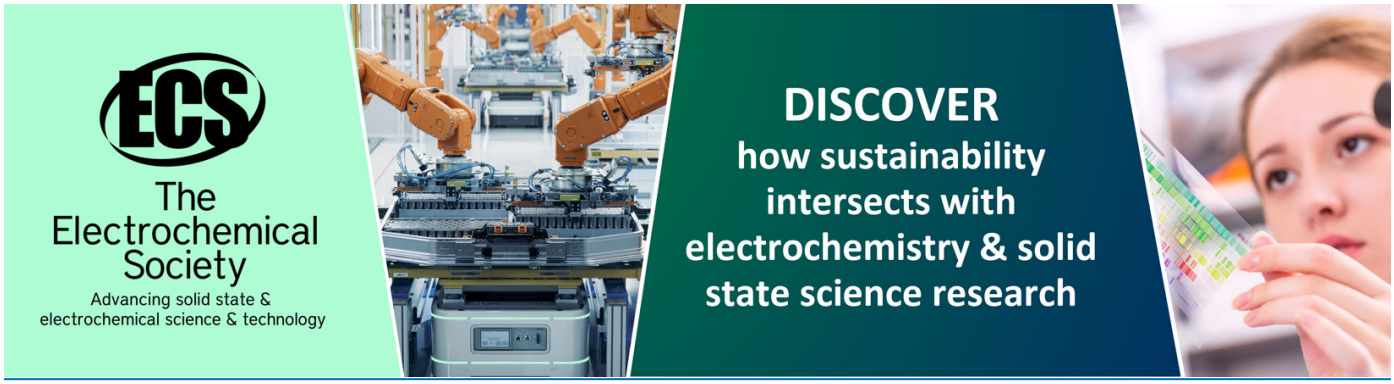
An analysis and design of fresh milk smart grading system based on internet of things



An analysis and design of fresh milk smart grading system based on internet of things

W Habsari*, F Udin and Y Arkeman

Department of Agroindustrial Technology, IPB University, Bogor, Indonesia 16680

E-mail: windihabsari@gmail.com

Abstract. The grading of fresh milk affects the quality classification in the dairy industry. This study aims to analyze and design a smart grading system using machine learning models to classify the grade of fresh milk. Business process analysis helped understand the capturing steps as the main elements, such as the smart grading system. The result of the requirement analysis showed how smart the grading system involved stakeholders. The machine learning model can help the Internet of Things system classify goods or services. *Artificial Neural Network* and K-means were designed to classify and group indicators of fresh milk quality. The variables used in this study consisted of pH, temperature, odour, turbidity, colour, fat, and taste values. The data were taken from the upstream dairy industry SAE Pujon. The classification result of fresh milk grades using ANN consisted of three low, medium, and high grades. The accuracy value of the classification obtained is 98.74%. The attributes used for grouping were temperature and colour. The best clusterization that used K-Means is the third cluster. Based on the data analysis, the smart grading system made users save time knowing the grade of fresh milk easier.

1. Introduction

Dairy cows produce fresh milk. Fresh milk is milk from dairy cattle whose natural content is not added or reduced and obtained using clean and correct milking [10]. Milk is the main product in a dairy farm that becomes a source of income for farmers in maintaining the sustainability of their livestock business. The requirement of fresh milk that can be consumed successfully goes through a causal test following SNI standards. The benefits of fresh milk can prevent various diseases and help relieve brain performance [13]. One of the largest dairy cattle businesses in East Java is located in Pujon Subdistrict, which is included in the Malang regency. According to the Central Statistics Agency (BPS), in 2017, milk production precisely in East Java Province reached 52,028,000, and the milk consumption of Indonesians amounted to 16.5 litres/capita/year. Fresh milk is a perishable ingredient [14]. Therefore, it is necessary to handle complexes ranging from the internal environment of cows to the cooperative environment. The problem in Pujon's "SAE" dairy cooperative is the lack of a new milk grading system to determine the quality of each farmer's fresh milk. A grading system can improve the quality of a product [10].

The grading system aims to classify the grade of fresh milk. A previous study by [2] used the multi-label decision tree to classify coconut oil quality using the classification of metagenome fragments with kmers as feature extraction [1]. The milk quality test's classification of fresh milk quality and selling price is determined by the milk quality test [10]. The important thing in testing the quality of fresh milk

* To whom any correspondence should be addressed.

using technology is the help of the internet in real-time. Internet of things, big data, and artificial intelligence are the latest technology trends [6].

IoT systems are widely applied to health, retail, logistics, smart transportation, smart homes, environmental monitoring, and agriculture [9]. A highly integrated intelligent Cyber-physical space opens the door to creating new businesses and market opportunities for manufacturing [16]. Monitoring systems improve effectiveness in a cold environment [14], while the IoT develops an intelligent environment, saving time, energy, and cost [4]. Any IoT-based object can be a smart object. Machine learning (ML) is a technique that can effectively recognize patterns [12]. This pattern analyzes the sample as the basis for an introduction to classification and identification. The ML method is one of the superior methods on the Internet of Things [14]. ML method can help IoT systems classify goods/services [6]. ML algorithms are used for K-Nears Neighbour, Support Vector Machine, Artificial Neural Network, and Convolutional Neural Network [8]. An artificial Neural Network (ANN) is a trained computer network similar to the human brain. Another example of machine learning algorithms is K-Means. The primary purpose of the K-Means algorithm is to allow methods in partitioning N dimensions into K sets that produce partitions that are quite effective in-class variance [16].

The requirement of fresh milk that can be consumed successfully goes through a causal test following SNI standards. Smart grading uses sensor tools to improve ease of activity in determining the quality of fresh milk. One of the advantages of a smart grading system is the information system to determine the grade of fresh milk per farmer, which helps stakeholders determine the purchase price of fresh milk. This information system uses machine learning algorithms to predict complex data input [5]. The internet of things helps facilitate the test of the quality of fresh milk. Therefore this research analyzes the system's needs and designs a smart grading system using machine learning models to classify the grade of fresh milk.

2. Methodology

2.1 Framework

The framework of this study is illustrated in Figure 1. The first step is to design an IoT-based fresh milk smart grading system, the second step is to predict the value and find the accuracy of the new milk grade prediction, and the third step is to determine the cluster.

Monitoring the temperature of fresh milk in temporary storage uses a DHT22 sensor. The data are collected using the internet, Arduino, and cloud computing. The design of an IoT system will help users monitor temperature and maintain each environment. It will help to determine the grade of fresh milk.

2.2 System Requirements

This study began by identifying problems and challenges that have been done. The stakeholders' requirements consisted of resources, opportunities, inputs, processes, and outputs. At the analysis stage of the system needs, all stakeholders were identified. The modelling systems used Business Process Modeling Notation (BPMN) and *Use Case Diagram.* BPMN was a model used to achieve a comprehensive system workflow and define system constraints graphically [7].

The primary purpose of BPMN was to provide a notation that was easy to understand by all people, especially software activists. A preliminary draft of the processes up to technical development covering the flow and work in models or notations was created from the existing business analysis. A business process is a set of operations with a start and endpoint and clearly defined inputs and outputs that focus on how an organization does a task [22].

2.3 Classifying Using Artificial Neural Network

The study used an artificial neural network to classify the grade of fresh milk. ANN was presented as an interconnected system of "neurons" from message exchange [11]. The value of fresh milk is determined based on input variables interpolated using artificial neural network algorithms.

ANN was used to predict based on the magnitude of non-linear input values. There were three stages to determine ANN value [11]:

1. Interconnection patterns between various neurons
2. Determine the interconnection weight
3. Activation function converts input weight value to output

According to [7] types of function composition used nonlinear weighting in equations (1)

$$f(x) = K\left(\sum\sum_{i=1} wi\ gi(x)\right) \quad (1)$$

F(x) is a neuron network function, k is an activation function, wi is a weight value, and g is a vector value. The software used to develop the ANN model was R Studio. The ANN model was developed through the design stage of the relationship formula between layers and neurons, compiling data sets for data tests and training tests. Then, calculate the model to produce accuracy. ANN uses multi-layer 7-5-3, with seven inputs, five hidden layers, and three outputs in the form of fresh milk grade. The training process uses 30% test data and 70% training data.

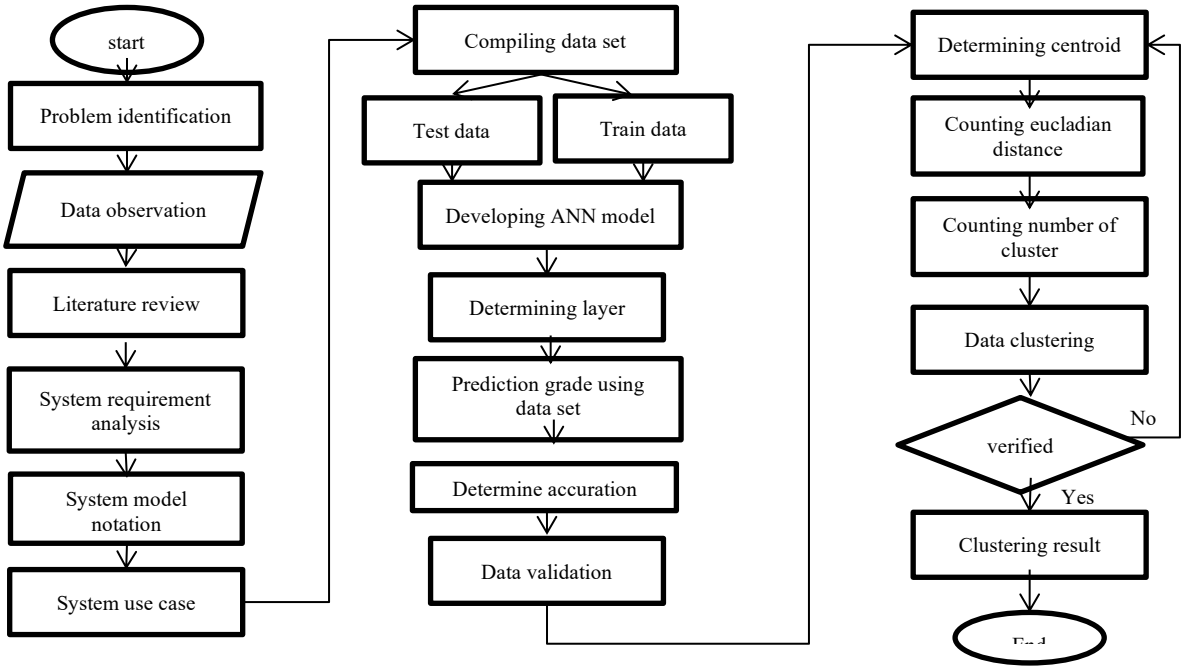


Figure 1. Study flow chart.

Table 1. Value of data attribute to prediction model (Modified from Kaggle and obsevation).

No	Indicator	Min	Max	Target
1	Temperature (°C)	10	37	Max
2	pH	6.3	6.8	Max
3	Turbidity	0	1	Min
4	Colour	0	1	Max
5	Odor	0	1	Min
6	Taste	0	1	Max
7	Fat (%)	0	3	Max

Data collection is taken from the literature and observation in SAE Pujon Malang. Then the data will be used for learning and testing models that will be developed (artificial neural network model). The data is preprocessed first before being used in the design phase of the model. Preprocessing is needed because the neural network can only accept numeric data input, so that the categorical attribute will be used as a dummy variable (unary encoding/numerical binary variable).
The formula used can be described as follows:
- Indicators used in this study are based on literature data. Data consists of 7 parameters (*pH, temperature, turbidity, colour, fat, odour, taste*) as independent variables (inputs), Between variables marked "+" to separate variable names.
- Target value (output) = grade of fresh milk. The output value is divided into three types, i.e., "2" indicates high grade, "1" indicates medium grade, "0" indicates low grade.

2.4 Clustering Using K-Means
In the K-Means method, the data were grouped. Each group has characteristics similar to others. This method minimizes differences between data in one cluster and maximizes differences with other clusters [16]. Grouping using K-Means can be seen in Figure 3. The initial centroid is performed randomly from multiple objects available in as many as k clusters. To calculate the centroid of the next i-cluster, use the following formula:

$$v = \frac{\sum_{i=1}^{n} \text{xi}}{\text{n}} \; i = 1,2,3,..n \qquad (2)$$

Where: v = centroid on cluster
xi: i-th object
Yi: i-i attribute data
n: number of objects or number of objects that are cluster members
Then calculate the distance between the object and centroid, in this study using Euclidian Distance.

$$d(x,y) = |x - y| = \sqrt{\sum_{i=1}^{n} (\text{xi} - yi)} \quad ^2 \; i = 1,2,3,..n \qquad (3)$$

3. Result and Discussions

3.1 Monitoring Based on Internet of Things
A cooling system is required to maintain the required temperature of fresh milk. A sensor is required to assemble with cold storage that requires temperature monitoring. DHT22 sensor monitors the temperature of fresh milk in temporary storage. In such cases, the sensor will provide an acceptable temperature reading. A digital signature has been used to enhance the property of authentication, integrity, and non-repudiation. End-users can easily identify any changes in the data. The design of Internet of Things architecture is illustrated in figure 2.

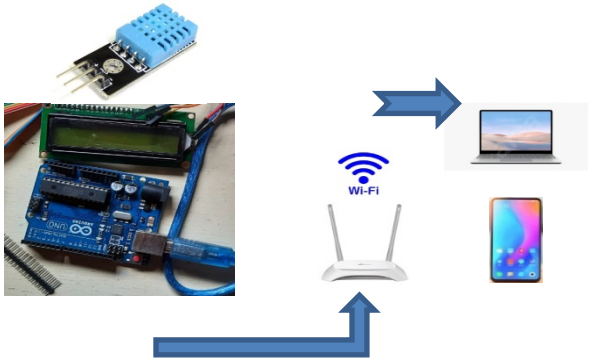


Figure 2. The connection on each device.

Environmental condition is a significant factor that needs to be controlled in fresh milk storage. The device can develop an environmental control smart grading system that will increase the checking of fresh milk. It is done by taking the environmental data and storing it using the internet. The data collected is temperature. It has helped the user monitor the temperature of milk during temporary storage and determine the grade of fresh milk. The control system inside the device is automatically triggered if the environmental conditions are not in optimum condition [6]. However, this system requires a stable internet connection to ensure that the data is sent to the internet. IoT from a supply chain perspective as*"Internet of Things* is a network of digitally connected physical objects to sense, monitor and interact within the company and between companies and the supply chain that enables tracking and sharing of information sources to facilitate the planning, control and coordination of timely supply chain processes". IoT can realize the integration of a wide range of manufacturing devices equipped with sensing, identification, processing, communication, actuation, and network capabilities. A highly integrated intelligent Cyber-physical space opens the door to creating new businesses and market opportunities for manufacturing [15], and the goal of IoT is to develop a smart environment, saving time, energy and cost [4]. IoT systems are widely applied to health, retail and logistics industries, smart transportation, smart homes, environmental monitoring and agriculture [9]. For example, as an electrical power monitoring system [11]. An example of previous research is monitoring potato yields [15] using a cargo monitoring system to improve effectiveness in the cold environment [14].

The machine learning method can help IoT systems classify goods/services [7]. This system has successfully implemented IoT and automated control in precision agriculture. This study is in accordance with [8] stating that The process of Internet of Things intrusion detection mainly includes data collection, data analysis and alarm response. The efficiency and performance of the Internet of Things intrusion detection system have also been significantly improved. Facing the problems of network complication and attack diversification, the research direction of scientific researchers is also developing towards intellectualization and distribution [22]. Machine learning (ML) is a technique that can effectively recognize patterns [16]. This pattern analyzes the sample as the basis for an introduction to classification and identification. The ML method is one of the superior methods on the Internet of Things [14]. An artificial Neural Network (ANN) is a trained computer network similar to the human brain. ANN is capable of handling large amounts of non-linear data interpolated. Another definition of an Artificial Neural Network is the paradigm of information processing inspired by the biological nervous system, such as the process of information in the human brain [13].

3.2 System Requirement of Smart Grading

At this stage, the system is designed to predict the grade of fresh milk. Analysis of the needs of IoT-based fresh milk grading systems was presented in figure 3. The use case diagram describes the functional requirements, as shown in figure 4. The BPMN diagram illustrates the activities of stakeholders, as shown in figure 5 and figure 6.

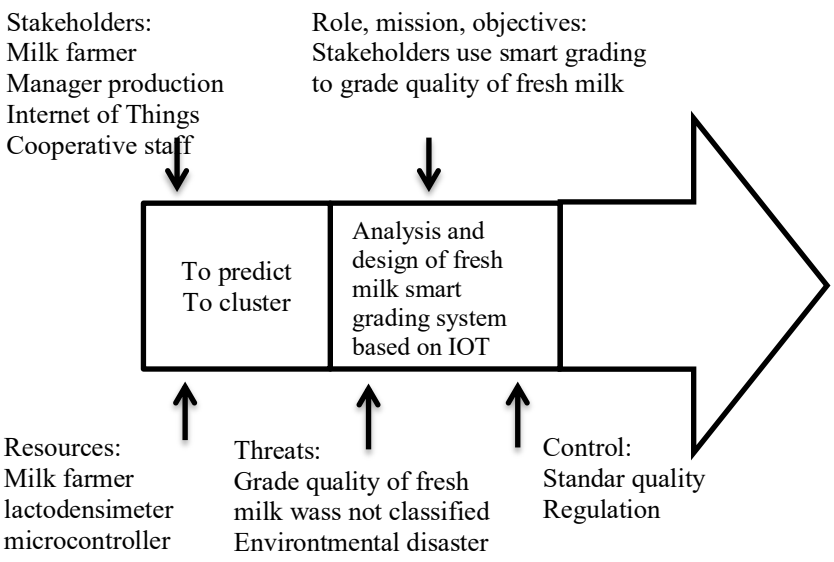


Figure 3. Analysis of fresh milk grading system.

System analysis is interpreted as an individual or organization that applies analytical methods and techniques (scientific, mathematical, statistical, financial, political, social, cultural, et cetera) to provide meaningful data to support informed decision making by missions planners, system operators, and system maintainers [15].

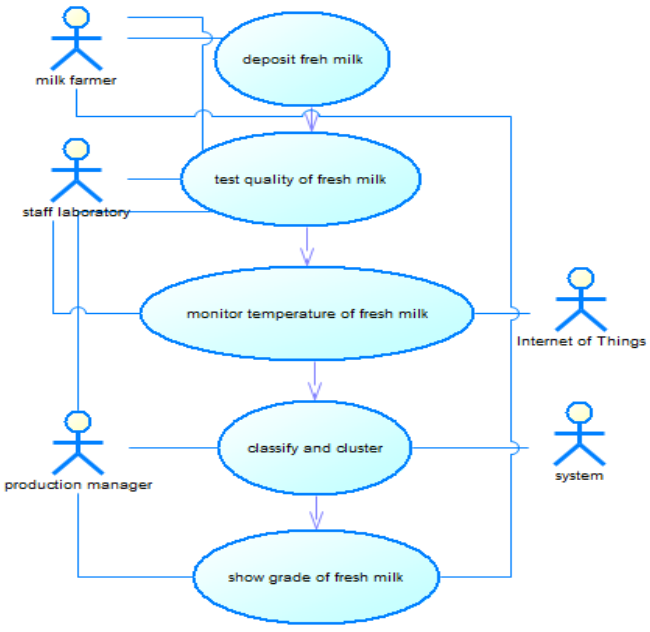


Figure 4. Use case diagram.

Unified Modelling Language can analyze and design a smart grading system appropriately. This modelling provides precise information on each system used. In this stage, the system will be designed to illustrate functional requirements in the classification smart grading systems. This study agrees with [7] that Unified Modelling Language (UML) is a system development method consisting of three stages: problem identification, system analysis, and system design.

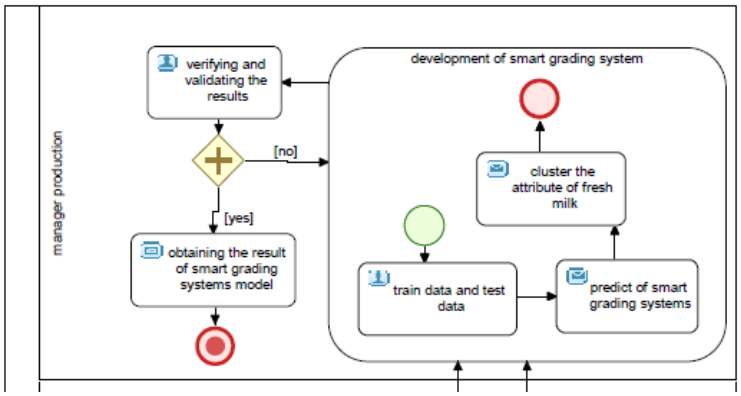


Figure 5. Process in classify of smart grading systems.

This study described the activity used to classify the quality of fresh milk. The system can be seen in figure 6. A milk farmer sent fresh milk, and then fresh milk was tested by specific gravity at cooperation. IoT systems have monitored the temperature of fresh milk in real-time. The quality data of fresh milk was collected and classified by the ANN model.

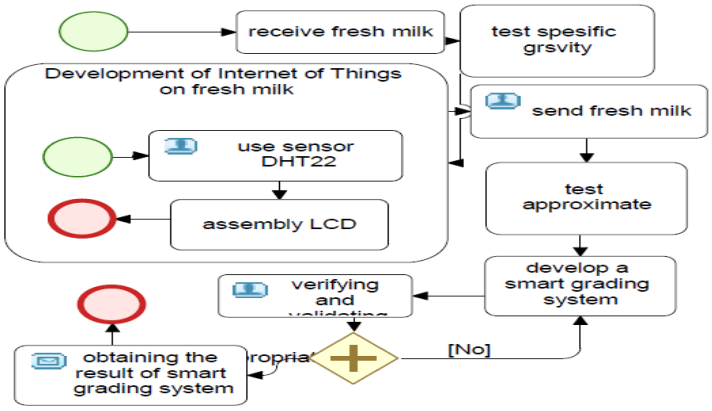


Figure 6. Process in smart grading based on Internet of things.

Business Process Modelling Notation (BPMN) diagrams helped know the stakeholders and their activity in the system. BPMN could describe the process and sub-process in a smart grading system of fresh milk based on IoT. BPMN represents all complicated relationships, processes, workflow, formulation, and stakeholders in the system, defined as implementing a business process model with a real-life business situation. Business process and methodology *Unified Modelling Language* (UML) was a system development method consisting of three stages: problem identification, analysis system, and design system [2]. *Smart* grading using sensor tools aims to improve ease of activity in determining the quality of fresh milk. One of the advantages of a smart grading system is the information system to determine the grade of fresh milk per farmer, which helps stakeholders determine the purchase price of fresh milk. This information system uses machine learning algorithms to predict using complex data input [5].

3.3 Classifying Using Artificial Neural Network

The design result with the ANN algorithm approach produced a predictive value, as shown in figure 6. The predicted accuracy value of fresh milk grade reached 98.74%. The accuracy value obtained from the prediction result of training data of 70% can be seen in table 2. In this study, the weight value on the

input affected the layer's output. The output of processed data produced a grade value of fresh milk. The value of Root Mean Square Error (RMSE) was 1,943.

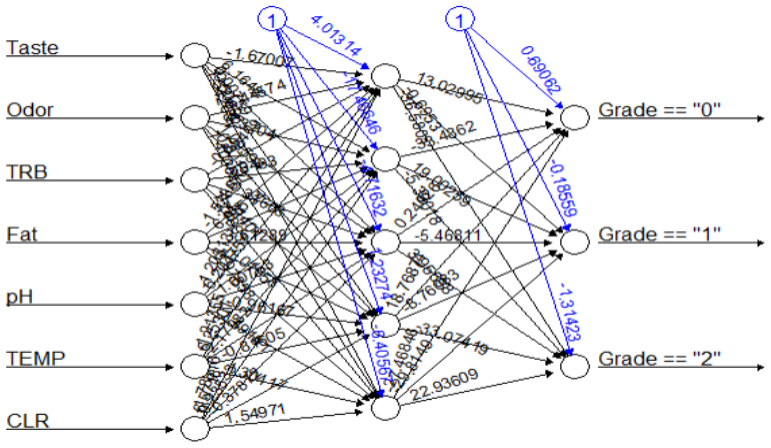


Figure 7. Network architecture prediction grade fresh milk.

An artificial Neural Network (ANN) is a trained computer network similar to the human brain. This study used supervised learning to reduce the error rate by adjusting the weight based on the difference between the target and the output to be achieved [18]. ANN was capable of handling large amounts of non-linear data interpolated. The hidden layer output showed three outputs.

This study agrees with [15] that the hidden layer is different from the input layer as it does not have any direct input. While the number of input and output layer designs and the number are determined by the inputs and outputs, respectively, finding the hidden layer design and number is not straightforward. The researchers have developed many design heuristics for the hidden layers; these different heuristics help the network behave the way they want it. Another sense of Artificial Neural networks was the paradigm of information processing inspired by the biological nervous system, such as the process of information in the human brain [13].

Tabel 2. Fresh milk grade prediction modeling results.

Testing data	Actual Value	Prediction Value
Grade 0 *(low)*	198	195
Grade 1 *(middle)*	304	300
Grade 2 *(high)*	238	235
Accuration		98.74%

The classification consisted of three types of grade: grade 0 means low-grade fresh milk, grade 1 mean medium grade fresh milk, and grade 2 means a high grade of fresh milk. Grade value is determined by the number of attribute values based on fresh milk quality standards. The data testing and training data provide an accuracy value of 98.74%. The accuracy value is generated from the number of predictive values divided by the actual value multiplied by 100%. ANN is important to mention here that neural networks are inspired by how the human brain learns. The recent development in these fields has led to the training of far dense neural networks, making it possible to capture signals of other machines [19].

3.4 Clustering of K-Means

The clusterization results in figure 7 show the optimal cluster value in cluster 5. Cluster 5 undergoes a defined valley tracing pattern that was likely to reach the optimal cluster [19]. Iteration results using the euclidean distance formula obtained two attribute types. Iteration calculations used *the euclidean distance* formula to find the final centroid value has been used as the cluster centre value [20].

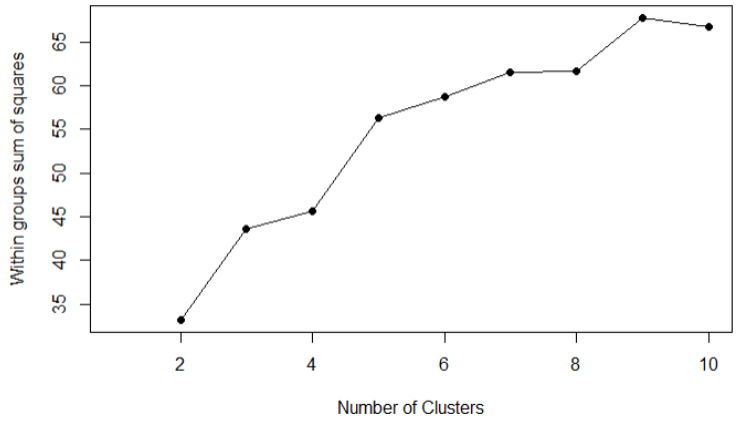


Figure 8. Optimal cluster value.

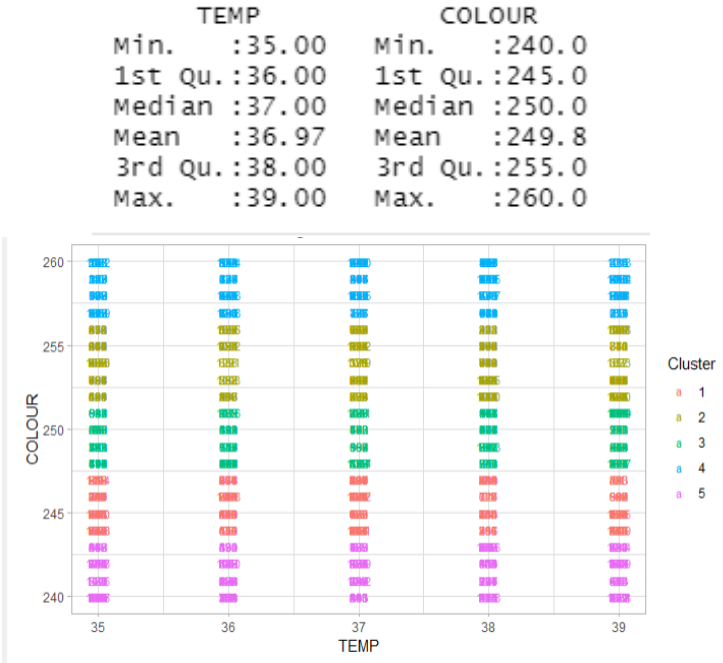


Figure 9. Relationship between two attributes.

K-Means is a method of analyzing data in which the modelling process is uncensored and is one of the methods that group data by partition [15]. The relationship between the two attributes, temperature and colour seen in figure 7. The average temperature was above 27.5, and the colour was 249.8 [1]. Attribute grouping resulted in five clusters. Cluster 3 showed temperature and colour values according to fresh milk quality standards. The clustering result is based on the value xi = 37 and the value yi = 250. The result presented that cluster 3 has the most efficient value. This study showed that K-Means could be polari by finding range values and improving better grouping capabilities [20].

4. Conclusion

The results of the analysis and design of the intelligent grading system of fresh milk were illustrated by the use case diagram and BPMN. Smart grading system design can be analyzed by stakeholders (dairy farmers, cooperative staff, IoT, laboratory staff, and production managers). Each actor has a task that helps determine the grade of fresh milk. Machine learning can develop smart grading systems with artificial neural networks and K-means algorithm models. ANN's algorithm could classify the grade of fresh milk based on quality standards that stakeholders have tested. The resulting grade of fresh milk consists of three types. There are low grades, middle grades, and high grades. The accuracy value of the ANN model development was 98.74%. Grouping using the K-means algorithm could cluster temperature and colour to determine the grade of fresh milk. The five types of clusters obtained the grouping results, and the best cluster was the third cluster with the best value following the quality standards of fresh milk. Based on the data analysis, the smart grading system made users save time knowing the grade of fresh milk easier.

5. References

[1] Badan Standarisasi Nasional 2011 *Susu Segar* SNI 01-3141-2011

[2] Choiriyati N, Arkeman Y, Kusuma W A, Arkeman Y, dan Kusuma W A 2020 Model deep learning for classify fragmen metagenom dengan spaced k- mers for extraction fitur Deep learning model for metagenome fragment classification using spaced k-mers 234-238

[3] Djatna T, Anggraeni E, dan Sailah I 2019 An analysis and design of downstreaming decision system on palm oil agroindustry based on multilabel classification *International Conference on Advanced Computer Science and Information Systems (ICACSIS)* 247-252

[4] Jain B R dan Dani P M 2020 Design internet of things for production planning small medium entreprises *Jurnal Sains Komputer & Informatika J-SAKTI* ***4*****(2)** 386-396

[5] Kendal, Booch 2007 *Analisis dan Perancangan System* (Jakarta: PT. Prehalindo)

[6] Kumar P, Ponnusamy V, Jain V 2020 Industrial internet of things and cyber-physical system: transforming the conventional to digital *IGI Global ACEE* ISBN 2327-039X

[7] Mahdavinejad M S, Rezvan M, Barekatain M., Adibi P, Barnaghi P, dan Sheth A P 2017 Machine learning for internet of things data analysis : *Digital Communications and Networks*

[8] Mahmud, Byamin, Mokdj 2018 Internet of Things based Smart Environmental Monitoring for Mushroom Cultivation *Indonesian Journal of Electrical Engineering and Computer Science* **10(3)** 847-852

[9] Maxwell A E, Warner T A, Fang F, Maxwell A E, Warner T A, Implementation F F, Maxwell A E, dan Warner T A 2018 Implementation of machine-learning classification in remote sensing : an applied review sensing : an applied review *International Journal of Remote Sensing* **39(9)** 2784-2817

[10] Misra N N, Dixit Y, Al-Mallahi A 2020 IoT, big data, and artificial intelligence in agriculture and the food industry *Internet of Things*

[11] Mursidah S 2020 Supply chain sustainability assessment system based on supervised machine learning techniques *: The Case for Sugarcane Agroindustry*

[12] Naik S dan Patel B 2017 Machine vision-based fruit classification and grading-a review In *International Journal of Computer Applications*

[13] Polepaka S, Swami D M, Ram K R P 2020 Internet of things and its applications: An overview *Lecture Notes in Electrical Engineering* **643** 67-75

[14] Pradika A Y, Chusniati S, Purnama M T E, Effendi M H, Yudhana A, Wibawati P A 2019 Uji total escherichia coli pada susu sapi segar di koperasi peternak sapi perah (kpsp) karyo ngremboko kecamatan purwoharjo kabupaten banyuwangi *Jurnal Medik Veteriner* **2(1)** 1

[15] Rakhmatulin I 2020 Neural network for automatic farm control *of Experimental & Theoretical Artificial Intelligence*
[16] Ramasubramanian K, Singh A 2019 *Machine Learning Using R* ISBN-13 (pbk): 978-1-4842-4214-8
[17] Reichenbach S E, Zini C A, Nicolli K P, Welke J E 2019 Benchmarking machine learning methods for comprehensive chemical fingerprinting and pattern recognition *Journal of Diary Farmer*
[18] Saragih J R, Hartama D, Wanto A 2020 *BACKPROPAGATION*
[19] Singh R, Dwivedi, Srivastava 2020 Internet of things based blockchain for temperature monitoring and counterfeit pharmaceutical prevention *MDPI Sensor* **20** 3950
[20] Tahsien S M, Karimipour H, Spachos P 2020 Machine learning-based solutions for the security of Internet of Things (IoT): A survey *Journal of Network and Computer*
[21] Waason C S 2006 *System Analysis, Design and Development Concepts, principals, and Practices* (John Willey Publication)
[22] Yang A, Zuangxun Y, Zhang 2019 Design of intrusion detection system for internet of things based on improved bp neural network *IEEE* **7**
[23] Yuan C 2019 *Study on K-Value Selection Method of K-Means Clustering Algorithm*
[24] Zhu Q, Pei J, Liu X, Zhou Z 2019 Analyzing commercial aircraft fuel consumption during descent : A case study using an improved K-means clustering algorithm Analyzing commercial aircraft fuel consumption during descent : A case study using an improved K-means clustering algorithm *Journal of Cleaner Production* 869-882

6. Acknowledgements

We sincerely thank Indonesia Endowment Fund for Education (LPDP - Lembaga Pengelola Dana Pendidikan), Ministry of Finance, the Republic of Indonesia, for granting the scholarship and supporting this research. The authors also would like to thank Prof Taufik Djatna from IPB University who provided insight and expertise that greatly assisted in conducting the research and writing this article.