p-ISSN 2089 – 0133 e-ISSN 2477-6416
URL : https://jurnal.uns.ac.id/ijap/article/view/85064
Indonesian Journal of Applied Physics (IJAP) Vol. 14 No. 2 page 301
DOI : https://doi.org/10.13057/ijap.v14i2.85064

MACHINE LEARNING-BASED COW MILK QUALITY CLASSIFICATION USING RECURSIVE FEATURE ELIMINATION CROSS-VALIDATION

Damar Wicaksono*[1], Affix Mareta[1], Ardy Erdiyanto[2], Nuzula Afianah[3], Rafly Ramadhani[4]

[1] Department of Information Technology, Faculty of Engineering, Universitas Tidar, Magelang Utara, Indonesia
2 Neurabot AI Laboratory, Jl. Bantul 9, Krandohan, Pendowoharjo, Sewon, Bantul, Yogyakarta, Indonesia
[3] Department of Engineering, Politeknik Negeri Jember, Sumbersari, Jember, Indonesia
[4] Department of Electrical Engineering, Faculty of Engineering, Universitas Tidar, Magelang Utara, Indonesia
*damar@untidar.ac.id

Received 29-02-2024, Revised 01-09-2024, Accepted 18-10-2024,
Available Online 18-10-2024, Published Regularly October 2024

ABSTRACT

Milk quality is of paramount importance as it directly impacts consumer health and well-being. High-quality milk is rich in essential nutrients such as calcium, protein, and vitamins, contributing to overall nutrition. Moreover, ensuring milk quality is vital for preventing the transmission of diseases and contaminants through dairy products. Therefore, research in this field is essential to guaranteeing the safety and nutritional value of milk consumed by individuals of all ages. In this paper, the design of machine learning-based grade measuring devices with recursive feature elimination with cross-validation (RFECV) is carried out as a guide in the design of a milk grade detection system. The milk is rated as low, medium, or high based on these criteria. The sensors will gather this information from the milk with the aid of the microcontroller. The algorithms utilized in this study and the results obtained from K-Nearest Neighbors (KNN) combined with the RFECV algorithm have a higher accuracy value: 5.10% higher than the support vector machine (SVM) model, 12.60% higher than single K-Nearest Neighbours (KNN) and 14.05% higher than the random forest (RF) model trained without RFECV. Using seven input features (pH, temperature, taste, odor, fat, turbidity, and color), the proposed model produces 96.27% accuracy.

Keywords: milk; grade; machine learning; algorithm; accuracy

Cite this as: Wicaksono, D., Mareta, A., Erdiyanto, A., Alfianah, N., & Ramadhani, R. 2024. Machine Learning-Based Cow Milk Quality Classification Using Recursive Feature Elimination Cross-Validation. *IJAP: Indonesian Journal of Applied Physics, 14*(2), 301-314. doi: https://doi.org/10.13057/ijap.v14i2.85064

INTRODUCTION

Cow milk is a fat phase in water, which contains several minerals such as protein in colloidal form, sugar, and salt [1]. So, the taste of pure cow milk that we often drink tastes sweet and salty because of the content in it and also because of the fat in the milk. The fat in cow milk is in the dispersion phase, which causes the cow milk fat to form grains called globules. Each grain of cow milk fat is surrounded by a thin protein membrane. This cow milk fat itself is always at the top of the cow milk when the cow milk has been milked due to adsorption. Cow milk has a bluish-white to brown colour, but the colour of cow milk can vary depending on the type of cow and also the type of food the cow eats [2].According to the Indonesian National Standards

Agency 2011 food number 3144.1, which concerns the quality of pure cow milk, cow milk must contain nutrition and also be food-safe [3-4].

Among them are the requirements for maximum microbial content, dangerous metal contamination contained in milk, and antibiotic residues. Apart from that, other factors must also be considered in order to get good-quality cow milk by paying attention to the cleanliness and sanitation of the dairy cow pen, the health of the cow, and the tools that will be used. After being used to milk, the cow must be clean [5-6].

A range of studies have utilized machine learning to assess cow milk quality. Slob [7] identified milking parameters and milk properties as key variables, while Frizzarin [8] found that statistical machine learning methods, such as random forests and support vector machines, improved prediction accuracy for milk quality traits. Mu [9] made it with three machine learning algorithms, namely logistic regression (LR), support vector machine (SVM), and random forest (RF), are used to construct the classification model of milk source (dairy farm) quality identification. Jiménez [10] explained about dairy products and the different multivariate analytical solutions to evaluate the quality and the authenticity of them by analytical solution was discussed. In addition, an overview of the applications of machine learning methods reported in the literature, such as artificial neural network, support vector machine, and random forest, to ensure the quality and authenticity of dairy products and milk is presented. While Fuentes [10] and Sugiono [12] both applied machine learning to predict milk yield and quality, with Fuentes focusing on the impact of heat stress and environmental parameters, and Sugiono investigating the influence of physiological and environmental factors. These studies collectively demonstrate the potential of machine learning in enhancing the assessment of cow milk quality.

By knowing the indicators, these indicators can be used as a reference to find out whether cow milk is good or not. Milk quality levels can be determined through the predictions from the readings in several sensors to read several parameters namely: pH, temperature, taste, odor, fat, turbidity, and colour that have gone through some algorithm. In response to this requirement, this study presents a Machine Learning (ML) approach for coping with the variability of lactation behaviours, and it is demonstrated how this approach contributes to the state of the art. This study aims to build a design of a system capable of predicting the quality of milk using a machine learning approach on datasets using RFECV method as a feature elimination method to increase the cost efficiency of the design as well as a method to increase the effectiveness of detection system. RFECV does not require a predefined number of features. Instead, it works iteratively by removing features and selecting the best subset based on model performance, thereby determining the number of features dynamically. Secondly, RFECV provides a more versatile possibility in evaluating feature significance. It allows the use of various machine learning models, such as linear regression, SVM, Extra-Trees, or Random Forest, leveraging their feature importance scores for evaluation. The iterative process continues until the specified stopping criteria are met. Additionally, RFECV is combined with cross-validation method to further enhance the robustness of feature selection. The studies and research will be carried out by making a tool to test the purity of cow milk. It is hoped that this tool can reduce public and consumer anxiety about consuming cow milk. By using this algorithm in selecting the design of the milk grade quality level detection system, the results obtained in selecting milk grade measurement instruments will be more effective, accurate and efficient.

METHOD

The milk grade instrument is a device used to determine the quality of milk through sensor readings that obtain data on parameters defining milk quality. RFECV has an important role in decreasing the number of sensors employed, as depicted. Where i denotes the number of

sensors and n_i represents each sensor index. In addition to influencing the precision of the ensuing detection process, the RFECV method also significantly impacts power reduction, increasing the efficiency of device power usage as shown in Figure 1.

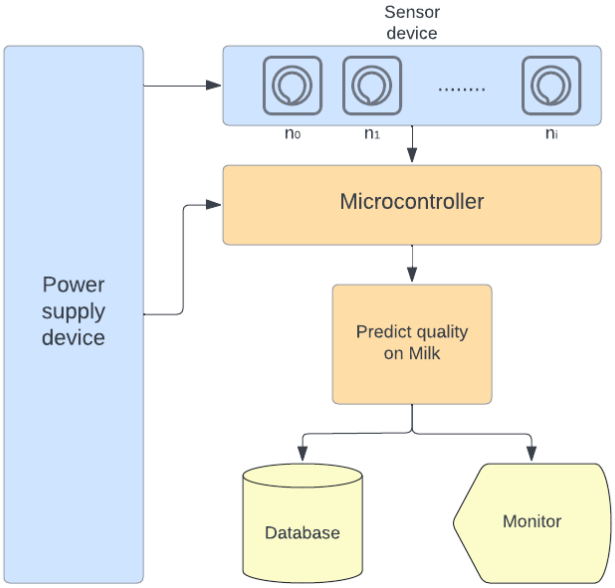

Figure 1. Proposed system

Dataset

The data used in this analysis is sourced from Kaggle, consisting of 3276 total data points. There are 7 parameters shown in [13]. The data is the result of synthesizing data used to measure milk grade. The classification of milk grade is divided into three cluster categories: low with a value of one (0), medium with value (1) and high with a value (2), in reference to milk quality standards. Synthetic data has been used in various machine learning training processes to prevent dataset imbalance and the presence of bias that tends to appear in real-world datasets [14]. Using synthetic datasets as primary training data has reliability in predicting data, actually with an insignificant decrease in accuracy and a low deviation [13-14].

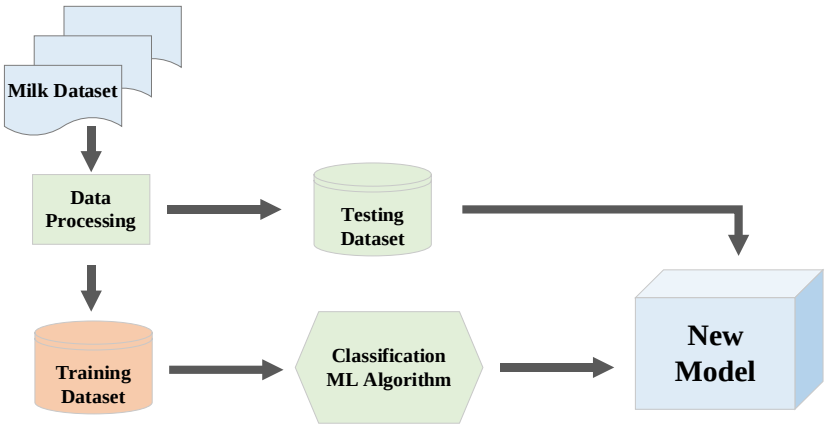

Figure 2. System model and dataset.

Prior to commencing model training, data preprocessing method such as imputation and standardization are applied to enhance the accuracy and performance of the results. Imputation is utilized to populate missing or blank values within the dataset using the mean value for each feature. Upon completion of the imputation stage, standardization then takes place. Standardization processes the data to give features zero mean and unit variance, which helps

optimization algorithms converge faster. These preprocessing steps optimize the data for training machine learning and deep learning models. Figure 2 shows a flowchart of the design determination process, which begins with the pre-processing stage.

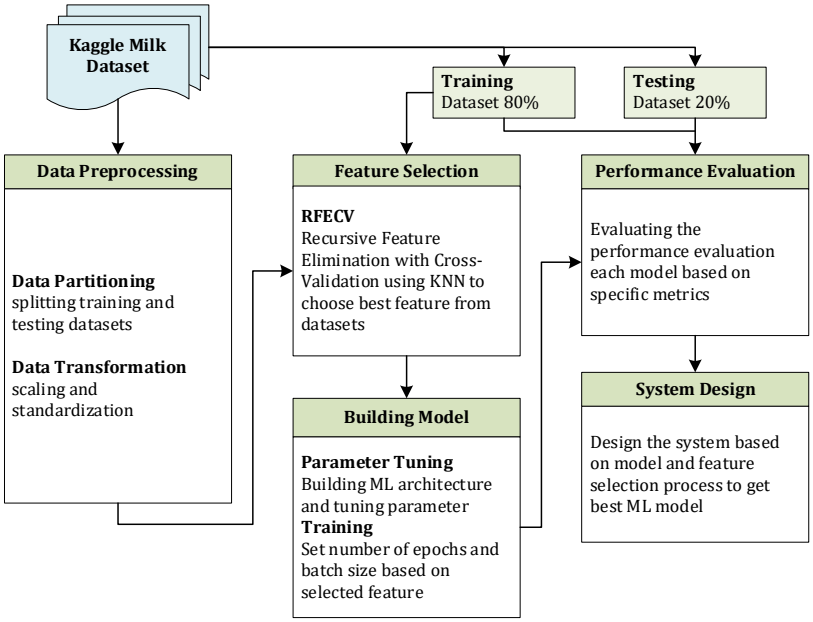

Figure 3. Data pipeline diagram.

Data Standardization

Data standardization is carried out to avoid over scaling the input data features and ensure each feature is on the same scale. The data standardization process is carried out by calculating the mean-$\bar{X}$ and standard deviation s values of each feature using (1) and (2); after the mean and standard deviation values are obtained, the z-score value z also Min-Max scaling normalization could be calculated using (3) and (4).

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \quad (1)$$

$$S = \sqrt{\frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1}} \quad (2)$$

$$z = \frac{X_i - \bar{X}}{s} \quad (3)$$

$$X_{scaled} = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (4)$$

Recursive Feature Elimination with Cross Validation (RFECV)

RFECV is a feature selection algorithm that combines the methods of recursive feature elimination and cross-validation to determine the optimal subset of features that best contribute to the performance of a machine learning model [15]. It starts by training a machine learning

model on the full set of features. Then, it recursively eliminates the least important features and retrains the model. This process is repeated until the optimal subset of features is identified, which maximizes the model performance based on a specified evaluation metric. It is particularly useful for dealing with high-dimensional data or datasets with a large number of

features, as it helps to reduce the dimensionality while still maintaining good predictive performance [16]. This algorithm plays a crucial role in feature selection by identifying the subset of features that are most informative and relevant for a given machine learning model. This helps to improve model performance by reducing overfitting, enhancing interpretability, and potentially increasing computational efficiency. Furthermore, RFECV incorporates cross-validation, which aids in the assessment of how well the selected features generalize to unseen data. It is an algorithm used to identify the optimal predictive features from a dataset to maximize accuracy [17]. By eliminating features that are not needed, RFECV can accelerate computational processing time and reduce resource requirements for model deployment. RFECV is performed through training a model on a dataset using cross validation. It is also a wrapper feature selection method that utilizes machine learning algorithms to identify the most pertinent features for intrusion detection. To ensure stability, RFECV combines recursive feature elimination and cross-validation to determine the ideal number of features that maximize model performance.

KNN is often used as the base RFECV model due to its strong performance on smaller datasets compared to other algorithms such as neural networks. The KNN training process begins by calculating the G_{ini} *impurity value* ($G_{ini_{t,j}}$) of each decision tree for randomly selected features (t) from a subset of the data. This determines which features will be used as root and leaf nodes by computing the probability ($P_{t,j}$) of each categorical target class value for each feature input to the target feature (j) out of the total unique categorical values (K) in the target class. The value $P_{t,j}$ is obtained by calculating the total number of each categorical value in the target class (n (t, j)) for a given categorical feature and dividing by the total number of all categorical target values (Ni) in class i, as shown in equations (5) and (6) [19].

$$P_{t,j} = \frac{n_{t,j}}{N_i} \tag{5}$$

$$G_{ini_{t,j}} = 1 - \sum_{j=1}^{K} P_{t,j}^2 \tag{6}$$

To determine the impact of each feature on the target value, the Gini impurity for each definite feature value is calculated using Equation 6. This results in a weighted sum of the Gini values which can be used to assess which feature plays an important role in defining the root node, child nodes, and leaf nodes of each decision tree. The analysis considers the total number of features (C) and total number of categorical target values in the class (Nt) as shown in equations (7) [17].

$$Weighted(G_{ini_t}) = \frac{1}{N_t} \sum_{i=1}^{C} N_i * G_{ini_{t,i}} \tag{7}$$

Out-of-bag (OOB) values refer to data not used during the decision tree training process. OOB error values are utilized to evaluate model performance and determine feature importance rankings. The OOB error is computed by inputting OOB data into a decision tree not involved in the training subset and comparing the prediction results against actual data. The OOB value is used to calculate the mean decrease in impurity by taking the average of the Gini impurity across each feature in each decision tree. RFECV is performed by specifying the k −value and

step parameters to employ. After selecting the step and k-value, RFECV partitions the data into k-folds with $k - 1$ folds as the test set. The average OOB error is computed across all k-folds.

The OOB error results are then used by grid search to determine optimal hyperparameters and identify how many and which features maximize accuracy [19].

A machine learning approach for designing a milk grade detection system can yield benefits in cost efficiency and predictive ability. One challenge is the need for large, high-quality data from system sensors. The proper sensor data allows for designs that are more flexible and robust under various conditions. System can be built using sensors as shown at Table 1 below.

Table 1. Sensor devices for each input parameter

Parameter	Sensor	References
pH	PH4502C, a separate electrode Ag/AgCl and KCl solution, with an accuracy up to 99.10%.	[24]
temperature	DS18B20 is a 1-wire programmable Temperature sensor from maxim integrated. It is used to measure temperature in hard environments like in chemical solutions, mines or soil etc. It can measure a wide range of temperature from -55°C to +125° with a decent accuracy of ±5°C.	[25]
taste	Taste-sensing TS-5000Z is a taste-sensing receptor membrane composed of a lipid, plasticizer, and polyvinyl chloride to detect taste substance.	[26]
odor	Smart Odor Sensor DT322 to measure elements: amine series, sulphur gas temperature measurement value that measured from -30~+85℃ (±0.3℃) with humidity measurement: 0-90%(±2%RH) and output level: 0-30PPM.	[27]
fat	Highly sensitive fiber optic milk fat sensor using U-bent plastic optical fiber (POF) probes based on the refractive index (RI) of milk, an inherent physicochemical property of milk, which is significantly influenced by the milk fat content. It has sensitivity of 0.15 ΔA/Δ% fat for the variation of fat content in the milk.	[28]
turbidity	Water quality sensor that utilizes an RS485 communication interface and the standard Modbus protocol. It uses an infrared LED with a wavelength of 860nm as the light source, which is not affected by the colour of the water sample with output High/Low level signal.	[**Error! Reference source not found.**]
color	TCS3200 is a colour sensor which can detect any number of colours with right programming. TCS3200 contains RGB (Red Green Blue) arrays. As shown in figure on microscopic level one can see the square boxes inside the eye on sensor.	[29]

Evaluation

The proposed performance model was compared against previous models developed using the same dataset based on evaluation metrics to determine the proposed model performance, as shown in Figure 4. This comparison revealed whether the use of RFECV provided advantages over prior models and explained why prior models achieved higher performance results than earlier models.

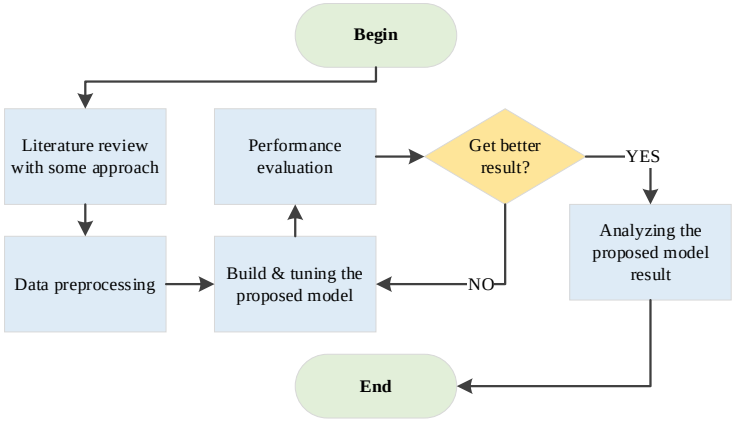

Figure 4. Evaluation diagram process

Model evaluation is a critical component of the testing process. The goal is to determine the model performance based on specific metrics. Commonly used metrics to measure performance include accuracy, precision, recall, and $F1$ score is commonly used [17]. These five metrics are calculated by comparing the predicted label to the actual known label. A true positive (TP) occurs when the predicted label matches the positive known label. A false positive (FP) is when the predicted label differs from the positive known label. A true negative (TN)occurs when the predicted label matches the negative known label. A false negative (FN) is when the predicted label differs from the negative known label [20].

$$precission = \frac{TP}{TP+FP} \tag{8}$$

$$accuracy = \frac{TP+TN}{TP+FP+TN+FN} \tag{9}$$

$$recall\ \ (TPR) = \frac{TP}{TP+FN} \tag{10}$$

$$F1 - \text{score} = \frac{2\,x\,precission\,x\,recall}{precission+\,recall} \tag{11}$$

Then AUC-ROC refers to the area under the receiver operating characteristic curve. It is basically a measurement of how well a model can distinguish between different classes. The closer the AUC-ROC is to 1, the better the model is at telling apart true positives from false positives. It is a common metric used in machine learning to evaluate binary classification models. The ROC curve plots the true positive rate against the false positive rate for a model using different classification thresholds. This indicates the model performance. The AUC is obtained by calculating the entire area under the ROC curve, which ranges from initial graph $(0,0)$ to $(1,1)$ on the x and y axes. The AUC provides a single measure of a model ability to discriminate between classes, aggregating performance over all possible classification

thresholds. It is a useful metric for comparing models as it summarizes a model performance in a single number between 0 and 1 [21].

$$sensitivity\ (FPR) = \frac{TP}{TP+FN} \quad (12)$$

RESULTS AND DISCUSSION

Exploratory Data Analysis

The collected data consists of seven numerical input variables and one categorical target variable for milk grade, totaling 1,059 data points. Data preprocessing identified empty values and found missing data in the pH, temperature, taste, odor, fat, turbidity, color features. A deletion process was carried out to fill in the blank data by imputing the missing values with the average value for each feature.

High accuracy cannot be achieved without correlation between parameters. Figure 5 shows the correlation between parameters and other parameters. It indicates strong correlation between pH and taste levels, pH and temperature, turbidity and temperature, and fat and taste, as visualized through heatmap graphics. The existence of correlation for each of these parameters will create patterns that can be utilized by machine learning method to determine the quality level of milk. The correlation between parameters as depicted in the heatmap and the aforementioned research illustrates that there are related factors which determine the level of milk quality.

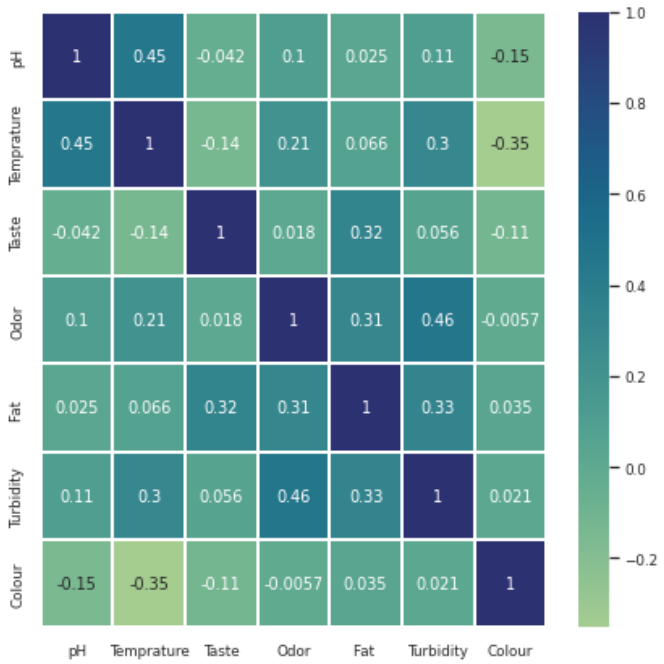

Figure 5. Correlations between each feature in predicting milk grade

A correlation matrix is a simple way to summarize the correlations between all variables in a dataset. The high accuracy cannot be separated from the correlation between parameters. Fig. 5 shows a strong correlation between the levels of pH and temperature score are 0.45 which indicates that they're strongly positively correlated, temperature and turbidity, and taste and fat. pH and turbidity levels score are 0.11 that they're fairly corelated, which are visualized through heatmap graphics. It also shows that the correlation between temperature and colour are 0.35, which indicates that they're weakly negatively correlated. The existence of a

correlation for each of these parameters will create a pattern that can be used by machine learning to determine the quality of the milk.

Some experiments were conducted with various k value, on this research from k=1 until k=7, then the optimum k=3 was chosen. The results obtained in the RFECV process by setting a value of k = 3 and a weight of step = 1 are shown in Fig. 6. Based on the data shown, the input color feature has the highest influence in determining the water quality and potability level of a milk, and the pH feature has the lowest feature in knowing the level of potability.

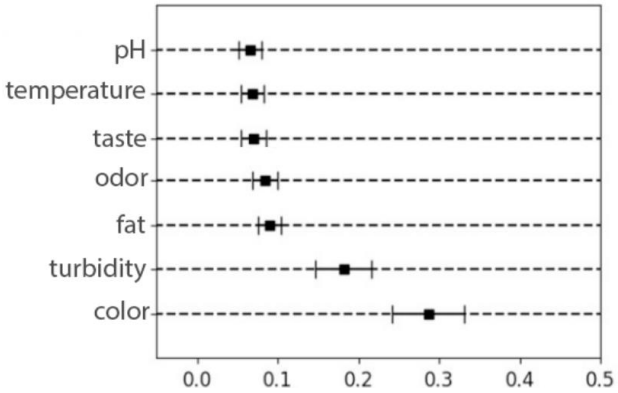

Figure 6. RFECV mean decrease impurity for each feature.

Training Stage

In the classification training stage using the K-Nearest Neighbour method, the k value approach has been carried out with k = 1,3,5,7. The best k-value approach is used to consider more accurate classification results. Predictions from training data are based on obtaining the highest accuracy by using the value of k in the K-Nearest Neighbour classification. In training for three classes with angle variations and values of k. The highest accuracy was obtained at k=3. However, this special case shows that the classification is predicted based on the nearest neighbour only, in other words, using the Nearest Neighbour algorithm itself [31]. The most optimal results from training for three classes were obtained at milk sampling time in pass 12 hours, pass 24 hours and pass 36 hours with a value of k = 3 with the acquisition of accuracy, sensitivity, and specificity of 96%, 94%, and 95%. The results of this training were used as a trial on 90 test data sets with 30 data points at each 12-hour, 24-hour, and 36-hour interval from fresh milk collection. The results of the comparison of time test with experimental values of k = 1,3,5,7 obtained from training for three classes can be seen in Table 2 below.

Table 2. Comparison of angle variations and k values for 3 class training

Time	Comparison time sampling variation and k-value											
	Accuracy (%)				Sensitivity (%)				Specificity (%)			
	1	3	5	7	1	3	5	7	1	3	5	7
12-hour	98	93	85	78	99	93	85	79	99	93	85	84
24-hour	99	96	85	81	98	94	86	88	99	95	86	82
36-hours	99	94	83	80	98	94	84	86	98	94	84	84

Testing Stage

The testing process for test data in the three-class classification of 150 sets of data points entered into the analysing program resulted in accuracy of 84.60%, sensitivity of 82.30%, and specificity of 80.50%. This process used a matrix of numbers that tell us where a model gets confused. It is a class-wise distribution of the predictive performance of a classification model. That is, the confusion matrix is an organized way of mapping the predictions to the original classes to which the data belong. The results obtained from RFECV process by setting a value of k = 3 and a step weight of 1. The results of the confusion matrix for three grade classes can be seen in Table below. Table 3 shows give a lot of information about the performance model. Adding the numbers in the first column, we see that the total samples in the low grade are 47. Similarly, adding the numbers in the second and third column gives us the number of samples in the medium grade and high grade, which are 53 and 50 in this case. The sum of the numbers in all the boxes gives the total number of samples evaluated. Further, the correct classifications are the diagonal elements of the matrix, namely: 41 for low grade, 44 for medium grade and 42 for high grade. Then, 2 sample M-low-medium that were expected were classified as the negative class by the model, because the model predicted negative, which was wrong same as M-low-high that contains value 4. Similarly, the others matrix (top-right box) was expected to be of medium and high grade but were classified as low and high grade. As usual, the diagonal elements are the correctly predicted samples. Amount of 150 samples were correctly predicted out of the total 127 samples.

Table 3. Confusion Matrix for three grade milk in quality

Confusion Matrix		**Expected Grade**		
		low	medium	high
Predicted Grade	low	41	4	4
	medium	2	44	4
	high	4	5	42

Model Effectiveness

Based on information from RFECV, training was carried out using five features with the other algorithm. The results obtained from training are shown in Figure 7. Through grid search on RFECV results [18]-[19], the highest accuracy was obtained when training used five features: pH, temperature, taste, odor, fat, turbidity, color, as shown in Figure 6 using equations (8) up to (12). Proposed method (KNN combined with RFECV) had a higher area for data testing under the curve (AUC) and accuracy result of 7,20% higher than the support vector machine (SVM) model with considered parametric tuning for C = 1.0, kernel = RBF, degree = 3, and gamma = auto, 15,37% higher than single K-Nearest Neighbours (KNN) with same previous value k, and 16.37% higher than the random forest (RF) model with hyperparameter tuning for n-estimator = 300, max depth = 8, max features = auto and with no minimum sample number.

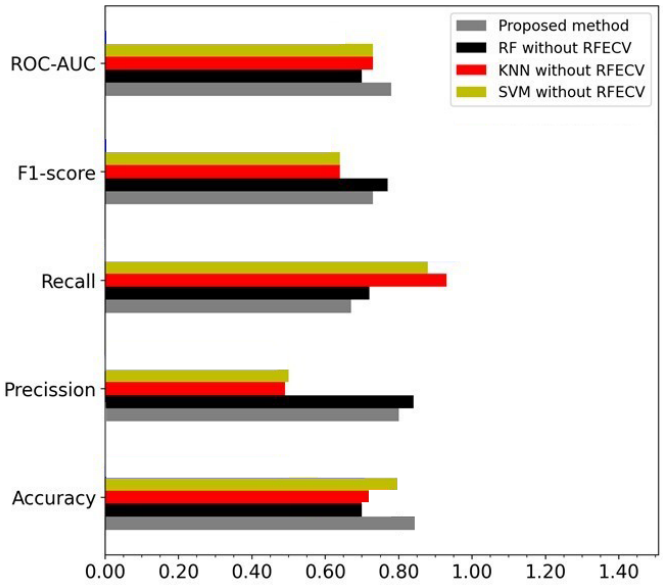

Figure 7. Comparison of proposed method results toward other ML approaches.

Models with high accuracy and AUC but lower precision, recall, and F1-score indicate that the model can accurately predict negative classes but struggles with positive classes [20]-[21]. There are several potential causes for this performance. An imbalanced dataset with respect to class distributions can impact predictability. A model that is overly complex, such as too deep, may lead to lack of generalization on test data. Data quality issues including noise and outliers can negatively influence evaluation metrics [22]-[22]. With RFECV process to set value k = 3 and step = 1, the input colour feature has the highest influence in determining the water quality and potability level of a milk, and the pH feature has the lowest feature in knowing the level of potability quality. And by some classification training stage using the K-Nearest Neighbour method, the k value approach has been carried out with k = 1,3,5,7. The most optimal results from training for three grade classes were obtained with a value of k = 3 with the acquisition of accuracy of 96% and training for three grade classes, namely sampling time in pass 12 hours, pass 24 hours and pass 36 hours while the testing accuracy for testing data obtained for three grade classes is 84.60%.

RFECV model demonstrates higher accuracy than a model trained on the full dataset for detection tasks. Reducing features in the dataset also decreases computation time, costs, and resource requirements. Therefore, in addition to enhancing effectiveness, the RFECV method has been shown to improve cost efficiency of the design. The potentiality level milk quality measurement approach uses machine learning methods such as evidenced by the results presented in Table 4, the proposed program has demonstrated the ability to distinguish between three categorical classes based on seven input parameters. The inclusion of additional features has validated the propensity to augment predictive accuracy when evaluating milk quality. Some other models e.g., decision trees not able to achieve higher accuracy in similar case, because trees are prone to overfit on the training data, if their growth is not restricted in some way, unstable to changes in the data and unstable to noise, non-continuous and complex calculations on larger datasets. While random forest can be less interpretable than a single decision tree because it involves multiple trees and it can be difficult to understand how the algorithm arrived at a particular prediction.

Table 4. Performance comparison with some of the existing approaches

No	Literature	Data	Method	Accuracy
1	Slob [7]	56	Decision tree (DT)	70.50%
			Artificial neural network (MLP-ANN)	83.55%
2	Frizzarin [8]	622	Random forests (RF)	70.55%
			Boosting decision trees (BDT)	78.85%
			Support vector machines (SVM)	76.95%
3	Mu [9]	105	Support vector machines (SVM)	75.15%
			Random forests (RF)	73.35%
4	Jiménez [10]	176	Artificial Neural Network (ANN)	78.00%
			Support vector machines (SVM)	74.65%
			Random forests (RF)	80.65%
5	This research	1059	Random forests (RF)	70.55%
			K-Nearest Neighbours (KNN)	72.00%
			Support vector machines (SVM)	79.50%
			Proposed model: (KNN with RFECV)	**84.60%**

CONCLUSION

Based on this research, the utilizing RFECV in the design selection process has a significant impact on system effectiveness and it can reduce many parameters to seven parameters. The most optimal results from training data for three classes were obtained at each data with a value of k = 3 with the acquisition of accuracy, sensitivity, and specificity of 96%, 94%, and 95% on getting data in 24-hours. Using proposed model (KNN with the RFECV) algorithm has a better accuracy result of 5.10% higher than the support vector machine (SVM) model, 12.60% higher than single K-Nearest Neighbours (KNN) and 14.05% higher than the random forest (RF) model, with values of accuracy, precision, recall, F1-score, and area under the curve (AUC) of 0.84, 0.80, 0.68, 0.75 and 0.81 respectively for testing data. The milk quality prediction design system grade is based on feature elimination results. It is also informed by a literature review on device selection. The application needs some input milk parameters from user value or farmers then can be implemented by a software or tools on a single computer using programming language or scripts. This model could be used in micro industrial that only need some appropriate data and need to be classified briefly.

REFERENCES

1 Vidhya, S., Siva Vadivu Ragavi, V., Monica, J.K., Kanisha, B... 2023. Milk Quality Prediction Using Supervised Machine Learning Technique. In: Rathore, V.S., Piuri, V., Babo, R., Ferreira, M.C. (eds) Emerging Trends in Expert Applications and Security. ICETEAS 2023. Lecture Notes in Networks and Systems, vol 682. Springer, Singapore.

2 Muchtadi. 1992. Petunjuk Laboratorium Ilmu Pengetahuan Bahan Pangan. Departemen Pendidikan dan Kebudayaan Direktorat Jenderal Pendidikan Tinggi Pusat Antar Universitas Pangan dan Gizi Institut Pertanian Bogor. 34-35.

3 Chandra, B. 2007. Pengantar Kesehatan Lingkungan. Jakarta, Penerbit Buku Kedokteran EGC
4 Wilshaw, GA, T. Cheasty, dan HR Smith, 2000. Escherichia coli. In: Lund, BM, Baird Parker, TC, Gould, GW (Eds.), The Microbiological Safety and Quality of Food II. Aspen Publishers Inc., Gaithersburg, Maryland, j.pp. 1136-1177. accessed in august 24, 2024
5 Brettschneider, K.C., Zettel, V., Vasafi, P.S. *et al.* 2022. Correction to: Spectroscopic-Based Prediction of Milk Foam Properties for Barista Applications. *Food Bioprocess Technol* 15, 1758–1759.
6 Tosca, M. A., Olcese, R., Trincianti, C., Naso, M., Schiavetti, I., & Ciprandi, G. 2024. Children with cow's milk allergy: prediction of oral immunotherapy response in clinical practice. *Allergo Journal International*, *33*(5), 168-169.
7 Slob, N., Catal, C., & Kassahun, A. 2021. Application of machine learning to improve dairy farm management: A systematic literature review. *Preventive Veterinary Medicine*, *187*, 105237.
8 Frizzarin, M., Gormley, I. C., Berry, D. P., Murphy, T. B., Casa, A., Lynch, A., & McParland, S. 2021. Predicting cow milk quality traits from routinely available milk spectra using statistical machine learning methods. *Journal of Dairy Science*, *104*(7), 7438–7447
9 Mu, F., Gu, Y., Zhang, J., & Zhang, L. 2020. Milk Source Identification and Milk Quality Estimation Using an Electronic Nose and Machine Learning Techniques. *Sensors*, 20(15), 4238
10 Jiménez-Carvelo, A. M., Sanae Bikrani, Mounir Nechar, Badredine Souhail, & Cuadros-Rodríguez, L. 2022. Machine learning–based chemometric methods for quality and authentication of milk and dairy products. *Elsevier EBooks*, *1*(4), 261–280.
11 Fuentes, S., Gonzalez Viejo, C., Cullen, B., Tongson, E., Chauhan, S. S., & Dunshea, F. R. 2020. Artificial Intelligence Applied to a Robotic Dairy Farm to Model Milk Productivity and Quality based on Cow Data and Daily Environmental Parameters. *Sensors (Basel, Switzerland)*, *20*(10).
12 Sugiono, S., Soenoko, R., & Riawati, L. 2017. Investigating the Impact of Physiological Aspect on Cow Milk Production Using Artificial Intelligence. *International Review of Mechanical Engineering-IREME*, *11*, 30-36.
13 Kaggle. Milk Dataset. 2024. Online: https://storage.googleapis.com/kagglesdsdata/datasets/2380569/4016630/milknew.csv
14 Jacobsen, B. N. 2023. Machine learning and the politics of synthetic data. *Big Data & Society*, 10(1).
15 Rankin, D., Black, M., Bond, R., Wallace, J., Mulvenna, M., & Epelde, G. 2020. Reliability of Supervised Machine Learning Using Synthetic Data in Health Care: Model to Preserve Privacy for Data Sharing. *JMIR medical informatics*, *8*(7), e18910.
16 Baressi Šegota, S., Anđelić, N., Šercer, M., & Meštrić, H. 2022. Dynamics Modeling of Industrial Robotic Manipulators: A Machine Learning Approach Based on Synthetic Data. *Mathematic*s, 10(7), 1174.
17 Mowri, Rawshan Ara, Siddula, M., & Roy, K. 2022. A Comparative Performance Analysis of Explainable Machine Learning Models with And Without RFECV Feature Selection Technique Towards Ransomware Classification. *ArXiv (Cornell University)*, 4(3).
18 Jadhav, A., Pramod, D., & Ramanathan, K. 2019. Comparison of Performance of Data Imputation Methods for Numeric Dataset. *Applied Artificial Intelligence*, 33(10), 913–933.
19 Huang, E.-H., Hu, H.-W., Jheng, W.-L., Chen, K.-Y., Liu, C.-H., Chi, H.-Y., Chang, T.-W., Chin-Yu, W., Un, C.-H., Lin, H.-M., Chen, C.-W., & Wang, J.-F. 2021. Feature Selection for Intradialytic Blood Pressure Value Prediction Using GRU-based Method Under RFECV algorithm. IEEE Xplore.
20 Bradley, A. P. 1997. The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognition*, 30(7), 1145–1159.
21 Erickson, B. J., & Kitamura, F. 2021). Magician's Corner: 9. Performance Metrics for Machine Learning Models. Radiology: *Artificial Intelligence*, 3(3), e200126.
22 Ivković, J., & Jelena Lužija Ivković. 2017. Analysis of the Performance of the New Generation of 32-bit Microcontrollers for IoT and Big Data Application. 2, 4(1), 330–336.
23 Goswami, S. 2021. Arduino-Based Milk Quality Monitoring System. *International Journal of Agriculture Environment and Biotechnology*, *14*(2).

24 Kokoulin, A. N., Tur, A. I., Yuzhakov, A. A., & Knyazev, A. I. 2019. Hierarchical Convolutional Neural Network Architecture in Distributed Facial Recognition System.
25 Permana, A. N., Wibawa, I. M. S., & Putra, I. K. 2021. DS18B20 sensor calibration compared with fluke hart scientific standard sensor. *International Journal of Physics & Mathematics*, *4*(1), 1–7.
26 Wang, J., Sakai, K., & Toshihiko Kiwa. 2023. All-in-One terahertz taste sensor: integrated electronic and bioelectronic tongues. *Sensors & Diagnostics*.
27 Natnaree Phukkaphan, Tanthip Eamsa-ard, Chalisa Chairanit, & Teerakiat Kerdcharoen. 2021. *The Application of Gas Sensor Array based Electronic Nose for Milk Spoilage Detection*.
28 Gowri, A., Rajamani, A. S., Ramakrishna, B., & Sai, V. V. R. 2019. U-bent plastic optical fiber probes as refractive index-based fat sensor for milk quality monitoring. *Optical Fiber Technology*, *47*, 15–20.
29 Hardiana, T. O., Hidayati, R. N., Anggoro, W., & Irawati, N. 2019. Detection of water turbidity using LDR sensor. I*n Third International Seminar on Photonics, Optics, and Its Applications (ISPhOA 2018,* Vol. 11044, pp. 66-69. SPIE.
30 Ziyaina, M., Rasco, B., Coffey, T., Ünlü, G., & Sablani, S. S. 2019. Colorimetric detection of volatile organic compounds for shelf-life monitoring of milk. *Food Control*, *100*, 220–226.
31 Lubis, Z. 2019. Optimasi Nilai K pada Algoritma K-NN dalam Clustering Menggunakan Algoritma Expectation Maximization. (Doctoral dissertation, Universitas Sumatera Utara)