

Improving Bread Sales Predictions through Extreme Learning Machine (ELM)

Nurhayati Sembiring¹, Erma Dwi Yanti², and Khairina Mahfuzah Sibarani³

^{1,2}Department of Industrial Engineering, Universitas Sumatera Utara, Medan, Indonesia

* Corresponding author: nurhayatipandia68@usu.ac.id

ABSTRACT

The bakery sector is experiencing growth, with enterprises like XY Bakery & Cake Shop providing their products directly to consumers. The company's bread sales have exhibited irregular pattern, rendering the task of establishing precise sales forecasts difficult. This inconsistency has resulted in financial losses, such as the spoilage of bread within three days, the squandering of raw materials, labor, and other forms of inefficiency. The objective of this study is to forecast bread sales and evaluate the accuracy of the predictions using the Extreme Learning Machine (ELM) approach. The accuracy will be assessed using the Mean Square Error (MSE) measure at AR Bakery. The ELM approach is utilized to predict bread sales, using normalized data divided into 80% for training and 20% for testing. The experimental tests conducted with 8 neurons produced a mean squared error (MSE) value of 0.27402, whereas utilizing 5 neurons resulted in an MSE of 0.28761. The minimum error value was achieved using 5 neurons. 10 predictions were made using a dataset consisting of 101 data points for each type of bread. In December, the sales reached their peak, with a total of 198,850 units sold.

Keywords: Bread industry, Sales prediction, Extreme Learning Machine (ELM), Mean Square Error.

1. INTRODUCTION

The food and beverage industry continues expanding with new enterprises varying in size. A key segment of the industry is bakeries, demonstrating reliable growth and importance over time [1].

Medan's proliferating bakeries has heightened competition within the industry through diversity in bread types, flavours, textures and affordably priced quality products. XY Bakery & Cake Shop exemplifies the growing bakery businesses in Medan, progressively expanding across North Sumatra and Aceh. The assortment of products sold include sweet buns, doughnuts and bread in varying flavors, bika ambon, traditional cakes, tarts, pies, and cakes. The company experiences fluctuating bread sales, making accurately predicting sales challenging, thus resulting in overproduction. Excess production every month can lead to a large number of leftover breads, incurring the company significant losses such as wastage of raw materials and labor costs. Bread in XY Bakery & Cake Shop is considered unfit for sale after three days as bread can spoil or mold within 2-5 days [2]. This is marked by the appearance of mold, resulting in an unpleasant odor and taste, hard texture, dark color, stickiness, and changes in the bread's color.

Extreme Learning Machine (ELM) is an advanced algorithm used to train single-layer feedforward neural networks, providing exceptional performance and high learning speed. ELM provides good generalization performance with a large number of hidden neurons by randomly assigning the input weights and hidden neuron biases, which remain unchanged during the training process, while the output weights are computed analytically [3].

A study in the field of forecasting focuses on sales forecasting for fashion products. This research, conducted by Yu, Choi, and Hui in 2011, is titled "An Intelligent Fast Sales Forecasting Model for Fashion Products." The study employs the ELM (Extreme Learning Machine) technology, specifically the Intelligent Fast Forecasting (IFSFM) technique, to anticipate fashion product sales. This prediction is based on sales data and incorporates many criteria such as clothing color, clothing size, and clothing price. Based on this research, testing conducted utilizing the initial data set demonstrated that ELM was capable of delivering high accuracy while adhering to the specified time limitations [4].

Another study conducted by Ayustina Giusti et al. (2018) titled "Prediction of Noodle Sales Using Extreme Learning Machine (ELM) Method at Kober Mie Setan Soekarno Hatta Branch" aims to forecast restaurant sales by employing the Extreme Learning Machine (ELM) method. The approach involved several steps: data normalization, training, testing, data denormalization, and error calculation using Mean Square Error (MSE) [5]. The ELM approach offers benefits in terms of rapid learning and minimal error rates.

A previous study conducted by Didik Siswanto et al. (2022), titled "Application of Apriori Algorithm in Developing an Application to Determine Bread Production Patterns," developed a web-based system that could identify patterns in bread production data for the purpose of devising marketing strategies [6]. This study focuses on forecasting bread sales through the utilization of the Extreme Learning Machine (ELM) methodology and aims to assess the accuracy of the predictions by calculating the prediction error rate using Mean Square Error (MSE). By utilizing this bread sales prediction, the company may accurately anticipate future bread sales, thus minimizing financial losses and maximizing profits.

2. METHODS

This research falls under the category of a case study. A case study focuses on the analysis of multiple variables within a limited number of units. Case studies are designed to thoroughly examine the context and relationships of a specific social entity being studied, offering a comprehensive account of its origins, characteristics, and distinctive features [7]. This study incorporates both dependent and independent variables as research factors. The dependent variable is the outcome that is influenced by the independent variable. In this study, the dependent variable is the predicted production amount for the year 2024 and the accuracy of the prediction [8]. The independent variable is the factor that directly causes or exerts impact on the emergence or outcome of the dependent variable. The independent variable in this study is the historical sales data of bread.

The data sources used in this research are primary and secondary data. The primary data used in this study includes sales data, production data, and sales residue data. Meanwhile, secondary data includes data obtained from literature studies, journals, books, company profile data, and both historical and current company data. The data collection methods employed in this research are observation, interviews, and literature study. The data processing method used in this research is the Extreme Learning Machine (ELM) algorithm.

2.1 Learning Machine

A learning machine is a computer program designed to achieve specific criteria or performance using training data or past data. According to Huang, G. B., Zhu, Q. Y., Siew, C. K (2004), "Learning Machine studies the theory to enable computers to "learn" from data. This learning involves various disciplines such as statistics, computer science, mathematics, and neurology. The models can be predictive (to forecast the future), descriptive (to gain knowledge from data), or a combination of both." Machine learning enables computers to solve various problems and learn independently, similar to humans. In general, machine learning algorithms are divided into three categories [9]:

1. **Supervised Learning** This algorithm instructs the computer to generate the desired output. It learns the knowledge of the relationship between the input and output. The potential outcomes are already known, and the data utilized for training is labelled. Supervised learning is a valuable method for generating predictions by utilizing training data.
2. **Unsupervised Learning** This algorithm uses data that has not been labeled. It utilizes unlabelled training data to model the structure of the data. Unsupervised learning is useful in finding implicit links within the data without any given labels. Hence, this method lacks the ability to forecast future outcomes due to the absence of any correlation between the input variable x and the output variable y .
3. **Reinforcement Learning (RL)** This algorithm incorporates feedback from the environment using iterative and adaptive learning techniques. RL is thought to bear a strong resemblance to human learning. In reinforcement learning, training data is not provided in advance; instead, data is acquired through experiential learning. This algorithm lacks a clearly defined objective; instead, it is compelled to acquire the ideal value through a process of experimentation and refinement.

2.2 Extreme Learning Machine (ELM) Method

The Extreme Learning Machine (ELM) method is a new approach in feedforward Artificial Neural Networks (ANN) with a single hidden layer, known as Single Hidden Layer Feedforward Neural Networks (SLFNs) [10]. This method is designed to improve the learning speed of feedforward ANN. The Extreme Learning Machine (ELM) algorithm does not train input weights or biases but trains to obtain output weights using the norm-least-squares solution and the Moore-Penrose inverse in a general linear system. The Extreme Learning Machine (ELM) is called a fast-learning machine because it uses random input weights and biases, resulting in higher learning speeds [11].

This study aims to estimate bread sales in a bakery by employing the Extreme Learning Machine (ELM) approach. The goal is to aid the company in accurately projecting sales. The input required for predicting bread sales using the ELM technique consists of sales data, variations in historical sales data features, and the number of neurons in the hidden layer. The objective of the prediction is to determine the most optimal alternative for maximizing bread sales and the company's production profits. The calculating parameters considered in this study include the amount of data, the features or variations used, the activation function, and the number of neurons in the hidden layer.

Once the bread sales data has been collected, the subsequent step involves creating training and testing data. It is essential for the data to be normalized using Min-Max Normalization beforehand. The data undergoes a training and testing phase, during which the prediction error is determined using the Mean Square Error (MSE) metric. Subsequently, the projected data is restored to its original form using denormalization. The output result is compared to the real bread sales value and analyzed using the Mean Square Error (MSE) method. This program produces two outputs: the Mean Square Error (MSE) value and the denormalized prediction value.

3. RESULTS AND DISCUSSION

The objective of forecasting bread sales is to determine the error rate produced by the Mean Square Error (MSE) when using the Extreme Learning Machine (ELM) method. The input data consists of previous monthly sales data. This study employs the Min-Max Normalization method for data normalization and utilizes the Sigmoid Activation Function as the activation function. The steps taken to predict bread sales using the Extreme Learning Machine (ELM) approach is as follows:

1. System input in the form of historical sales data, the number of neurons in the hidden layer, and the number of features. This study uses the sigmoid activation function.
2. Initialization of input weights and biases.
3. Data normalization using Min-Max Normalization.
4. The data normalized data is utilized for both the training and testing processes. The outcomes of this testing procedure are predictions that are unreadable because they have not been denormalized.
5. The prediction results after the testing process, before being renormalized, are evaluated by calculating the error value using Mean Square Error (MSE).
6. The system output is the prediction results that have been renormalized.

3.1 System Implementation

This section will discuss the system implementation process, including the code that outlines the processes for applying the Extreme Learning Machine (ELM) method to predict bread sales.

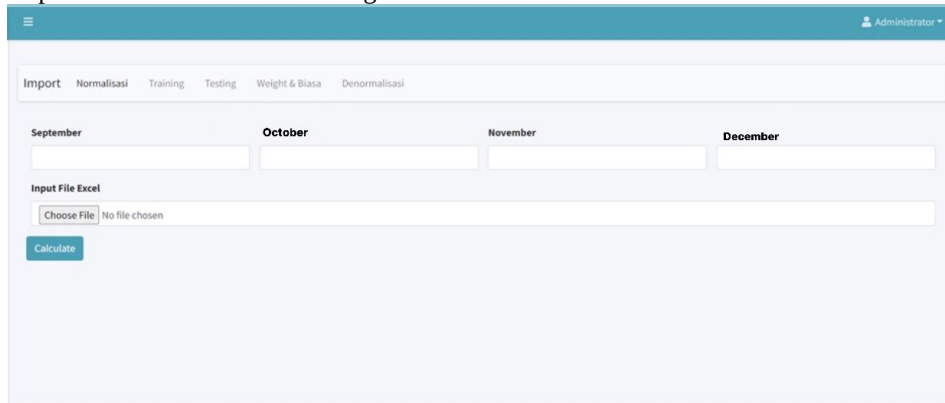
1. **Implementation of Data Normalization Process** During the data normalization process, the system employs all data with predefined features. It involves identifying the minimum and maximum values for each feature in the initial dataset and then normalizing the data based on these values. The normalized data is subsequently utilized in both the training and testing phases.
2. **Implementation of Input Weight and Bias Initialization Process** During the implementation stage of initializing input weights and biases, these parameters are randomly selected within a range of -1 to 1. The number of input weights and biases is determined by the number of neurons entered by the user into the system.
3. **Implementation of Hidden Layer Output Calculation Process** At this stage, calculations involving the normalized data and the input weights and biases are performed. This procedure is applied during both the training and testing phases.

4. **Implementation of Hidden Layer Output Calculation** The hidden layer output calculation involves utilizing the sigmoid activation function to compute the hidden layer output. This procedure is applied during both the training and testing phases.
5. **Implementation of Output Weight Calculation Process** During the implementation of the output weight calculation process, the training data undergoes a series of steps as previously described.
6. **Implementation of Mean Square Error (MSE) Calculation** The process of calculating the Mean Square Error (MSE) involves analyzing the predicted data, where the error value is obtained by comparing the normalized target with the predicted values, following the previously described method.
7. **Implementation of Data Denormalization Process** During the data denormalization process, the predicted data is converted back to normal data.

3.2 Interface Implementation

This research utilizes web-based implementation adhering to the design and testing phases. The application consists of several pages, including data import, normalization, training, testing, denormalization, and weight & bias pages.

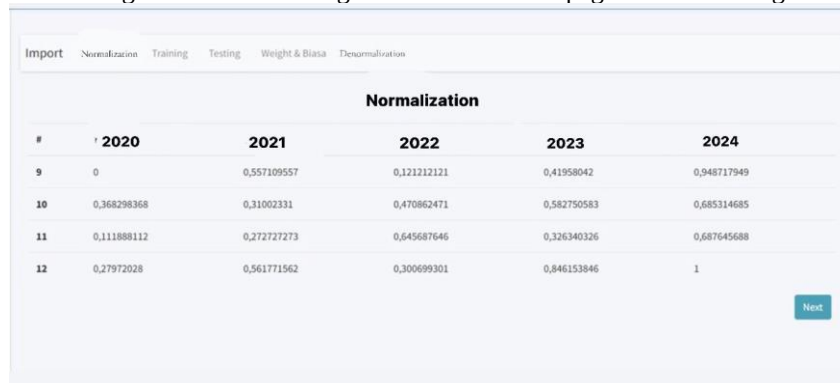
1. **Data Import Implementation** The data import page prompts users to enter various datasets, including historical bread sales data. Users are also required to provide variations or features from this historical data and specify the number of neurons in the hidden layer. The data import implementation can be seen in Figure 1 as follows.



Source: Data Processing

Figure 1. Data Import Implementation

2. **Normalization Implementation** The normalization page displays the results of normalizing the initial data input by the user, including historical bread sales data. The normalization process involves 101 records from January 2023, encompassing bread sales data from XY Bakery & Cake Shop. The system utilizes 8 neurons in the hidden layer, with the data split into 80% for training and 20% for testing. The normalization page is shown in Figure 2 as follows.



#	2020	2021	2022	2023	2024
9	0	0,557109557	0,121212121	0,41958042	0,948717949
10	0,368298368	0,31002331	0,470862471	0,582750583	0,685314685
11	0,111888112	0,272727273	0,645687646	0,326340326	0,687645688
12	0,27972028	0,561771562	0,300699301	0,846153846	1

Source: Data Processing

Figure 2. Normalization Page Implementation

3. **Implementation of Training Page** The training page displays the results of the data training process. This involves several steps, including calculating the hidden layer output, which is then processed through the sigmoid activation function. The weight output from this training process is derived from 101 records of January 2023 bread sales data. The implementation of the training page can be seen in Figure 3 below.

Training					
H	1	2	3	4	5
9	0,60737047	0,524772906	0,642780211	0,567549107	0,60737047
10	0,60737047	0,568896305	0,695671264	0,492296456	0,621105931
11	0,621105931	0,580424444	0,691008818	0,465152505	0,616805718
12	0,649580247	0,564998804	0,758422702	0,378778107	0,649580247

Source: Data Processing

Figure 3. Training Page Implementation

4. **Implementation of Testing Page** The testing page displays the results of the data testing process. This involves several steps, including calculating the hidden layer output, which is then processed through the sigmoid activation function. The weight output from this testing process is derived from 101 records of January 2023 bread sales data. The implementation of the testing page is shown in Figure 4 below.

Testing					
#	2020	2021	2022	2023	2024
9	0,436272439	0,099172826	0,587451602	0,271858501	0,436272439
10	0,436272439	0,277349525	0,826768743	-0,030816614	0,494244984
11	0,494244984	0,324516015	0,804839845	-0,139616332	0,476011968
12	0,61719465	0,261474833	1,14405139	-0,494737768	0,61719465

Source: Data Processing

Figure 4. Implementation of Testing Page

5. **Implementation of Weight Page** The of weight page serves to display the values of the input weights and biases, which are randomly generated by the system. The weight page of the system can be seen in Figure 5 below.

Weight				
β	1	2	3	4
9	0,070877426	0,071130896	0,070732488	0,070522987
10	1,608817613	1,614551504	1,605503605	1,600773325
11	1,324778712	1,329254219	1,321734813	1,317531819
12	2,625177885	2,637111951	2,623078055	2,618183103
13	0,07060679	0,07085982	0,070463146	0,070255158

Source: Data Processing

Figure 5. Implementation of Weight Page

6. **Denormalization Page Implementation** The denormalization page serves to display the process of converting the prediction results obtained from the testing process back to their original form. The denormalization page implementation is shown in Figure 6 below.

Import	Normalization	Training	Testing	Weight & Bias	Denormalization
Denormalization					
#	2020	2021	2022	2023	2024
9	410	649	462	590	817
10	568	543	612	660	704
11	458	527	687	550	705
12	530	651	539	773	839

Source: Data Processing

Figure 6. Denormalization Page Implementation

3.3 Testing and Analysis of Calculation Results Using Extreme Learning Machine (ELM) Method

The objective of this testing stage is to achieve the smallest or most optimal Mean Square Error (MSE). This study conducted three types of experiments: varying the number of neurons in the hidden layer, the amount of historical sales data, and the number of features from historical sales data. Each experiment is conducted ten times, and the results are averaged to obtain the MSE

1. Testing the Number of Neurons in the Hidden Layer

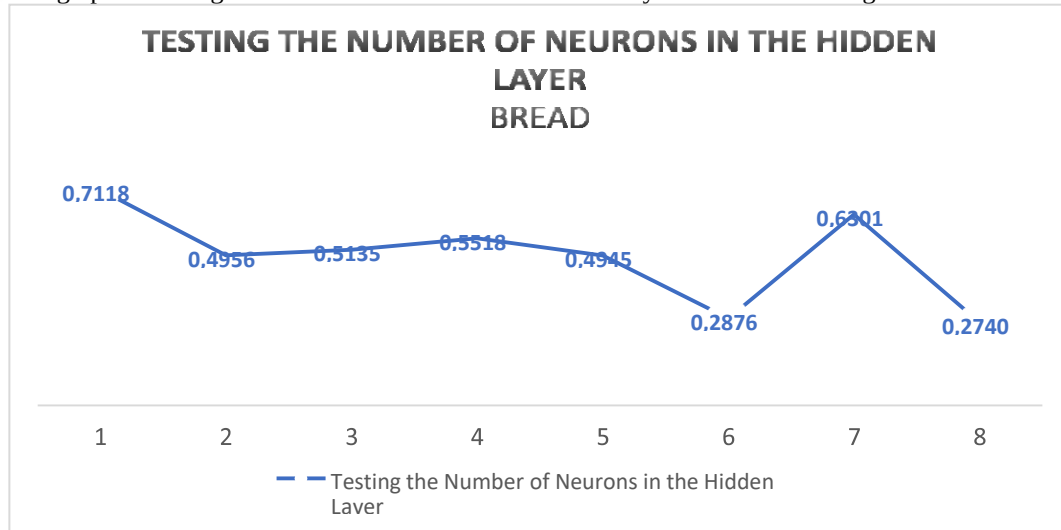
This experiment aims to determine the smallest or optimal Mean Square Error (MSE) by varying the number of neurons in the hidden layer. The experiments range from 1 to 8 neurons. Historical sales data from January 2023, comprising 101 data points for each type of bread, is utilized. The study tests four features, including historical sales data from the previous four days. The testing of the number of hidden layers is shown in Table 1 below.

Table 1. Testing the Number of Neurons in the Hidden Layer with Sigmoid Function

X	X1	X2	X3	X4	X5	X6	X7	X8
1	0,726432	0,460705	0,437606	0,556601	0,531962	0,283257	0,643828	0,230365
2	0,680495	0,528146	0,512396	0,537101	0,472269	0,304898	0,619421	0,278373
3	0,73019	0,505655	0,547597	0,538029	0,498624	0,305596	0,597749	0,285116
4	0,789499	0,505469	0,558009	0,558453	0,477207	0,198258	0,666023	0,184287
5	0,712684	0,468528	0,503133	0,572823	0,505129	0,285906	0,638655	0,297775
6	0,750852	0,506988	0,556089	0,551338	0,487426	0,267974	0,641019	0,259917
7	0,68028	0,486561	0,49668	0,538021	0,503414	0,362325	0,616398	0,307553
8	0,657569	0,441832	0,449384	0,575762	0,51142	0,322541	0,625647	0,327586
9	0,571076	0,477588	0,390516	0,552128	0,545029	0,392175	0,59206	0,406661
10	0,714746	0,542686	0,564658	0,549699	0,465045	0,267415	0,632147	0,277398
11	0,720356	0,529314	0,588955	0,541303	0,445655	0,269302	0,621637	0,265443
12	0,807978	0,494718	0,557632	0,55117	0,491	0,191617	0,666974	0,167797
AVE	0,711846	0,495683	0,513555	0,551869	0,494515	0,287605	0,63013	0,274023

Source: Data Processing

The experiment utilizes data from day h-4 to day h-1 preceding the target prediction and employs the sigmoid activation function. The ratio of training to testing data is 80% to 20%, respectively. The graph of testing the number of neurons in the hidden layer can be seen in Figure 8.



Source: Data Processing

Figure 8. Graph of Testing the Number of Neurons in the Hidden Layer with Sigmoid Function

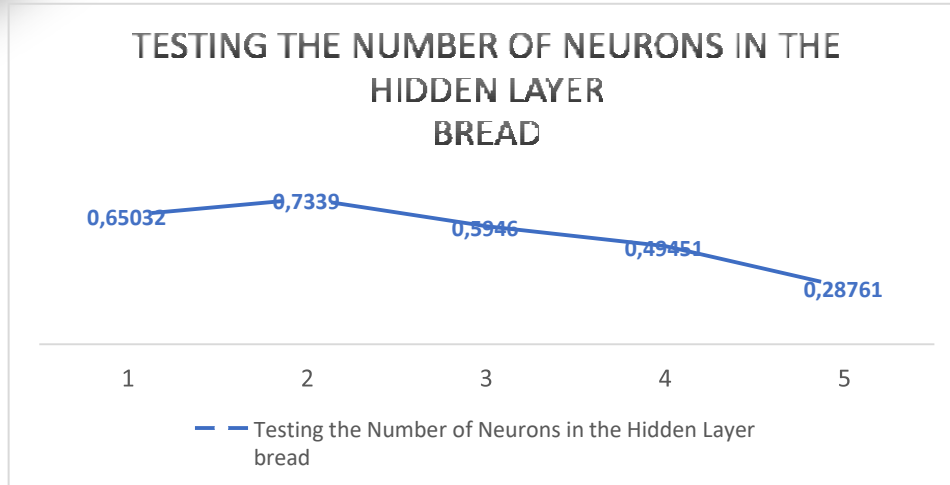
Each test for the number of neurons in the hidden layers is conducted ten times. In each trial, the system displays the MSE value. The average MSE value from these ten trials is then compared with the results from other tests with different numbers of neurons in the hidden layers, as shown in Table 2 below.

Table 2. Testing the Number of Hidden Layers

X	X1	X2	X3	X4	X5
1	0,649973	0,742935	0,623361	0,531962	0,283257
2	0,641253	0,727329	0,611592	0,472269	0,304898
3	0,643017	0,725365	0,583889	0,498624	0,305596
4	0,700467	0,804139	0,609039	0,477207	0,198258
5	0,64491	0,732785	0,577379	0,505129	0,285906
6	0,682824	0,755662	0,590525	0,487426	0,267974
7	0,639673	0,698103	0,586889	0,503414	0,362325
8	0,597961	0,689	0,561081	0,51142	0,322541
9	0,569361	0,617931	0,605345	0,545029	0,392175
10	0,669392	0,766068	0,614871	0,465045	0,267415
11	0,655792	0,748	0,570294	0,445655	0,269302
12	0,709159	0,799437	0,600954	0,491	0,191617
AVE	0,650315	0,733896	0,594602	0,494515	0,287605

Source: Data Processing

Based on the data presented in Table 5.9 and Figure 5.9, the best MSE value for the bread type is achieved when using 8 neurons in the hidden layer, with an MSE value of 0.287605.



Source: Data Processing

Figure 9. Graph of Testing the Number of Hidden Layers

The results of the interface system testing indicate that when the test identifies the smallest Mean Square Error (MSE) or produces a low error value, the interface system displays the highest sales predictions in units of pcs, as shown in Figure 6 and Table 2. Table 2 displays the results of testing the number of neurons in the hidden layer. The interface system displays the best (smallest) MSE value of 0.287605.

This smallest MSE value results in a prediction of 4,850 pcs of bread in the denormalization stage of the system, as shown in Figure 6. The system indicates the highest sales in December as 4,850 pcs of bread per month for one outlet, and for all outlets, it is shown in Table 3 below.

Table 3. Predicted Bread Sales Results in the System

Month	Bread Sales /Outlet (psc)	Bread Sales /Totals Outlet (psc)
January	3530	144.730
February	3530	144.730
March	3903	160.023
April	4720	193.520
May	3547	145.427
June	4100	168.100
July	3200	131.200
Agust	2890	118.490
September	2109	86.469
October	4054	166.214
November	4105	168.305
December	4850	198.850

Source: Data Processing

This demonstrates that the developed forecasting model performs well, providing accurate predictions and effectively explaining the forecasted data.

4. CONCLUSION

The implementation of the Extreme Learning Machine (ELM) approach was carried out through an application featuring several interface displays:

1. The Extreme Learning Machine (ELM) algorithm can predict the sales volume with low error values. The normalized data is divided into two parts: 80% for training data and 20% for testing data. The forecasted bread production amounts for 2024 using Extreme Learning Machine (ELM) are as follows: January 144.730 pcs, February 144.730 pcs, March 160.023 pcs, April 193.520 pcs, May 145.427 pcs, June 168.100 pcs, July 131.200 pcs, August 118.490 pcs, September 86.469 pcs, October 166.214 pcs, November 168.305 pcs, December 198.850 pcs.
2. The prediction matrix values for bread production in 2024 are derived from the Mean Squared Error (MSE). Testing with 8 neurons resulted in an MSE of 0.27402, while testing with 5 neurons yielded an MSE of 0.28761. The tests, conducted 10 times with 101 data points per bread type, indicated that using 5 neurons produced the smallest error value. December was the highest-selling month, with 198,850 pcs sold across all outlets. This indicates that the forecasting model built performs well, enabling accurate predictions and effective explanation of the forecasted data.

REFERENCES

- [1] Misran, N. N., Sharif, M. S. M., Saad, M., Karim, S. A., & Ramly, A. S. (2021). Family Traditional Bakery Business Sustainability: An Analysis on its Contributing Success Factors. *International Journal of Academic Research in Business and Social Sciences*, 11(9), 244–260.
- [2] Rustanto, D., Anam, C., & Parnanto, N. H. R. (2018). Chemical Characteristics and Determination of Shelf Life of White Bread with Addition of Calcium Propionate and Nipagin. *Journal of Food Science and Agricultural Results*, 2(2), 121-133.
- [3] Khalil, A. A., et al. (2018). Two Swarm Intelligence Approaches For Tuning Extreme Learning Machine. *Neurocomputing*, 9, 1271-1283.
- [4] Yu, Y., Choi, T.M. and Hui, C.L., 2011. An intelligent fast sales forecasting model for fashion products. *Expert Systems with Applications*, 38(6), 7373-7379.
- [5] Giusti, A., et al. (2018). Sales Prediction Using Extreme Learning Machine (ELM) at Kober Mie Setan Soekarno Hatta Branch. *Journal of Information Technology and Computer Science Development*, Malang.
- [6] Chandra and K. R. Dewi, "Implementation of Data Mining Sales of Milk Using Apriori Algorithm Method," *IOPConf. Ser. Mater. Sci. Eng.*
- [7] Woodside, A. G. (2010). *Case study research: Theory, methods and practice*. Emerald Group Publishing.
- [8] Alwiyah, Toto Thinakone Louangdy, and Aulia Yolandari. (2018). Relation of Relationship Between Research Theory and Variable with Management Case Study. *Ilearning Journal Center*.
- [9] Huang, G. B., Zhu, Q. Y., Siew, C. K., Extreme Learning Machine : a New Learning Scheme of Feedforward Neuralnetworks, in : *Proceedings of International Joint Conference on Neural Networks, IJCNN2004*, page : 25-29, July 2004, Budapest, Hungary.
- [10] Huang, G. B., Zhu, Q. Y., & Siew, C. K. (2006). Extreme Learning Machine: Theory and Applications. *Neurocomputing*, 70, 489-501.
- [11] Humain, Q. (2015). Extreme Learning Machine (ELM) Artificial Neural Network for Weather Prediction in Malang Region. Department of Mathematics, Maulana Malik Ibrahim State Islamic University.