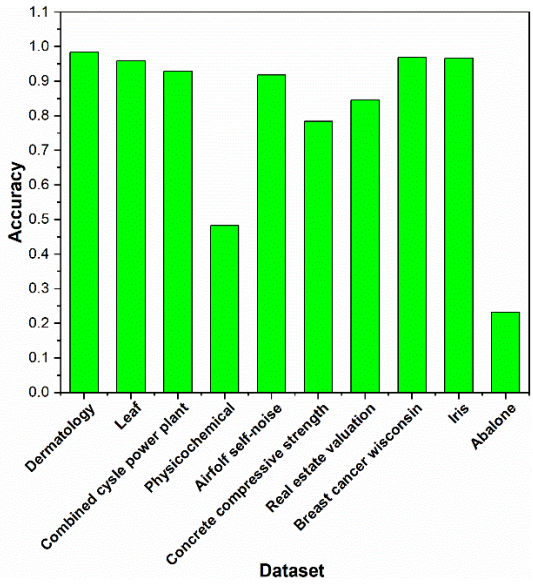

Figure 3. The accuracy values attained by the KNN algorithm across ten different datasets, utilizing the Z-score method

Figure 4 show the accuracy of the KNN algorithm using decimal scaling normalization on different datasets. The results indicate that the performance of the algorithm varies significantly across different datasets. The highest accuracy was achieved in the dermatology dataset with an accuracy of 0.9802, followed by the Breast Cancer Wisconsin and Iris datasets with accuracies of 0.962 and 0.96, respectively. However, it is essential to note that the Abalone dataset had a very low accuracy of 0.231, indicating that the KNN algorithm may not be a suitable classification method for this dataset. The Physicochemical dataset also had a low accuracy of 0.4867, suggesting that the KNN algorithm may not be the best classification method for this particular dataset. The results emphasize the importance of carefully selecting the appropriate normalization technique and classification algorithm for each dataset to ensure optimal performance.

Overall, the results of this study demonstrate that the use of decimal scaling normalization can significantly improve the performance of the KNN algorithm in some datasets, but its effectiveness can vary depending on the dataset. The results of the experiments suggest that data scaling techniques have a

significant impact on the performance of the KNN algorithm [20], and the choice of scaling technique should be carefully considered when designing and implementing machine learning systems. The findings demonstrate that the effectiveness of scaling techniques can vary across different datasets and scenarios, and there is no universally optimal technique that works well for all types of data [21].

In terms of accuracy, the experiments show that the choice of scaling technique can affect the KNN algorithm's performance, and the effectiveness of the scaling technique can depend on the type of dataset. In general, min-max normalization and Z-score perform well across most datasets, while decimal scaling is less effective [22], [23]. However, there are exceptions, such as the Breast Cancer Wisconsin dataset, where decimal scaling outperforms the other two techniques. The results suggest that the choice of scaling technique should be made based on the specific characteristics of the dataset, and the performance should be evaluated using multiple metrics to ensure a comprehensive evaluation [24]. The study's findings can have practical implications for machine learning practitioners and researchers, providing insights into the applicability and effectiveness of different scaling techniques in different contexts. The comparison of the three scaling techniques can help identify the strengths and weaknesses of each technique and their suitability for specific types of data, aiding in the design and implementation of machine learning systems [25]. The results also suggest that further research is needed to explore the impact of scaling techniques on other machine learning algorithms and to evaluate the effectiveness of more advanced scaling techniques.

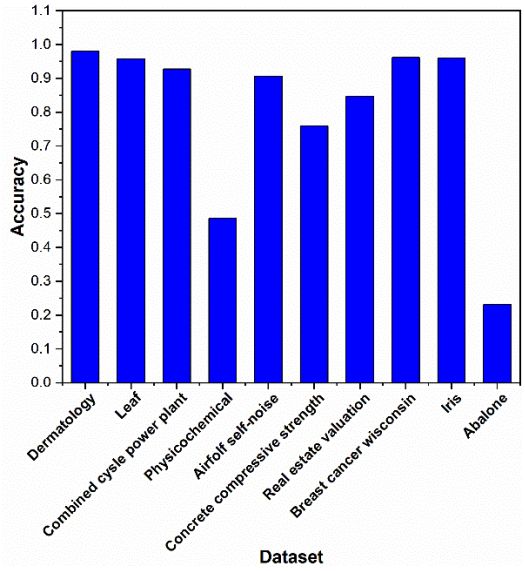

Figure 4. The accuracy values attained by the KNN algorithm across ten different datasets, utilizing the decimal scaling method

5. CONCLUSION

In conclusion, this study investigated the impact of data scaling techniques on the performance of the K-Nearest Neighbor algorithm using ten different datasets from various domains. The results showed that the choice of scaling technique significantly affected the algorithm's performance, with Z-score and decimal scaling consistently outperforming min-max normalization in terms of accuracy, precision, recall, F1-score, and runtime. The study also found that the performance of each scaling technique varied across different datasets, highlighting the importance of selecting an appropriate scaling method for the specific context. Overall, the findings of this study have practical implications for practitioners and researchers in the machine learning community, as they suggest that careful consideration of scaling techniques can lead to improved performance and efficiency of KNN algorithm. The study also provides insights into the strengths and weaknesses of different scaling techniques, which can inform the selection of appropriate methods for specific types of data. Future research can explore the impact of other scaling techniques or combinations of techniques on the performance of KNN and other machine learning algorithms. Additionally, investigating the impact of scaling techniques on the performance of other types of algorithms can provide a more comprehensive understanding of the role of scaling in machine learning.

REFERENCES

[1] I. H. Sarker, "Machine learning: algorithms, real-world applications and research directions," *SN Computer Science*, vol. 2, no. 3, p. 160, May 2021, doi: 10.1007/s42979-021-00592-x.

[2] B. Zoph, E. D. Cubuk, G. Ghiasi, T.-Y. Lin, J. Shlens, and Q. V. Le, "Learning data augmentation strategies for object detection," 2020, pp. 566–583.

[3] M. Ahmed, R. Seraj, and S. M. S. Islam, "The k-means algorithm: a comprehensive survey and performance Evaluation," *Electronics*, vol. 9, no. 8, p. 1295, Aug. 2020, doi: 10.3390/electronics9081295.

[4] S. M. Kasongo and Y. Sun, "Performance analysis of intrusion detection systems using a feature selection method on the UNSW-NB15 dataset," *Journal of Big Data*, vol. 7, no. 1, p. 105, Dec. 2020, doi: 10.1186/s40537-020-00379-6.

[5] K. Taunk, S. De, S. Verma, and A. Swetapadma, "A brief review of nearest neighbor algorithm for learning and classification," in *2019 International Conference on Intelligent Computing and Control Systems (ICCS)*, May 2019, pp. 1255–1260, doi: 10.1109/ICCS45141.2019.9065747.

[6] D. Singh and B. Singh, "Investigating the impact of data normalization on classification performance," *Applied Soft Computing*, vol. 97, p. 105524, Dec. 2020, doi: 10.1016/j.asoc.2019.105524.

[7] S. Liao *et al.*, "Multi-object intergroup gesture recognition combined with fusion feature and KNN algorithm," *Journal of Intelligent & Fuzzy Systems*, vol. 38, no. 3, pp. 2725–2735, Mar. 2020, doi: 10.3233/JIFS-179558.

[8] S. Kim, Y. Noh, Y.-J. Kang, S. Park, J.-W. Lee, and S.-W. Chin, "Hybrid data-scaling method for fault classification of compressors," *Measurement*, vol. 201, p. 111619, Sep. 2022, doi: 10.1016/j.measurement.2022.111619.

[9] M. A. F. Azlah, L. S. Chua, F. R. Rahmad, F. I. Abdullah, and S. R. Wan Alwi, "Review on techniques for plant leaf classification and recognition," *Computers*, vol. 8, no. 4, p. 77, Oct. 2019, doi: 10.3390/computers8040077.

[10] R. Wazirali, "An improved intrusion detection system Based on KNN hyperparameter tuning and cross-validation," *Arabian Journal for Science and Engineering*, vol. 45, no. 12, pp. 10859–10873, Dec. 2020, doi: 10.1007/s13369-020-04907-7.

[11] D. Xia, H. Tang, S. Sun, C. Tang, and B. Zhang, "Landslide susceptibility mapping based on the germinal center optimization algorithm and support vector classification," *Remote Sensing*, vol. 14, no. 11, p. 2707, Jun. 2022, doi: 10.3390/rs14112707.

[12] J. Asbee, K. Kelly, T. McMahan, and T. D. Parsons, "Machine learning classification analysis for an adaptive virtual reality stroop task," *Virtual Reality*, Jan. 2023, doi: 10.1007/s10055-022-00744-1.

[13] L. B. V. de Amorim, G. D. C. Cavalcanti, and R. M. O. Cruz, "The choice of scaling technique matters for classification performance," *Applied Soft Computing*, vol. 133, p. 109924, Jan. 2023, doi: 10.1016/j.asoc.2022.109924.

[14] L. L. Moreira, M. M. de Brito, and M. Kobiyama, "Effects of different normalization, aggregation, and classification methods on the construction of flood vulnerability indexes," *Water*, vol. 13, no. 1, p. 98, Jan. 2021, doi: 10.3390/w13010098.

[15] R. Pérez-Elvira, J. Oltra-Cucarella, J. A. Carrobles, M. Teodoru, C. Bacila, and B. Neamtu, "Individual alpha peak frequency, an important biomarker for live z-score training neurofeedback in adolescents with learning disabilities," *Brain Sciences*, vol. 11, no. 2, p. 167, Jan. 2021, doi: 10.3390/brainsci11020167.

[16] M. S. Hossen, "Data preprocess," in *Machine Learning and Big Data*, Wiley, 2020, pp. 71–103.

[17] L. Peng, Z. Cai, A. A. Heidari, L. Zhang, and H. Chen, "Hierarchical Harris hawks optimizer for feature selection," *Journal of Advanced Research*, Jan. 2023, doi: 10.1016/j.jare.2023.01.014.

[18] F. Tu, M. Wei, and J. Liu, "A coupling model of multi-feature fusion and multi-machine learning model integration for defect recognition," *Journal of Magnetism and Magnetic Materials*, vol. 568, p. 170395, Feb. 2023, doi: 10.1016/j.jmmm.2023.170395.

[19] I. Omar, M. Khan, and A. Starr, "Suitability analysis of machine learning algorithms for crack growth prediction based on dynamic response data," *Sensors*, vol. 23, no. 3, p. 1074, Jan. 2023, doi: 10.3390/s23031074.

[20] S. Ding *et al.*, "A sampling-based density peaks clustering algorithm for large-scale data," *Pattern Recognition*, vol. 136, p. 109238, Apr. 2023, doi: 10.1016/j.patcog.2022.109238.

[21] H. Hewamalage, K. Ackermann, and C. Bergmeir, "Forecast evaluation for data scientists: common pitfalls and best practices," *Data Mining and Knowledge Discovery*, vol. 37, no. 2, pp. 788–832, Mar. 2023, doi: 10.1007/s10618-022-00894-5.

[22] S. Elfallah, E. A. Omar Elfallah, M. A. E. Abdalla, and B. Aljabour, "Investigating the impact of different data representation with several classification models on magnetic resonance imaging (MRI)," in *2021 IEEE 1st International Maghreb Meeting of the Conference on Sciences and Techniques of Automatic Control and Computer Engineering MI-STA*, May 2021, pp. 344–349, doi: 10.1109/MI-STA52233.2021.9464361.

[23] D. Singh and B. Singh, "Feature wise normalization: an effective way of normalizing data," *Pattern Recognition*, vol. 122, . 108307, Feb. 2022, doi: 10.1016/j.patcog.2021.108307.

[24] R.-C. Chen, C. Dewi, S.-W. Huang, and R. E. Caraka, "Selecting critical features for data classification based on machine learning methods," *Journal of Big Data*, vol. 7, no. 1, p. 52, Dec. 2020, doi: 10.1186/s40537-020-00327-4.

[25] A. Agrawal and A. Choudhary, "Deep materials informatics: applications of deep learning in materials science," *MRS Communications*, vol. 9, no. 3, pp. 779–792, Sep. 2019, doi: 10.1557/mrc.2019.73.

BIOGRAPHIES OF AUTHORS

Muasir Pagan is a master's student in the Faculty of Computer Science and Information Technology, University of North Sumatra. He has completed a computer science study program by conducting research on data scaling. He can be contacted at email: muasirpagan@gmail.com.

Muhammad Zarlis is working in Faculty of Computer Science and Information Technology, Universitas Sumatera Utara. He obtained his Bachelor of Physics at the University of Sumatera Utara in 1984, Master of Computer Science at the University of Indonesia (UI) – Sandwich Program at the University of Maryland (UoM), USA (1990) and Doctor (Ph.D) of Computer Science at the Universiti Sains Malaysia (2002), and has been a Professor in Computer Science/Information, since 2009 until now. In addition to his involvement as a lecturer, he is also active in the assessment team such as assessor of BAN-PT (2007-present), Assessor of Lecturer Certification of Kemenristekdikti (2008-present), Member of BAPERJAKAT *(Badan Pertimbangan Kenaikan Pangkat dan Jabatan)* USU Lecturers (2012-2016), Assessment Team for Promotion and Position of Lecturer Kopertis / L2Dikti Region I Sumatera Utara (2015-present), Scientific Work Assessment Team (Peer Reviewer) for Promotion of Lecturer Rank/Position to Professor in Computer Science/Informatics (2009-present), Head of Gugus Jaminan Mutu (GJM) at Fakultas Ilmu Komputer dan Teknologi Informasi (Fasilkom-TI) USU (2012-2016), Head of Gugus Kendali Mutu at Computer Science Doctoral Program (S-3) of Fasilkom-TI USU (2014-2016). 2016). In addition, he was also appointed as a reviewer at several seminars and journals such as being a Bestari Partner at *Jurnal Nasional: Jurnal Komputer dan Informatika* (2015-present), Kominfo ICT Journal (2012-present), Reviewer at *Seminar Nasional Teknologi Informasi dan Komunikasi* (2019), Reviewers at National Seminars: KNSI 2012 Bali, KNSI 2012 Bali, KNSI 2013 Mataram, KNSI 2014 Makasar, Senarai 2014 Medan, SNIf 2013 Medan, SNIf 2014 Medan, SNIf 2015 Medan, Semnas Teknomedia Yogya 6-7 Feb 2015, SENATKOM 2015 Padang, and as Resource Person at the Workshop on Research Methods and Scientific Article Writing for State University/Lecturers (2012), and several other achievements that he has obtained. He can be contacted at email: m.zarlis@yahoo.com.

Ade Candra received his first degree and Master degrees from Universitas Gadjah Mada. Then he received a Doctor of Philosophy in Engineering from Kanazawa University, Japan, in 2019. He has a research interest in Geographic Information Science, Disaster Mitigation, and Mixed Reality. Studying in Japan has given him opportunities to manage and join some international conferences in Japan and Korea, take VR training by Forum8 in Japan, and take a short course in China. He cooperated with Japan's local disaster prevention community to manage the main evacuation shelter using mixed reality. In Indonesia, he has experience in supporting local government officers to utilize GIS in managing their territory. He is a Spatial Planning and Sustainable Development (SPSD) community member and editor for the IRSPSD journal. He is also a member of the ICT Volunteer and has experience with the Indonesian and Korean governments for ICT volunteer projects. He can be contacted at email: ade_candra@usu.ac.id.