 ISSN 2579-7298

Salt Quality Classification Using Backpropagation Neural Network and K-Nearest

Anas Mahmudi [a,1,*], Zainul Abidin [b,2], Angger Abdul Razak [b,3]

[a] Department of electrical engineering, Brawijaya University Malang, 65145, Indonesia.
[1] *anasmahmudi176@student.ub.ac.id* *; [2] zainulabidin@ub.ac.id; [3] *aarazak@ub.ac.id*

ARTICLE INFO

Article history
Received 13 July 2025
Revised 12 October 2025
Accepted 12 December 2025
Available online 30 December 2025

Keywords
Salt Quality Classification,
Backpropagation,
Neural Network,
K-Nearest Neighbor,
Machine Learning

ABSTRACT (10PT)

Salt quality plays a vital role in determining its usability across various sectors, including food, pharmaceuticals, and industrial applications. Traditional methods of classifying salt quality, which rely heavily on manual inspection and laboratory testing, are often time-consuming, costly, and prone to human error. In response to these limitations, this study explores the implementation of machine learning techniques—specifically, Backpropagation Neural Network (BPNN) and K-Nearest Neighbor (K-NN)—to classify salt quality based on its physical and chemical properties. The features used in this research include NaCl concentration, moisture content, magnesium levels, sulfat, insoluble, calcium, NaCL(wb) and NaCL(db) which are commonly used indicators of salt purity and grade. The BPNN model is designed to handle complex and non-linear relationships within the dataset by adjusting weights through iterative backpropagation during training. Meanwhile, the K-NN algorithm serves as a simpler, instance-based learning method that classifies samples based on the majority class of their nearest neighbors in the feature space. Comparative experiments were conducted to evaluate the classification and computational efficiency of both models. Results indicate that both methods are effective in classifying salt into predefined quality categories. However, BPNN consistently outperforms K-NN in terms of time efficiency and generalization, particularly when handling noisy or overlapping data. The findings underscore the potential of integrating artificial intelligence into quality control systems in the salt industry, offering a faster, more objective, and scalable solution for ensuring product standards.

1. Introduction `

Salt is an important mineral that is widely used in various sectors, including the food industry, pharmaceuticals, agriculture, and chemical manufacturing. The quality of salt plays a crucial role in determining its functionality and value, especially for applications that require quick turnaround times. Traditionally, salt quality classification is done through manual calculations where parameters are computed one by one and laboratory tests are conducted, which often take a long time, are subjective, and prone to human error. This conventional method is increasingly challenged by the demand for faster, more accurate, and consistent quality control in modern production environments.

To address these limitations, computational approaches based on machine learning algorithms have been introduced into quality assessment systems. Among the various available methods, *Backpropagation Neural Network* (BPNN) and *K-Nearest Neighbor* (K-NN) are two widely used and proven techniques for solving classification problems [2]. BPNN is capable of learning complex, non-linear relationships within data using a multi-layer network structure, while K-NN provides a simpler approach based on the proximity of feature values between data points.

This research aims to develop a salt quality classification system using BPNN and K-NN algorithms. The system utilizes salt parameters such as moisture content, insoluble substances, calcium, magnesium, sulfate, Nacl (wb), and Nacl (db). Through comparative analysis, this study evaluates the speed and performance of both methods to determine the most effective approach in supporting automated and efficient salt quality classification.

Classification is the process of grouping objects or data into specific classes or categories based on their characteristics or features**.** In the context of machine learning, classification refers to the use of algorithms to predict the label or class of a given data point based on previously available training data. In the field of salt processing, classification plays a crucial role due to the presence of seven distinct parameters and varying quality levels of salt. Through classification, each type of salt quality can be aligned with established quality standards and tailored to meet the diverse demands of the market—whether for household use, the food industry, pharmaceuticals, or non-food industries. This process not only aids producers in planning production and distribution more accurately but also provides consumers with confidence in selecting the appropriate product. Furthermore, classification supports better decision-making. With a classification system in place, identifying and managing salt quality becomes more effective and efficient in determining product standards.

Therefore, this study focuses on the classification of salt to determine the accuracy level in assessing its quality. Evaluating salt quality requires several important parameters, including moisture content, insoluble substances, calcium, magnesium, sulfate content, and NaCl levels in both wet basis (wb) and dry basis (db) forms. Based on these parameters, salt is classified into four quality levels: Quality 1 (K1), Quality 2 (K2), Quality 3 (K3), and Quality 4 (K4). To be categorized as Quality 1, specific parameter values must be met, such as a moisture content of 9.73, insoluble matter at 56.57, calcium at 67.27, magnesium at 68.00, sulfate at 60.32, NaCl (wb) at 84.10, and NaCl (db) at 93.10. Salt classified as K1 is generally used for consumption purposes, particularly as a food seasoning. K2 is also used for consumption, but the main difference between K1 and K2 lies in their market value, with K1 having a higher selling price. Meanwhile, K3 salt is commonly used in industrial applications such as pharmaceutical preservation and textile manufacturing. K4 is allocated for other industrial needs, including livestock feed and various non-food sectors. Among the four categories, K1 represents the highest quality due to its superior parameter values, while K4 is considered the lowest, reflecting inferior quality standards. In addition to the challenges of producing high-quality salt, accurately classifying salt quality also presents a significant challenge. Therefore, this study conducts a comparative analysis between two classification methods Backpropagation neural network and K-Nearest neighbor, to determine which provides better accuracy in assessing salt quality.

The structure of this paper is arranged as follows: Section II discusses the methods related to BPNN and K-NN. Section III presents the results, discussion, and evaluation. Finally, Section IV provides the conclusion along with recommendations for future research.

2. Method

The methods and approaches that will be used in this research will be explained. It will also discuss the techniques to be employed, the types and methods of data acquisition, analysis, research steps, and testing.

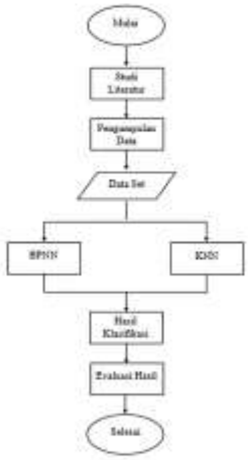


Fig. 1. The proposed research flowchart has several stages, namely

Proposed research flow diagram

1. The first stage of data collection, this data is obtained from PT. located in Sumenep district. The data obtained is then classified to achieve efficiency from that classification.
2. The second stage of the data set will be classified using two methods. These two methods are Backpropagation neural network and K-nearest neighbor.
3. In the third stage, the classification process will enhance time efficiency for the quality of salt that has been classified using Backpropagation neural network and K-nearest neighbor methods, covering seven parameters that standardize the quality of the salt.
4. In the evaluation stage of the results, the results from the Backpropagation neural network method and K-nearest neighbor method are compared determine which of the two methods is more efficient in terms of the time achieved.

Backpropagation neural network

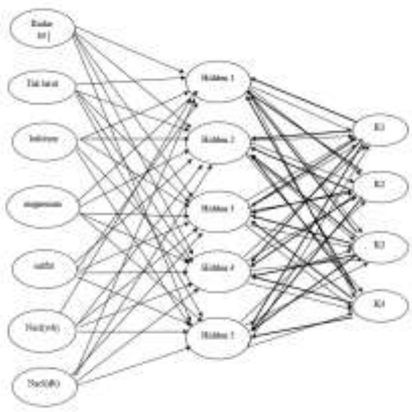


Fig. 2. Diagram Alir Proces Backpropagation Neural Network

The Backpropagation neural network, as shown in Figure… consists several stages, which are as follows:

1. Training and Testing Data Division
   After the data is processed, it is divided into two sets: training data and testing data. Training data is the data used to train the model, while testing data is the input data that will be compared with the training data. The data used in this study is salt production data from Sumenep. The training and testing data are divided with a 70% training data and 30% testing data ratio.
2. Input Parameters
   Input the parameters based on the training data that has been divided earlier.
3. Hidden Neurons, Error Tolerance, Learning Rate
   Determine the number of hidden neurons, error tolerance, and learning rate based on the previously processed data.
4. The next step is Hidden Layer Activation and Output Activation, which are determined based on the previous data using the sigmoid function.
5. Optimization
   The next stage is the optimization process, where actions are performed to achieve the optimal results.
6. Testing (Backpropagation)
   After optimization, the data undergoes testing.
7. BPNN Calculation
   The data is processed and calculated using the Backpropagation Neural Network (BPNN) method.
8. Results After the calculations are completed using the BPNN method, the results are obtained.
9. Analysis and Evaluation
   The results obtained are then analyzed and evaluated further.

Backpropagation neural network is a type of supervised learning algorithm that consists of multiple layers. This method adjusts the weights by propagating the output error backward through the network. To compute this error, the forward propagation process must be completed first (Andrijasa & Mistianingsih, 2010).

Recognition method is the process of initializing data that will later be processed by the backpropagation neural network. The data to be recognized is presented in the form of vectors, each with a corresponding target, also represented as a vector. The target, or reference output, serves as a character map indicating the position of the input vector. Meanwhile, the training method involves learning to recognize the data and storing the acquired knowledge or information into the network's weights.

There are three phases in training a BPNN the feedforward phase, the backpropagation phase, and the weight adjustment phase. In the feedforward phase, the input pattern is processed forward from the input layer through to the output layer. During the backpropagation phase, each output unit receives the target pattern associated with the input and calculates the error. This error is then propagated backward through the network. The purpose of the weight adjustment phase is to reduce the calculated error. These three phases are repeated continuously until a stopping condition is met (Jumarwanto, 2009).

K-nearest neighbor

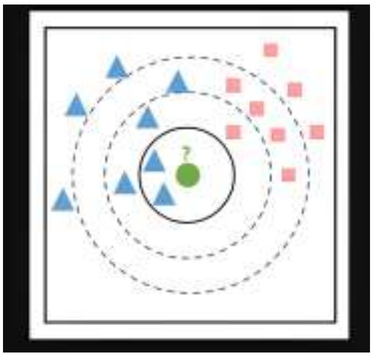


Fig. 3. Diagram Alir Proces K-nearest Neighbor.

The K-nearest neighbor, as shown in Figure 3 consists several stages, which are as follows:

1. K-Nearest neighbors algorithm workflow
2. Data Collection
   Collect and prepare the dataset that includes input features and their corresponding class labels (for classification) or values (for regression).
3. Preprocessing
   Clean the data and normalize or scale the features if necessary. This step helps improve the accuracy of distance calculations.
4. Select the number of neighbors (K) to consider. This is a key parameter that affects the algorithm's performance.
5. Calculate Distance
   For each test data point, compute the distance between it and all data points in the training set. Common distance metrics include Euclidean distance
6. Find K-Nearest Neighbors
   Identify the training data points that have the smallest distances to the test point.
7. Make a Prediction
   Classification: Use majority voting among the K nearest neighbors (the class that appears most frequently).
   Regression: Compute the average
8. Return the result
9. Evaluate performance

Confusion Matrix

The confusion matrix is a very popular calculation for solving classification problems. It can be used for binary class classification as well as multi-class classification. The confusion matrix represents the calculation of predicted values and actual results. There are four cells in the confusion matrix. (Sharma et al, 2022).

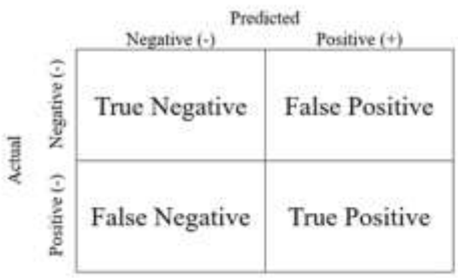


Fig. 4. Fig.4.Confusion Matrix

. True Negative represents data that has a negative value but is predicted correctly.
. True Positive represents data that has a positive value and is predicted correctly.

 ISSN 2579-7298

. False Positive represents data that has a positive value but is predicted incorrectly.

. False Negative represents data that has a negative value and is predicted incorrectly.

Accuracy

Accuracy shows the number of correct predictions from the tested model. It measures how well the model can correctly classify the classes in the dataset. Accuracy can be calculated by adding true negatives and true positives, then dividing by the total dataset.

$$Accuracy = \frac{TP+TN}{TP+TN+FN+F}$$

Precision (positive Predictive value)

Precision indicates the evaluation of the results of the model that are positive and predicted as positive. This metric helps in measuring how precise the model is in classifying the positive class.

$$Precision = \frac{TP}{TP+FP}$$

Recall atau Sensitivy (True Positive Rate)

Recall indicates the evaluation of the model's results that are actually positive from the data that is truly positive.

$$Recall = \frac{TP}{TP+FN}$$

F1-Score

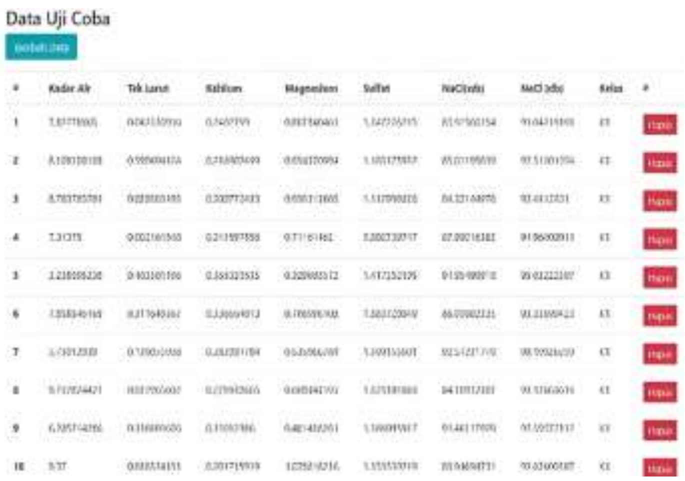
F1-score is a combined metric of Precision and Recall. This metric emphasizes both factors of Precision and Recall.

$$F1\text{-}Score = \frac{Precision\ x\ Recall}{Precision+Recall}$$

3. Results and Discussion

Testing data

Data used in this research is about the quality of salt obtained from PT. with a total of 350 data points and 7 features or criteria. Below is an example of the test data used in this research.


Data Uji Coba

#	Kadar Air	Tak larut	Kalsium	Magnesium	Sulfat	NaCl(wb)	NaCl (db)	Kelas	#
1	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	Hapus
2	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	Hapus
3	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	Hapus
4	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	Hapus
5	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	Hapus
6	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	Hapus
7	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	Hapus
8	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	Hapus
9	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	Hapus
10	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	Hapus

Fig.5.Test Data Clasification

Test data is a set of data used to evaluate the performance or accuracy of a machine learning model after it has been trained using training data. The primary purpose of test data is to assess how well the model can generalize to new, unseen data.

System testing

System testing is a critical aspect aimed at evaluating the performance of the developed system in classifying salt quality. The testing process involves comparing two methods: Backpropagation Neural Network and K-Nearest Neighbor. These methods are assessed by calculating precision, recall, and F1-score values. The following section presents the testing scenarios used in this study.

Table 1.

scenario	**Table Column Head**		
	Hidden Layer	***Learning Rate***	***Evaluasi***
Scenario 1………….dst	1………………5	0,1…………..0,9	Presisi, Racall, F1 Score

Testing method scenario Backpropagation Neural Network

In the testing phase, using a hidden layer of 4 and a learning rate between 0.7 and 0.9, the model achieved a precision score of 55%, a recall of 67.70%, and an F1-score of 60.69%.

Table 2.

scenario	**Table Column Head**	
	Nilai K	***Evaluasi***
Scenario 1………….dst	Bernilai ganjil tidak > data training	Presisi, Racall, F1 Score

Testing Method Scenario K-Nearest Neighbor

Based on the test results shown in Table 5.6, the input value for K-1 achieved a precision of 76%, a recall of 76%, and an F1-score of 76%

Confusion Matrix

Confusion Matrix

	K1	K2	K3	K4
K1	3	3	2	0
K2	2	2	0	2
K3	0	0	3	0
K4	0	0	0	3

Fig.6.Confusion Matrix Backpropagation Neural Network

Based on the values obtained from the confusion matrix, calculations were performed to evaluate the model's performance using precision, recall, and F1 score. The results showed a precision of 55%, recall of 67.70%, and an F1 score of 60.69%. Each metric serves a specific purpose: precision indicates how accurately the model predicts positive data, recall reflects how well the model identifies actual negative data, and the F1 score measures the balance between precision and recall for the BPNN model.

Confussion Matrix

	K1	K2	K3	K4
K1	88	15	1	9
K2	14	108	9	2
K3	5	0	26	1
K4	15	3	0	16

Fig.7.Confusion Matrix K-nearest neihbor

Using the values from the confusion matrix, the model's performance was evaluated based on precision, recall, and F1 score. The results showed a precision of 76%, a recall of 76%, and an F1 score of 76%. Each metric has a specific role: precision measures how accurately the model predicts positive instances, recall measures the model’s ability to correctly identify negative instances, and the F1 score reflects the balance between precision and recall for the KNN model.

4. Conclusion

Based on the research results, the K-Nearest Neighbor (KNN) method performed better than the Backpropagation Neural Network (BPNN) in classifying salt quality.

However, KNN has a weakness in determining the K value. As the K value increases, precision, recall, and f1-score tend to decrease. Therefore, it is important to compare distance calculation methods to improve classification accuracy.

References

[1] Erza Anggara Verbiawan, Ketut Sumaba, “Spray nozzle technology to accelerate the evaporation of seawater in conventional salt production (Chemical Engineering, Faculty of Engineering, National Development University "Veteran" East Java, Surabaya) Chemical Engineering Journal Vol. 18, No. 1, October 2023. http://ejournal.upnjatim.ac.id/

[2] Dilla Kurniati, Application of probabilistic neural networks and backpropagation neural networks for the classification of salt production in Indonesia. https://www.academia.edu/

[3] Asroni, A., Fitri, H., dan Prasetyo, E, Application of clustering methods using the k-means algorithm for grouping data of new student candidates at Muhammadiyah University of Yogyakarta. https://doi.org/10.18196/st.211211

[4] Wishnu Padma, The use of table salt as a stabilizing material for shear parameters of Lampung soil (Civil Engineering, Faculty of Engineering, Muhammadiyah University of Surakarta)

[5] Elfiana, mukhlis, Classification of feasibility for providing capital to salt farmer groups using K-Nearest Neighbor in the context of community economic empowerment in Bireun District (Faculty of Agriculture, Almuslim University).

[6] Galih Wahyu Pratama, The effectiveness of guided inquiry on acid-base salt materials in improving classification and communication skills (FKIP University of Lampung, Jl. Prof. Dr. Soemantri Brojonogoro No.1)

[7] Elfiana, mukhlis, Classification of feasibility for providing capital to salt farmer groups using K-Nearest Neighbor in the context of community economic empowerment in Bireun District (Faculty of Agriculture, Almuslim University)

[8] Wishnu Padma, The use of table salt as a stabilizing material for shear parameters of Lampung soil (Civil Engineering, Faculty of Engineering, Muhammadiyah University of Surakarta)

[9] Andriyanto, E., dan Melita, Y, Introduction to human characteristics through palm line patterns using probabilistic neural network methods. Asian Journal of Information Technology, 7(2), 1–31

[10] Asroni, A., Fitri, H., dan Prasetyo, The application of the clustering method with the k-means algorithm in grouping data of prospective new students at Muhammadiyah University Yogyakarta (case study: Faculty of Medicine and Health Sciences, and Faculty of Social and Political Sciences). Semesta Teknika, 21(1), 60–64.

[11] Chauhan, H., dan Chauhan, A, Implementation of the C4.5 decision tree algorithm. International Journal of Scientific and Research Publications, 3(10), 1–3.

[12] Ciarelli, P. M., Oliveira, E., Badue, C., dan De Souza, Multi-label text categorization using a probabilistic neural network. International Journal of Computer Information Systems and Industrial Management Applications (IJCISIM), 1, 133–144.

[13] Lestari, N., dan Van Fc, L. L, Implementation of artificial neural networks to assess the feasibility of final projects for students (case study at AMIK Bukittinggi). Digital Zone: Journal of Information and Communication Technology, 8(1), 10–24.

[14] Luvia, Y. S., Windarto, A. P., Solikhun, S., dan Hartama, The application of the C4.5 algorithm for classifying student success predictions at AMIK Tunas Bangsa. Jurasik (Journal of Information Systems Research and Informatics Engineering), 1(1), 75–79.

[15] Mustakim,). Effectiveness of k-means clustering to distribute training data and testing data on k-nearest neighbor classification. Journal of Theoretical and Applied Information Technology, 95(21), 5693-5700.

[16] Nurbaiti, F. A. S., Midyanti, D. M, Identification of seedlings in peatland plants based on leaf shape using a website-based probabilistic neural network (seedling age 2 months-1 year). Coding Journal of Computer and Applications, 5(1).

[17] Repi Ramadani, Elvia Budianita, Febi Yanto, and Siska Kurnia Gusti, Classification of Coronary Heart Disease Using Backpropagation Neural Network Method. National Seminar on Information Technology, Communication, and Industry (SNTIKI) ISSN (Printed): 2579-727 ISSN (Online) 2579-5406.

[18] Mohd. Azhima, Iis Afrianty, Elvia Budianita, Siska Kurnia Gusti, Application of Backpropagation Neural Network Method for Stroke Disease Classification. KLIK: Scientific Study of Informatics and Computer ISSN 2723-3898 https://djournals.com/klik.

[19] Saifur Yusuf Kurniawan, Classification of Drinking Water Quality Using Backpropagation Neural Network Based on Missing Value Handling and Normalization Journal of Information System Research E-ISSN: 2686-228X.

[20] Wise Herowati, Ricardus Anggi Pramunendar, and Harun Al Azies, Optimization of the BPNN (Backpropagation Neural Network) method Using GA (Genetic Algorithm) in Determining the Offset Direction in the GLCM (Gray Level Co-occurrence Matrices) Feature Extraction Method by Badroe Zaman, Badroe Zaman, and. BINA INSANI ICT JOURNAL ISSN: 2355-3421 (Print) ISSN: 2527-9777.

[21] Erfan Hasmin Hasmin, Nurul Aini, Data Mining For Inventory Forecasting Using Double Exponential Smoothing Method. IEEE Access, Conference Paper. International Conference on Cybernetics and Intelligent System (ICORIS).

[22] Yusuf, M, R., Valensyah, M, A., Gunawan, W, Application of the K-Nearest Neighbor (KNN) Algorithm in Predicting and Calculating the Accuracy Level of Weather Data in Indonesia. Engineering and Science Journal, No. 2, Vol. 2, 11-16.

[23] Naja, M, M., Harliana, Sukerti, S., Herdian, R,M, Application of K-Nearest Neighbor (KNN) Algorithm to Predict Stroke Disease. Intech Scientific Journal, No. 1, Vol. 4, 130-140.

[24] N. F. Romdhoni, K. Usman, and B. Hidayat, Detection of Soybean Quality Through Digital Image Processing using Gray-Level Co-Occurrence Matrix (GLCM) Method and Decision Tree Classification. Pros. Semin. Nas. Ris. Dan Inf. Sci., vol. 2, pp. 132–137.

[25] M. Ramadhani, Classification of Acne Types Based on Texture Using the GLCM Method," e-Proceeding of Engineering, vol. 5, no. 1, pp. 870–876.