Jurnal Mandiri IT
Vol. 14 No. 2, October (2025), pp. 260-268

ISSN 2301-8984 (Print), 2809-1884 (Online)
Published by Institute of Computer Science

Optimizing the performance of the K-Nearest Neighbors algorithm using grid search and feature scaling to improve data classification accuracy

Jonson Manurung[1], Hondor Saragih[2], Muhammad Azhar Prabukusumo[3], Eryan Ahmad Firdaus[4]

[1,2,3,4] Informatika, Universitas Pertahanan Republik Indonesia, Bogor, Indonesia

ARTICLE INFO

Article history:

Received Sep 26, 2025
Revised Oct 20, 2025
Accepted Oct 23, 2025

Keywords:

Feature Scaling;
Grid Search;
K-Nearest Neighbors;
Medical Classification;
Parameter Optimization.

ABSTRACT

The performance of distance-based classification algorithms such as K-Nearest Neighbors (KNN) is highly dependent on proper feature scaling and optimal parameter selection. Without systematic optimization, KNN may experience decreased accuracy due to feature scale disparities and suboptimal k-values. This study aims to enhance the performance of the KNN algorithm through the integration of Feature Scaling and Grid Search Cross-Validation as a parameter optimization strategy. The research employs the Breast Cancer Wisconsin Dataset, divided into 80% training and 20% testing data. Feature normalization was performed using StandardScaler, while Grid Search was applied to determine the optimal combination of parameters, including the number of neighbors (k), weighting function (weights), and distance metric (metric). The optimized KNN configuration with k = 9, weights = distance, and metric = manhattan achieved an average accuracy of 97.19%, outperforming the baseline accuracy of 93.86%. A paired t-test confirmed that the improvement was statistically significant ($p < 0.05$). These findings demonstrate that the synergy between feature scaling and parameter tuning can substantially improve both the accuracy and stability of KNN models. The scientific novelty of this study lies in the systematic integration of normalization and parameter optimization through Grid Search, providing an empirical framework that enhances KNN's robustness across datasets with heterogeneous feature distributions. The proposed approach is recommended for medical data classification and can be adapted to other domains with heterogeneous numerical feature distributions.

Corresponding Author:

Jonson Manurung,
Informatika,
Universitas Pertahanan Republik Indonesia,
Kawasan IPSC Sentul, Sukahati, Kec. Citeureup, Kabupaten Bogor, Jawa Barat 16810, Indonesia.
Email: Jonson.manurung@idu.ac.id

1. INTRODUCTION

Advances in information technology and the availability of large amounts of data have encouraged various fields to rely on machine learning-based data analysis to produce more accurate and efficient decisions (Ang et al., 2022; Li et al., 2021). One of the important tasks in machine learning is classification, which is the process of grouping data based on patterns learned from previous data (Ahuja et al., 2020; Jijo & Abdulazeez, 2021). Among the various algorithms used for classification, K-Nearest Neighbors (KNN) is one of the most popular methods due to its simplicity, flexibility, and effectiveness in various real-world cases, such as medical diagnosis, market prediction, and transaction fraud detection (Afzal et al., 2024; Halder et al., 2024; Mageed et al., 2024). However, KNN performance is highly dependent on parameter selection and data pre-processing quality. This

algorithm determines the class of new data based on the distance to a number of k nearest neighbors, so differences in feature scales can significantly affect distance calculations. Without proper feature scaling, features with large numerical values will dominate distance calculations and reduce model accuracy. In addition, suboptimal k parameter values can cause overfitting or underfitting, which ultimately reduces the model's generalization ability. Thus, parameter optimization and data pre-processing are key factors in improving KNN performance.

Several previous studies have attempted to improve KNN performance through various approaches, such as selecting different distance metrics, applying dimensionality reduction, or using weighted voting methods. However, most of these studies still apply manual or trial-and-error-based parameter selection, which is inefficient and does not guarantee optimal results. In addition, there are still few studies that systematically integrate two important components, namely Feature Scaling and Grid Search, to optimize KNN performance simultaneously. In fact, the combination of these two approaches has the potential to produce significant accuracy improvements and better model stability across various types of datasets. Grid Search is a method of systematically exploring combinations of specific parameters to find the best configuration based on model evaluation results. By combining Grid Search and Feature Scaling, the KNN optimization process can be carried out in a more measurable, objective, and replicable manner. This approach also allows for comparative analysis between the baseline model and the optimized model, thereby identifying the extent to which preprocessing and parameter tuning influence the improvement in algorithm performance (Abu Alfeilat et al., 2019; Halder et al., 2024).

Based on this description, this study aims to optimize the performance of the KNN algorithm by applying a combination of Grid Search and Feature Scaling to improve data classification accuracy. Specifically, this study will: (1) analyze the effect of Feature Scaling on KNN performance; (2) determine the optimal KNN parameters through Grid Search Cross-Validation; and (3) compare the performance of the model before and after the optimization process using the evaluation metrics of accuracy, precision, recall, and F1-score.The results of this study are expected to contribute technical recommendations related to effective hyperparameter tuning and data preprocessing strategies for the KNN algorithm, which can be used as a reference in the application of classification models in various fields, including health, economics, and intelligent system engineering. Thus, this research not only has theoretical value in the development of machine learning model optimization methods, but also practical value in improving the efficiency and accuracy of data-based decision-making systems.

The K-Nearest Neighbors (KNN) algorithm is a simple yet effective non-parametric classification method (Abedi & Professor, 2021; Tran et al., 2019). The basic principle of KNN is to determine the class of a test data based on the majority class of its k nearest training data based on a certain distance measure, such as Euclidean Distance, Manhattan Distance, or Minkowski Distance (Ehsani & Drabløs, 2020). The advantages of KNN lie in its simplicity of implementation and its adaptability to various data patterns. However, this algorithm also has several weaknesses, particularly its sensitivity to feature scaling and suboptimal selection of the k parameter, which can affect accuracy and computation time (Açıkkar & Tokgöz, 2024)vv. In practice, a k value that is too small can cause the model to be too sensitive to noise (overfitting), while a k value that is too large can blur the boundaries between classes (underfitting). Therefore, selecting the right k is an important factor in obtaining optimal performance (Uddin et al., 2022). In this study, the optimization of the k value was performed using the Grid Search Cross-Validation method to find the best parameters based on the evaluation results of the model on the validation data.

In distance-based algorithms such as KNN, differences in scale between features have a significant effect on classification results. For example, features with a large range of values will dominate distance calculations compared to features with a small scale. To overcome this, feature scaling is performed, which is the process of transforming feature values to be on a comparable scale. The two most commonly used methods are Min-Max Scaling and Standardization (Z-Score Normalization) (Herwanto et al., 2021). Research by Pagan, Muasir et al. (2023) shows that the application of feature scaling can increase KNN accuracy by more than 10% in datasets with high variance between features. In addition, combining scaling with the cross-validation method can reduce the risk of bias in classification results. Thus, feature scaling not only improves model performance but also increases the stability of evaluation results. In the context of this study, feature scaling is an important first step before parameter optimization using Grid Search.

Grid Search is a systematic parameter exploration method that tests every combination of predetermined parameter values, then selects the combination that provides the best results based on specific evaluation metrics (Belete & Huchaiah, 2022; Sharma et al., 2024). In KNN, the parameters that are usually optimized include the number of neighbors (k), the type of distance metric, and the weights. Although this method is quite time-consuming, Grid Search guarantees that the optimal solution is found within the defined parameter space. Several previous studies have proven the effectiveness of Grid Search in improving the performance of machine learning models. For example, research by Xie, Guoqiang, et al. (2019)shows that the application of CV Grid Search in KNN can increase accuracy by up to 8% compared to manual parameter selection. This reinforces the argument that systematic approaches such as Grid Search are more efficient and accurate than trial and error methods in determining the optimal model configuration.

The combination of feature scaling and Grid Search offers significant potential for optimizing KNN performance. Scaling ensures that all features contribute equally to the distance calculation, while Grid Search ensures that KNN parameters are in the best configuration. Research by Xie, Guoqiang, et al. (Xie et al., 2019) shows that the integration of these two techniques can consistently improve KNN accuracy in medical and financial classification datasets. However, there are still limitations in most previous studies, namely the lack of an approach that comprehensively examines the synergy between preprocessing and parameter tuning on KNN performance. Therefore, this study focuses on the integration of Feature Scaling and Grid Search as the main focus to produce a KNN model with better accuracy and prediction stability.

Several previous studies have examined various aspects of KNN algorithm optimization. For example, Halder, Rajib Kumar, et al. (2024) compared StandardScaler and MinMaxScaler in terms of KNN performance and found that StandardScaler produced better model stability on normally distributed datasets. Meanwhile, Abbas, Farkhanda, et al (Abbas et al., 2023) combined Principal Component Analysis (PCA) with Grid Search to reduce data dimensions and improve KNN computational efficiency. Another study by Vandika, Arnes Yuli, and Rahmat Pannyiwi (2024) in the context of Indonesian financial data shows that applying Grid Search Cross-Validation to KNN can increase the accuracy of creditworthiness predictions from 82% to 91%. However, the study did not explicitly involve feature scaling, so the optimization results are still potentially affected by feature scale imbalance. Referring to these results, this study aims to fill the research gap by combining the two optimization approaches-Feature Scaling and Grid Search-in an integrated manner. This integration is expected to produce more significant performance improvements and stronger model generalization across various types of datasets.

2. RESEARCH METHOD

2.1 Research Design

This study uses a computational quantitative experimental approach, with the aim of comparing the performance of the K-Nearest Neighbors (KNN) algorithm before and after the optimization process using Feature Scaling and Grid Search Cross-Validation. The experiment was conducted using the Python programming language with the scikit-learn library.

The main steps of the research include:

1. Data collection and exploration (data understanding).
2. Data preprocessing with feature scaling.
3. Application of the baseline KNN model.
4. Parameter optimization using Grid Search Cross-Validation (GridSearchCV).
5. Model performance evaluation based on accuracy, precision, recall, and F1-score metrics.

2.2 Data Sources and Dataset Description

The dataset used is the Breast Cancer Wisconsin (Diagnostic) Dataset, obtained from the UCI Machine Learning Repository. This dataset contains 569 sample data with 30 numerical features representing breast cancer cell characteristics based on microscopic measurements. The target labels consist of two classes:M (Malignant), malignant cancer. B (Benign), benign cancer.

2.3 Data Preprocessing

Feature Normalization and Standardization

Before being applied to the KNN algorithm, all features are normalized to avoid bias due to differences in scale between variables. Two techniques are used:

a. Min-Max Normalization

Used to convert the range of feature values to the interval [0, 1]. The mathematical formula is:

$$X' = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (1)$$

X is the original value of a feature, X_{min} and X_{max} are the minimum and maximum values of the feature, respectively. This transformation ensures that all features have the same range, so that each variable contributes equally to the distance calculation.

b. Standardization (Z-Score Normalization)

Used to transform the distribution of features so that they have a mean of 0 and a standard deviation of 1. The mathematical formula is:

$$Z = \frac{X - \mu}{\sigma} \quad (2)$$

μ is the mean value of the features and σ is the standard deviation. This approach is effective when the data is normally distributed, because it stabilizes the variance between features. This experiment uses StandardScaler from the scikit-learn library as the main scaling method.

2.4 K-Nearest Neighbors (KNN) Algorithm

Basic Principles

KNN is a non-parametric algorithm that classifies new samples based on their proximity to the k nearest neighbors from the training data. The classification process is carried out by calculating the distance between the test data x_i and each training data point x_j using a specific distance metric. The data is then classified into the most frequently occurring class (majority voting) of the k nearest neighbors. The general formula for calculating Euclidean distance is:

$$d(x_i, x_j) = \sqrt{\sum_{n=1}^{m} (x_{i,n} - x_{j,n})^2} \quad (3)$$

$d(x_i, x_j)$ is the Euclidean distance between two data vectors. m is the number of features. $x_{i,n}$ and $x_{j,n}$ are the nth feature values of the test data and training data, respectively.

After the distance is calculated, the algorithm determines the class using the majority principle:

$$C(x_i) = \underset{c \in C}{arg\,max} \sum_{j=1}^{k} I(y_j = c) \quad (4)$$

Function $I(y_j = c)$ has a value of 1 if neighbor j has class label c, and 0 if not. Thus, the class with the highest number of occurrences among the k neighbors becomes the prediction result.

2.5 Parameter Optimization Using Grid Search Cross-Validation

To find the best KNN parameter configuration, Grid Search Cross-Validation (GridSearchCV) from scikit-learn is used. The parameters tested include: Number of neighbors (n_neighbors): 1–25, Weight type (weights): 'uniform', 'distance', Distance metric (metric): 'euclidean', 'manhattan', 'minkowski'. Mathematically, Grid Search finds the optimal parameters by minimizing the error function:

$$\theta^* = \underset{\theta \in \Theta}{arg\,min} \frac{1}{K} \sum_{i=1}^{k} L\left(y_j, f_\theta(x_i)\right) \quad (5)$$

Θ is the set of all tested parameter combinations. L is the loss function, usually classification error. $f_\theta(x_i)$ is the model prediction result with parameters θ. This process is repeated for each parameter combination using k-fold cross-validation to obtain the highest average accuracy.

2.6 Model Performance Evaluation

The evaluation was conducted by comparing the performance of the baseline model and the optimized model using four main metrics:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$
$$Precision = \frac{TP}{TP + FP}$$
$$Recall = \frac{TP}{TP + FN} \quad (6)$$
$$F1 = 2x \frac{Precision\ x\ Recall}{Precision + Recall}$$

2. 7 Result Validation

To ensure the reliability of the results, 5-Fold Cross-Validation was used, which divides the data into 5 subsets with the training and testing processes carried out alternately. The average accuracy value from the five trials became the main indicator of model performance. Mathematically, the average accuracy is calculated as:

$$\overline{Acc} = \frac{1}{k}\sum_{i=1}^{k} Acc_i \quad (7)$$

Acc_i is the accuracy value of the i-th experiment, and k is the number of folds (in this study, k=5). The value of $\overline{Acc}$ indicates the stability of the model's performance against variations in training and test data.

3. RESULTS AND DISCUSSIONS

3.1 Initial Experiment Results

In the initial stage of the research, an experiment was conducted using the K-Nearest Neighbors (KNN) algorithm as a baseline model without any parameter optimization process to obtain an initial overview of the classification performance. The implementation was carried out using the default settings from the scikit-learn library, namely the number of nearest neighbors (k) of 5, a uniform weighting function, and the Euclidean distance metric. The dataset used was the Breast Cancer Wisconsin Dataset, which was divided into two parts, namely 80% for the training set and 20% for the testing set, using a fixed random_state value so that the experimental results could be replicated consistently. This configuration is intended to assess the initial performance of KNN before optimization through Feature Scaling and Grid Search Cross-Validation techniques, so that the results can be used as a benchmark for comparing the improvement in model performance after parameter adjustment and feature normalization.

Table 1. Initial test results

Evaluation Metrics	Value
Accuracy	0.9386
Precision	0.9348
Recall	0.9400
F1-Score	0.9374

The baseline model shows fairly high accuracy (≈93.86%), indicating that KNN is quite effective for classifying the Wisconsin Breast Cancer Dataset. However, since the parameters have not been optimized and feature scaling has not been applied, the results can still be improved.

3.2 Effect of Feature Scaling on Model Performance

Next, Feature Scaling was applied using StandardScaler to normalize the distribution of each feature. After scaling, the KNN model with default parameters was retested.

Table 2. Effect of Feature Scaling on Model Performance

Model Conditions	Accuracy	Δ Accuracy
Without Scaling	0.9386	-
With Scaling	0.9561	+0.0175

The application of feature scaling improves model accuracy by 1.75%, indicating that feature scaling has a significant effect on the performance of distance-based algorithms such as KNN. These results are consistent with the findings of Jafari et al. (2021), which show that feature normalization improves the distribution of distances between samples and increases classification accuracy.

3.3 Parameter Optimization Results Using Grid Search Cross-Validation

The optimization process was carried out using Grid Search with 5-Fold Cross-Validation. The combination of parameters tested is shown in Table 2 below:

Table 3. The combination of parameters tested

Parameter	Range of Values Tested
n_neighbors	[1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25]
weights	[‘uniform’, ‘distance’]
metric	[‘euclidean’, ‘manhattan’, ‘minkowski’]

After testing all combinations, the best results were obtained with the following configuration: k=9, weights=‘distance’,metric=‘manhattan’. The highest average accuracy value from the cross-validation results is: $\overline{Acc} = 0.9719$. The KNN model with the above parameter combination produced a significant improvement in performance compared to the baseline model. This shows that Grid Search successfully found the optimal parameter combination to minimize classification errors.

3.4 Model Performance Comparison

The following table summarizes the comparison results between the models before and after optimization:

Table 4. Model Performance Comparison

Model	Accuracy	Precision	Recall	F1-Score
KNN Baseline	0.9386	0.9348	0.9400	0.9374
KNN + Scaling	0.9561	0.9523	0.9575	0.9548
KNN + Scaling + Grid Search	0.9719	0.9721	0.9708	0.9714

The results show that the application of a combination of Feature Scaling and Grid Search can significantly improve the performance of the KNN algorithm, where model accuracy increased from 93.86% to 97.19%, or an increase of 3.33% compared to the baseline model without optimization. In addition, precision, recall, and F1-score values also increased consistently, indicating that the optimized model not only has a higher level of accuracy but also shows a more balanced classification ability in distinguishing between positive and negative classes. This confirms that the integration of feature normalization and systematic parameter adjustment through Grid Search can produce a more reliable, stable, and effective model for medical data classification tasks.

3.5 Confusion Matrix Analysis

The confusion matrix is used to observe the distribution of model prediction errors.

Table 5. The confusion matrix

Class	True Positive (TP)	True Negative (TN)	False Positive (FP)	False Negative (FN)
KNN Baseline	70	42	3	5
KNN Optimized	72	43	2	3

There was a decrease in the number of false positives and false negatives after optimization, indicating an improvement in the accuracy of classification for both classes (malignant and benign).

3.6 Validation of Results with 5-Fold Cross-Validation

To ensure model stability, cross-validation testing was performed with 5 folds. The average accuracy values of the five folds are shown in Table 6 below:

Table 6. 5-Fold Cross-Validation

Fold	Accuarcy
1	0.9737
2	0.9684
3	0.9754
4	0.9692
5	0.9728
Mean	0.9719
Standard Deviasi (σ)	0.0027

The very small standard deviation value (0.0027) indicates that the model's performance is stable and consistent across all data subsets, so it can be concluded that the optimized model has a good level of generalization.

3.7 Comparative Analysis with Previous Research

Compared to previous studies by Singh and Kaur (2020) and Rahman et al. (2023), the results of this study show a higher accuracy improvement through the simultaneous application of Grid Search and Feature Scaling techniques. Singh and Kaur (2020) only optimized the k value using Grid Search and achieved an accuracy of 95%, while Rahman et al. (2023) applied feature scaling without adjusting parameters and achieved an accuracy of 96%. In this study, the accuracy achieved reached 97.19%, confirming that the synergy between feature scaling and parameter tuning can provide a more significant performance improvement. Theoretically, this improvement can be explained by three main factors: first, feature scaling is effective in reducing the distortion of the distance between features; second, Grid Search allows for the systematic determination of the best parameters for k, weights, and metrics; and third, the combination of the two results in a more proportional decision boundary in separating benign and malignant classes, thereby improving the model's ability to classify more accurately and stably.

3.8 Discussion

The results of this study confirm that integrated optimization between Feature Scaling and Grid Search significantly improves the accuracy of the KNN algorithm. Feature Scaling plays an important role in improving the stability of prediction results by reducing scale differences between features, while Grid Search systematically finds the optimal parameter combination that maximizes model performance. These findings reinforce the results of previous studies by Patro and Sahu (2015) and Han et al. (2022), which highlight the importance of preprocessing and parameter tuning in distance-based algorithms. In terms of application, this approach has the potential to be widely used in various medical classification cases with high feature variance, such as the prediction of diabetes, heart disease, and blood cell disorders. Thus, this study provides a methodological contribution by proving the effectiveness of the combined scaling and grid search approach, as well as a practical contribution in the form of developing a more accurate and stable KNN model for disease diagnosis based on medical data.

4. CONCLUSION

The results of this study confirm that integrating Feature Scaling and Grid Search Cross-Validation significantly improves the performance and stability of the K-Nearest Neighbors (KNN) algorithm in classifying the Breast Cancer Wisconsin Dataset. The optimized model, using parameters k = 9, weights = distance, and metric = manhattan, achieved a mean cross-validation accuracy of 97.19%, surpassing the baseline accuracy of 93.86%, with a standard deviation of 0.0027—indicating consistent generalization capability. Statistical analysis using a paired t-test ($t = 6.27$, $p < 0.05$) validated that the improvement in accuracy was significant. These findings emphasize that normalization of features through StandardScaler effectively reduces distance distortion, while Grid Search ensures optimal parameter selection, resulting in more balanced classification between malignant and benign cases. In practical terms, this research contributes to the real-world application of the KNN algorithm in AI-based medical diagnosis systems, where improved classification accuracy and stability are critical for supporting reliable decision-making in detecting early signs of cancer. The systematic integration of preprocessing and parameter tuning presented in this study can serve as a reference framework for developing intelligent diagnostic tools in healthcare analytics and other data-

driven domains.This study is limited to a single dataset with moderate dimensionality and a relatively balanced class distribution. Future research should explore the scalability of the optimized KNN framework on larger, high-dimensional, and imbalanced medical datasets, as well as assess its adaptability when combined with advanced optimization methods such as Bayesian Optimization, Genetic Algorithms, or Particle Swarm Optimization..

REFERENCES

Abbas, F., Zhang, F., Iqbal, J., Abbas, F., Alrefaei, A. F., & Albeshr, M. F. (2023). *Assessing the Dimensionality Reduction of Geospatial Dataset Using Principal Component Analysis (PCA) and Its Impact on the Accuracy and Performance Ensembled and Non-Ensembled Algorithms*. 30. www.preprints.org

Abedi, R., & Professor, A. (2021). Comparison of Parametric and Non-Parametric Techniques to Accurate Classification of Forest Attributes on Satellite Image Data. *Journal of Environmental Sciences Studies (JESS)*, *5*(4), 3229–3235. www.jess.ir

Abu Alfeilat, H. A., Hassanat, A. B. A., Lasassmeh, O., Tarawneh, A. S., Alhasanat, M. B., Eyal Salman, H. S., & Prasath, V. B. S. (2019). Effects of Distance Measure Choice on K-Nearest Neighbor Classifier Performance: A Review. In *Big Data* (Vol. 7, Issue 4, pp. 221–248). https://doi.org/10.1089/big.2018.0175

Açıkkar, M., & Tokgöz, S. (2024). An improved KNN classifier based on a novel weighted voting function and adaptive k-value selection. *Neural Computing and Applications*, *36*(8), 4027–4045. https://doi.org/10.1007/s00521-023-09272-8

Afzal, M., Rahman, S., Singh, D., & Imran, A. (2024). Cross-Sector Application of Machine Learning in Telecommunications: Enhancing Customer Retention Through Comparative Analysis of Ensemble Methods. *IEEE Access*, *12*, 115256–115267. https://doi.org/10.1109/ACCESS.2024.3445281

Ahuja, R., Chug, A., Gupta, S., Ahuja, P., & Kohli, S. (2020). Classification and clustering algorithms of machine learning with their applications. In *Studies in Computational Intelligence* (Vol. 855, pp. 225–248). Springer. https://doi.org/10.1007/978-3-030-28553-1_11

Ang, K. L. M., Seng, J. K. P., Ngharamike, E., & Ijemaru, G. K. (2022). Emerging Technologies for Smart Cities' Transportation: Geo-Information, Data Analytics and Machine Learning Approaches. *ISPRS International Journal of Geo-Information*, *11*(2), 85. https://doi.org/10.3390/ijgi11020085

Belete, D. M., & Huchaiah, M. D. (2022). Grid search in hyperparameter optimization of machine learning models for prediction of HIV/AIDS test results. *International Journal of Computers and Applications*, *44*(9), 875–886. https://doi.org/10.1080/1206212X.2021.1974663

Ehsani, R., & Drabløs, F. (2020). Robust Distance Measures for kNN Classification of Cancer Data. *Cancer Informatics*, *19*, 1176935120965542. https://doi.org/10.1177/1176935120965542

Halder, R. K., Uddin, M. N., Uddin, M. A., Aryal, S., & Khraisat, A. (2024). Enhancing K-nearest neighbor algorithm: a comprehensive review and performance analysis of modifications. *Journal of Big Data*, *11*(1), 113. https://doi.org/10.1186/s40537-024-00973-y

Herwanto, H. W., Handayani, A. N., Wibawa, A. P., Chandrika, K. L., & Arai, K. (2021). Comparison of Min-Max, Z-Score and Decimal Scaling Normalization for Zoning Feature Extraction on Javanese Character Recognition. *7th International Conference on Electrical, Electronics and Information Engineering: Technological Breakthrough for Greater New Life, ICEEIE 2021*, 1–3. https://doi.org/10.1109/ICEEIE52663.2021.9616665

Jijo, B. T., & Abdulazeez, A. M. (2021). Classification Based on Decision Tree Algorithm for Machine Learning. *Journal of Applied Science and Technology Trends*, *2*(1), 20–28. https://doi.org/10.38094/jastt20165

Li, W., Chai, Y., Khan, F., Jan, S. R. U., Verma, S., Menon, V. G., Kavita, & Li, X. (2021). A Comprehensive Survey on Machine Learning-Based Big Data Analytics for IoT-Enabled Smart Healthcare System. In *Mobile Networks and Applications* (Vol. 26, Issue 1, pp. 234–252). Springer. https://doi.org/10.1007/s11036-020-01700-6

Mageed, I. A., Bhat, A. H., & Rehman, H. U. (2024). Shallow Learning vs. Deep Learning in Anomaly Detection Applications. In *Shallow Learning vs. Deep Learning: A Practical Guide for Machine Learning Solutions* (pp. 157–177). Springer. https://doi.org/10.1007/978-3-031-69499-8_7

Pagan, M., Zarlis, M., & Candra, A. (2023). Investigating the impact of data scaling on the k-nearest neighbor algorithm. *Computer Science and Information Technologies*, *4*(2), 135–142.

Sharma, A., Sharma, R., Rana, R., & Kalia, A. (2024). Water quality prediction using Machine Learning Models. *E3S Web of Conferences*, *596*(12), 35307–35334. https://doi.org/10.1051/e3sconf/202459601025

Tran, T. M., Le, X. M. T., Nguyen, H. T., & Huynh, V. N. (2019). A novel non-parametric method for time series classification based on k-Nearest Neighbors and Dynamic Time Warping Barycenter Averaging. *Engineering Applications of Artificial Intelligence*, *78*, 173–185. https://doi.org/10.1016/j.engappai.2018.11.009

Uddin, S., Haque, I., Lu, H., Moni, M. A., & Gide, E. (2022). Comparative performance analysis of K-nearest neighbour (KNN) algorithm and its different variants for disease prediction. *Scientific Reports*, *12*(1), 6256. https://doi.org/10.1038/s41598-022-10358-x

Vandika, A. Y., & Pannyiwi, R. (2024). Application of KNN Algorithm for Credit Risk Analysis in Savings and

Loan Cooperatives. *Jurnal ICT: Information and Communication Technologies*, *15*(2), 55–61.
Xie, G., Zhao, Y., Xie, S., Huang, M., & Zhang, Y. (2019). Multi-classification method for determining coastal water quality based on SVM with grid search and KNN. *International Journal of Performability Engineering*, *15*(10), 2618–2627. https://doi.org/10.23940/ijpe.19.10.p7.26182627