talenta PUBLISHER

Jurnal Sistem Teknik Industri
Journal homepage: https://talenta.usu.ac.id/jsti

The Design Of A Demand Forecasting Model Of Glass Bottled Tea Products With Machine Learning Approach

Said Munal Akid[1] , Aulia Ishak[*1] , Sukaria Sinulingga[1]
[1]*Department of Industrial Engineering, Faculty of Engineering, Universitas Sumatera Utara, Jl. Dr. T. Mansur No.9, Medan, 20155, Indonesia*
[*]*Corresponding Author: aulia.ishak@usu.ac.id*

ARTICLE INFO

Article history:
Received 30 June 2024
Revised 2 December 2024
Accepted 2 January 2025
Available online 29 April 2025

E-ISSN: 2527-9408
P-ISSN: 1411-5247

How to cite:
Akid, S. M., Ishak, A., & Sinulingga, S. (2025). The Design Of A Demand Forecasting Model Of Glass Bottled Tea Products With Machine Learning Approach. *Jurnal Sistem Teknik Industri*, 27(2), 57-65.

http://doi.org/10.26594/register.v6i1.idarticle

ABSTRACT

An accurate sales forecasting is crucial to the profits earned because it affects the company's stock management. With computational support, machine learning and artificial intelligence can continuously and automatically recognize patterns in data, thereby reducing the risk of demand unpredictability. PT XYZ is one of the companies in industrial sector that produces various beverage products. The factory in Medan. One of the products is the tea glass bottle. At PT XYZ, there are frequent differences between forecasting data and sales data, causing high error rates in production planning accuracy. This study aims to analyze the most effective model for forecasting future sales by comparing the accuracy of a Machine Learning-based forecasting model with the existing forecasting method currently employed at PT XYZ. This research was conducted using the Recurrent Neural network (RNN) method as part of the Machine Learning approach. The data that was inputted to the programme was weekly demand data, calendar day off data, temperature data, and population data. The forecasted data is weekly demand. Based on the company's historical data, a demand graph is obtained which has a cyclical pattern. From the results of forecasting using Machine Learning, an accuracy value of 99.47% is obtained with an error rate of 0.53%, which is still below the tolerance limit set by the company. The error rate shows a decrease of 14.72% compared to the error value in the previous company model. This decrease is expected to help control inventory more effectively.
Keyword: Forecasting, Machine Learning, Demand, Recurrent Neural Network

ABSTRAK

Peramalan penjualan yang akurat sangat penting bagi keuntungan perusahaan karena berpengaruh langsung terhadap manajemen persediaan barang. Machine Learning merupakan kecerdasan buatan yang dapat mengurangi resiko ketidakpastian permintaan dengan cara melakukan pengenalan data secara terus menerus dan otomatis dengan bantuan komputasi. PT. XYZ adalah perusahaan besar yang memproduksi berbagai jenis minuman di pabriknya yang berlokasi di Medan. Salah satu produk yang diproduksi adalah teh botol kaca. Pada PT. XYZ sering terjadi selisih data peramalan dengan data penjualan sehingga menyebabkan tingginya niai error pada akurasi perencanaan produksi. Penelitian ini bertujuan untuk menganalisis model terbaik untuk memprediksi penjualan kedepan dengan membandingkan nilai akurasi model forecasting Machine Learning dengan model forecasting existing pada PT. XYZ. Penelitian ini dilakukan menggunakan metode Recurrent Neural network (RNN) sebagai bagian dari pendekatan Machine Learning. Data yang dimasukan pada program adalah data permintaan mingguan, data hari libur kalender, data suhu, dan data penduduk. Data yang diprediksi adalah permintaan mingguan. Berdasarkan data historis perusahaan, didapatkan grafik permintaan yang memiliki pola siklis. Dari hasil peramalan menggunakan Machine Learning, diperoleh nilai akurasi sebesar 99,47% dengan tingkat error sebesar 0,53%, yang masih berada di bawah batas toleransi yang ditetapkan oleh perusahaan. Tingkat error tersebut menunjukkan penurunan sebesar 14,72% dibandingkan dengan nilai error pada model perusahaan sebelumnya. Penurunan ini diharapkan dapat membantu mengendalikan inventori secara lebih efektif
Keyword: Peramalan, Machine Learning, Permintaan, Recurrent Neural Network

1. Introduction

Competition in the world of business is increasingly competitive, requiring each company to improve their performance in order to compete with other similar companies. Improving the performance of the company can be done by improve the effectiveness and efficiency of the company which is done with a systematic and comprehensive approach. Nowadays, business competition requires businesses to be more sensitive to the needs and demands of consumers for the products that they offer. So it takes a management that plays a role in determining the number of sales. The successful of a company is reflected by the ability of planning at management to optimally utilise opportunities in order to produce sales and profits according to the expected results. [1]. The company's ability at management will reflect the success in utilising business opportunities as optimally as possible, so as to produce optimal production and profit as expected. The utilisation of business opportunities is approached by examining current and past situations and looking at how they will affect the future situation. An important task of company management is to planning the futures so that all predicted possibilities and opportunities can be actualised [2]. The prediction results can be used as considerations for making decisions in business management planning [3].

PT XYZ as one of the largest beverage production company in Indonesia requires an analysis of its production sales, especially for maintaining production stability and avoiding losses. Accurate sales forecasting is crucial to the profits earned by retail companies because it affects the company's stock management. A good stock management will minimise the existence of Out of Stock and Over Stock items which will directly affect the profits earned by the company. Therefore, forecasting is quite important for the retail business process. It is a problem at PT XYZ that there are frequent discrepancies between forecasts and sales data, which results in a high level of production planning error. The difference can cause problems in the production department. One of the problems caused by this difference is the management of raw materials and packaging materials. Demand forecasting is a key part of activity in the supply chain. It is an important process to estimate the total demand for goods or services in a given period, as well as to determine appropriate marketing targets [4].

In general, there are two types of methods in time series forecasting, namely statistical methods (Statistical Time Series) and computational intelligence or Machine Learning methods[5] .The implementation of artificial intelligence is often applied in various cases in life, such as Artificial Neural Network (ANN). An information processing system that shares traits with biological neural networks is called an artificial neural network (ANN), often referred to simply as a neural network (NN) [6]. In fact, the data obtained is most often non-linear. With non-linear data that can cause market demand to changing can be solved using Machine Learning methods because it can minimise the risk of demand uncertainty by identifying the data continuously and automatically with the help of computations. The Deep Learning method, which is a subset of Machine Learning, has a better ability to capture non-linear patterns in data, in addition, if the data is quite large, Deep Learning is more capable to enhance their performance when compared to statistical methods that do not provide a better forecasting result even with more data availability [7][8].

One of the algorithms that belongs to the Deep Learning family is RNN (Recurrent Neural network). RNN is a class of ANN (Artificial Neural network), where the hidden layer architecture allows the previous output to become the current input in the current state [9]. RNNs can recognise temporal dependencies and is suiTable for forecasting sequential data [10]. Two important concepts in Machine Learning model development are data training and data testing. Training is the process of model building, while testing is the process of testing the performance of the learning model [11]. Machine Learning's main advantage is its ability to learn and update knowledge automatically. This allows AI researchers to avoid the need to manually set rules, instead letting Machine Learning models automatically learn patterns and relationships between variables. This provides effective and efficient advantages in many areas of life. One major drawback of Machine Learning models is their dependence on data. Since Machine Learning models operate by analyzing training data, poor data quality such as noise, outliers, and others can significantly affect the performance of Machine Learning models [12].

In this study, a model design was carried out with several factors, namely temperature, population and day off factors. An increase in population will have an impact on increasing demand for goods. When the population of an area increases, the demand for goods will also increase [13]. The number of day off can affect sales such as a case study conducted by Ilham et al, [14][15]. PT XYZ stated that day off affect demand. The more day off, the demand for bottled tea drinks will decrease. Temperature in an area can

significantly influence consumers in their decisions on a product [16]. Based on the explanation mentioned earlier, it is recommended to revise and improve on PT XYZ by designing a demand forecasting model with Machine Learning to forecast demand with a better accuracy.

2. Methodology

The aim of this research is to enhance the precision of the existing forecasting model by designing a new one utilizing a Machine Learning approach based on Recurrent Neural Networks (RNNs). This research is an experimental research with a quantitative approach. This research emphasises the analysis of processed data by analyzing the variables of day off, temperature and population. This research is simulation-based to determine the accuracy value of the forecasting model that is designed for the work system at PT XYZ. The forecasting model referred to is to determine if the new variables and new methods that are used in the design of the model can solve the existing problems at PT. XYZ. The data being collected are sales data, temperature data, population data, day off data for 3 years. The results of the forecasting model were analysed using two methods, which is the Mean Squared Error (MSE) and Root Mean Squared Error (RMSE) methods. The experimental methodology used in this study allows researchers to accurately plan future demand by considering new variables. The findings of this study are expected to contribute to improving operational efficiency in areas such as inventory management, production, and distribution within the company.

In forecasting research using the Machine Learning approach, there are several stages of data processing that must be carried out, including processing raw data to be input to google colab, coding Machine Learning, simulating with Machine Learning. The following is a more detailed processing stage:

1. Data Plotting
   The first step in data processing is data plotting. data plotting is done in order to determine the appropriate forecasting method for a data pattern.
2. Processing Raw Data Before Machine Learning
   Before the research proceeds to the Machine Learning stage, data processing must first be carried out according to what is needed so that it can be input in google colab. Google colab is used as a means to do Machine Learning. In this research, the data needed before being inputted into google colab is data on the number of products in the provinces of Aceh, North Sumatra, West Sumatra, and parts of Riau from 2019 to 2023, sales data from 2019 to 2023, day off data from 2019 to 2023, and weather temperature data in Indonesia from 2019 to 2023.
3. Coding Machine Learning
   Machine Learning coding is done according to research needs. The model used in this research is the Recurrent Neural network method. The process that is passed in coding Machine Learning is carried out in several stages, namely: entering the required library, entering data, dividing training data and test data, performing model fit, making predictions, and determining forecasting accuracy.
4. Simulating with Machine Learning
   After making Machine Learning coding, then simulation with Machine Learning is carried out. The model that has been created will be simulated to determine the results of future forecasting and determine the error test value of the designed forecasting model. An error test is carried out in order to determine the accuracy value of a forecasting model. The results obtained at this stage are used to test the Machine Learning forecasting model with the company's forecasting model.

RNN (Recurrent Neural Network) is one of the algorithms that belong to the Deep Learning family. RNN is a development of Artificial Neural Network (ANN), where there is a loop that allows the output of the network to become the input to the next iteration, which then produces a new output. RNN has the ability to maintain an internal memory that allows it to recognize dynamic patterns or patterns. RNN has a looping pattern that allows it to store temporary memory that is used in subsequent data processing. RNN possesses a distinct feature of incorporating at least one feedback loop in its architecture, enabling it to retain information within its network structure. The prediction performance of an RNN is shaped by its weights and architectural configuration. [17]. In addition, the feedback connections in an RNN carry noise information from the previous input that will be accommodated in the next input [6].

Computers can learn from data without being explicitly programmed using predetermined rules. In essence, Long Short-Term Memory (LSTM), a type of Recurrent Neural Network (RNN), introduces long-term and short-term memory mechanisms to address the vanishing gradient problem, and operates based on a set of simple mathematical formulas.

1. Forget Gate

The forget gate decides which parts of the previous cell state should be discarded:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (1)$$

2. Input Gate

It is the input gate that decides which fresh data should be introduced into the cell state:

$$it = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (2)$$

$$C_t = tanh(W_c \cdot [h_{t-1}, X_t] + b_{ic}) \quad (3)$$

3. Cell State Update

The cell state C_t is updated by combining the effects of the forget gate and the input gate:

$$C_t = f_t \cdot C_{t-1} + i_t \cdot C_t \quad (4)$$

4. Output Gate

The portion of the cell state used to compute the hidden state is controlled by the output gate:

$$O_t = \sigma\,(W_o \cdot [h_{t-1}, X_t] + B_o) \quad (5)$$

$$h_t = O_t \cdot Tan\,(C_t) \quad (6)$$

The model that has been made will be simulated to determine the future forecasting results and determine the error test value. The following is the error test formula for the Mean Square Error (MSE) and Root Mean Square Error (RMSE) methods.

$$MSE = \sum \frac{(X_t - F_t)^2}{n} \quad (7)$$

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(P_i - A_i)^2} \quad (8)$$

3. Results and Discussion

The forecasting of product demand is performed to be able to predict future events using past product demand data and projected into the future with a systematic approach model. The existing product demand data is plotted in graphical form to determine the existing data pattern. By knowing the type of data pattern that exists, it can be determined which type of forecasting is in line with the existing demand data. Plotting of demand data is done to see the trend of demand from time-to-time. This helps in understanding whether the demand is increasing, decreasing, or remaining sTable over time. Below is a graph of the sales data pattern for glass bottle tea products from PT XYZ from 2021 to 2023, shown in Figure 1.

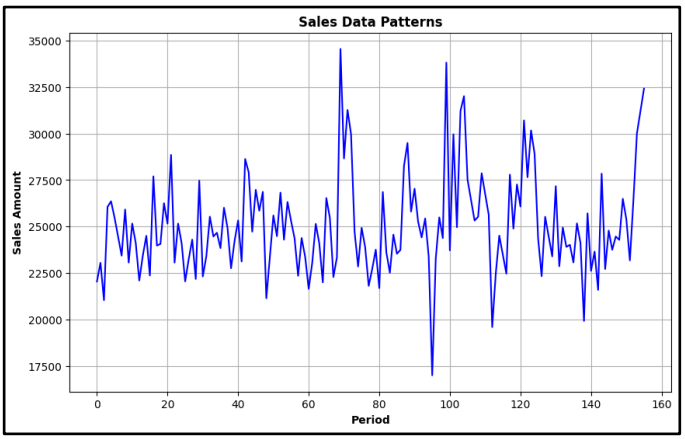

Figure 1. Demand Data Pattern of Glass Bottled Tea Products in 2021-2023.

The weekly demand data, as illustrated in Figure 1, exhibits erratic fluctuations and noticeable swings. It is evident from the figure that the data is non-stationary and displays a cyclical pattern. These large variations in demand contribute to production instability.

The data that is entered in the Machine Learning programme is weekly demand data, calendar day off data, temperature data, population data. The forecasted data is weekly demand. The data used to predict is weekly demand data, calendar day off data, temperature data, and population data for the next period. The Machine Learning process is carried out after all the required data is complete. An example of some of the data that will be input to Machine Learning is shown in Figure 2.

A	B	C	D	E	F
early_Period	final_period	total_days_off	temperature	total_population	total_product
1/1/2021	7/1/2021	2	27.35	14038916	22046
8/1/2021	14/1/2021	1	27.35	14038916	23049
15/1/2021	21/1/2021	1	27.35	14038916	21044
22/1/2021	28/1/2021	1	27.35	14038916	26055
29/1/2021	4/2/2021	1	27.93	14038916	26359
5/2/2021	11/2/2021	1	28.5	14038916	25475
12/2/2021	18/2/2021	2	28.5	14038916	24456
19/2/2021	25/2/2021	1	28.5	14038916	23437
26/2/2021	4/3/2021	1	28.3	14038916	25915

Figure 2. Example of data to be input to Machine Learning

The programming of Machine Learning in this research uses an algorithm that belongs to the Deep Learning family, namely RNN (Recurrent Neural Network). This research uses Google Collab for Machine Learning programming. The following model of forecasting demand of glass bottled tea products with Machine Learning is shown in Figure 3.

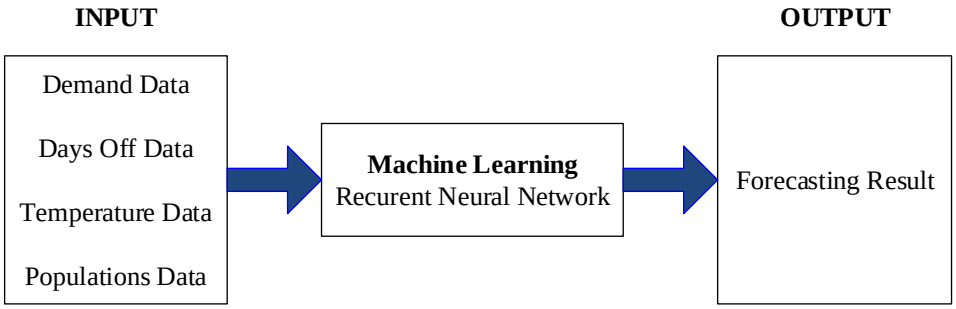

Figure 3. Glass Bottle Tea Product Demand forecasting model with Machine Learning

In this model, we can see the input, process and output in the model used. By utilizing Machine Learning simulations, various insights are obtained, including the outcomes of the Machine Learning forecasting model ChatGPT said:

By utilizing the Recurrent Neural Network method, the accuracy and error test results of the forecasting model are obtained. The outcomes of the Machine Learning forecasting model using the Recurrent Neural Network method are displayed in Figure 4..

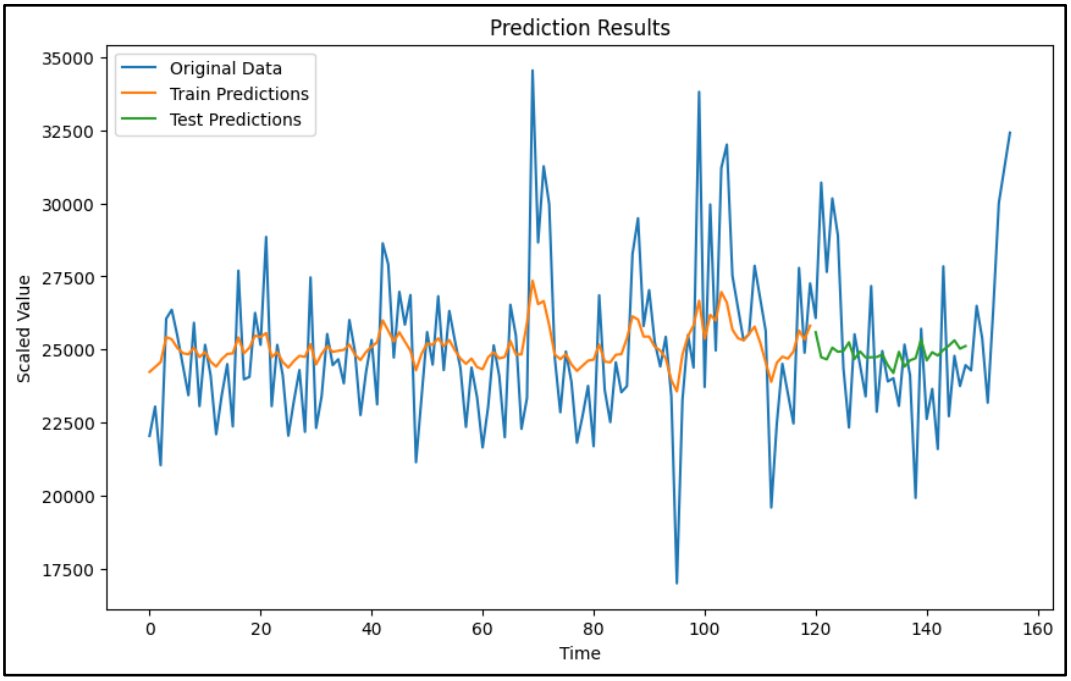

Figure 4. Machine Learning Forecasting Model Results

In Figure 4, the training data is represented in orange, the testing data in green, and the demand realization data in blue. The training data refers to the process of building the model, while the testing data is used to evaluate the performance of the learning model. The model's results are derived from the computational output generated by analyzing and processing the input data within the Machine Learning model. The data inputs include demand data, day off data, temperature data, and population data. In Figure 4 the orange line is the constructed Machine Learning model that runs from the earliest period to the 120th period with 120 data. The green line is the process of testing the performance of the learning model with 28 data.

The forecasting model score accuracy is an evaluation metric used to quantify by how well the forecasting model makes predictions on data that is not seen before. this is done to determine the accuracy value and the forecasting error value. In this study, the accuracy score is calculated from Machine Learning testing data consisting of 28 data. Here is Table 1 of the average accuracy score and model error results.

Table 1. Comparison of Accuracy Score and Model Error

Forecasting Model	Accuracy Score (%)	Error (%)
Recurrent Neural Network	99.47	0.53
Existing Company	84.75	15.25

From the Table 1, it can be seen that the forecasting model using Recurrent Neural Network gets an accuracy score value of 99.47% with an error value of 0.53%. The company's existing model gets an accuracy score value of 84.75% and an error value of 15.25%, which means it is still below the tolerance limit given by the company. A smaller forecasting error value indicates that the forecast created is close to the actual value of the observed data. The error value refers to the difference between the value predicted by the forecasting model and the actual value of the observed data. The smaller the error value, the more accurate the forecasting is presumed to be.

In a Machine Learning simulation, the error test of the demand forecasting model is carried out in order to determine whether the results of the forecasting model during a certain period are too high or too low. The Mean Squared Error (MSE) and Root Mean Squared Error (RMSE) error tests were made using google collab with 28 Machine Learning testing data. The following Table shows a comparison of the Mean Squared Error (MSE) and Root Mean Squared Error (RMSE) values of the Company's existing models and the Machine Learning model.

Table 2. Forecasting Model Error Test Value

Forecasting Model	MSE Value	RMSE Value
Existing Company	21790802.75	4668.06
Recurrent Neural Network	5153408.71	2270.11

Table 2 presents the error test results for both the corporate model and the Machine Learning model utilizing the Recurrent Neural Network (RNN) approach. The corporate model yields an RMSE of 4,668.06 and an MSE of 21,790,802.75. In contrast, the RNN-based machine learning model achieves a lower MSE of 5,153,408.71 and an RMSE of 2,270.11. From these results, the Machine Learning model gets a smaller error value than the company model, which means that the Machine Learning forecasting model Recurrent Neural Network method is better than the company forecasting model. Here is Table 3 of Machine Learning data testing results, company forecasting models and demand realization data.

Table 3. Machine Learning Data Testing Results, Company Forecasting Model and Demand Realization Data

Demand Realization (Pcs)	Machine Learning Model (Pcs)	Company Model (Pcs)
25.523	25.592	27.859
24.460	24.735	26.699
23.396	24.664	25.538
27.170	25.063	32.060
22.869	24.923	27.605
24.948	24.949	30.115
23.908	25.247	28.860
24.019	24.677	29.210
23.072	24.934	28.246
25.169	24.729	30.814
24.121	24.736	29.530
19.926	24.742	24.394
25.709	24.832	32.195
22.624	24.460	28.332
23.653	24.198	29.620
21.596	24.921	27.044
27.843	24.415	34.601
22.719	24.630	28.124
24.784	24.701	30.681
23.751	25.341	29.403
24.464	24.629	29.057
24.286	24.907	27.363
26.493	24.795	29.851
25.389	24.986	28.607
23.182	25.133	26.119

Demand Realization (Pcs)	Machine Learning Model (Pcs)	Company Model (Pcs)
26.411	25.314	28.760
30.012	25.031	32.682
31.212	25.118	33.989

From Table 3, it can be seen the difference in demand realisation data, Machine Learning forecasting model and company forecasting model. The forecasting results of the Machine Learning model are closer to the actual demand realization values.

4. Conclusion

Based on the company's historical data, a forecasting graph is obtained which has a cyclical pattern characterized by a pattern or trend that repeats regularly within a certain time span. Using the forecasting method currently used by the company, there is a difference between demand realization and forecasting results with an average error of 15.25%. The Machine Learning forecasting model of the Recurrent Neural Network method is better than the company's forecasting model. This is indicated by the low test error value of the Machine Learning model compared to the company model. The accuracy value of the Machine Learning forecasting model reaches 99.47% with an error value of 0.53%. With a low error value, it will have a positive impact such as companies can manage their supply chain more efficiently so that companies can reduce inventory costs, avoid shortages or excess stock, and optimize company production and distribution.

References

[1] D. A. Wibowo, "Prediksi Penjualan Obat Herbal Hp Pro Menggunakan Algoritma Neural Network," *Technol. J. Ilm.*, vol. 9, no. 1, p. 33, 2018, doi: 10.31602/tji.v9i1.1100.

[2] M. Layakana and S. Iskandar, "Penerapan Metode Double Moving Average dan Double Eksponential Smoothing dalam Meramalkan Jumlah Produksi Crude Palm Oil (CPO) pada PT. Perkebunan Nusantara IV Unit Dolok Sinumbah," *Karismatika*, vol. 6, no. 1, pp. 44–53, 2020.

[3] E. P. Cynthia, N. Yanti, Y. Fitriani, and M. Yusuf, "Penerapan Metode Elman Recurrent Neural Network (ERNN) Untuk Peramalan Penjualan," *J. Educ. Inform. Technol. Sci.*, vol. 1, no. 2, pp. 49–61, 2019.

[4] A. Roihan, P. A. Sunarya, and A. S. Rafika, "Pemanfaatan Machine Learning dalam Berbagai Bidang : Review paper," vol. 5, no. April, pp. 75–82, 2020.

[5] T. Ma, C. Antoniou, and T. Toledo, "Hybrid machine learning algorithm and statistical time series model for network-wide traffic forecast," *Transp. Res. Part C Emerg. Technol.*, vol. 111, no. December 2019, pp. 352–372, 2020, doi: 10.1016/j.trc.2019.12.022.

[6] S. Aisyah Fany Achmalia , Walid, "BACKPROPAGATION NEURAL NETWORK DAN RECURRENT NEURAL," *UNNES J. Math.*, vol. 9, no. 1, pp. 6–21, 2020.

[7] H. Abbasimehr, M. Shabani, and M. Yousefi, "An optimized model using LSTM network for demand forecasting," *Comput. Ind. Eng.*, vol. 143, no. March, p. 106435, 2020, doi: 10.1016/j.cie.2020.106435.

[8] A. Cecaj, M. Lippi, M. Mamei, and F. Zambonelli, "Comparing deep learning and statistical methods in forecasting crowd distribution from aggregated mobile phone data," *Appl. Sci.*, vol. 10, no. 18, 2020, doi: 10.3390/APP10186580.

[9] R. B. Radite Putra and H. Hendry, "Multivariate Time Series Forecasting pada Penjualan Barang Retail dengan Recurrent Neural Network," *INOVTEK Polbeng - Seri Inform.*, vol. 7, no. 1, p. 71, 2022, doi: 10.35314/isi.v7i1.2398.

[10] L. Wiranda and M. Sadikin, "Penerapan Long Short Term Memory Pada Data Time Series Untuk Memprediksi Penjualan Produk Pt. Metiska Farma," *J. Nas. Pendidik. Tek. Inform.*, vol. 8, no. 3, pp. 184–196, 2019.

[11] J. Wira and G. Putra, *Pengenalan Konsep Pembelajaran Mesin dan Deep Learning*, vol. 4. 2020.

[12] G. N. E. T. S. B. Pardamean, *PENGENALAN KONSEP MACHINE LEARNING*. INSTIPER PRESS, 2022.

[13] Y. S. J. Nasution, "Analisis Pengaruh Jumlah Penduduk Dan Harga Terhadap Permintaan Air Minum PDAM Tirtanadi Cabang Medan," *Kitabah*, vol. 3, no. 1, pp. 82–87, 2019.

[14] W. Ilham, N. Fajri, and K. Cirebon, "Analisis Jumlah Produksi Tahu Wawan Menggunakan Metode Fuzzy Tsukamoto AIDA," *JURSISTEKNI (Jurnal Sist. Inf. dan Teknol. Informasi)*, vol. 5, no. 2, pp. 225–238, 2023.

[15] J. Huber and H. Stuckenschmidt, "Daily retail demand forecasting using machine learning with emphasis on calendric special days," *Int. J. Forecast.*, vol. 36, no. 4, pp. 1420–1438, 2020, doi: 10.1016/j.ijforecast.2020.02.005.

[16] A. Abdurrahim and A. Hartono, "Pengaruh Cuaca Terhadap Perilaku Belanja Konsumen Minimarket: Studi Pada Minimarket Indomaret," *EKOMBIS Rev. J. Ilm. Ekon. dan Bisnis*, vol. 10, no. 2, pp. 1109–1122, 2022, doi: 10.37676/ekombis.v10i2.2649.

[17] Walid, "Peramalan penjualan harga saham PT Bank Rakyat Indonesia (Persero) Tbk (BBRI) dengan Menggunakan Recurrent Neural Network (RNN)," *Prisma*, vol. 2, pp. 139–147, 2019, [Online]. Available: https://journal.unnes.ac.id/sju/index.php/prisma/