

COMPARISON OF LINEAR REGRESSION, NEURAL NET, AND ARIMA METHODS FOR SALES PREDICTION OF INSTRUMENTATION AND CONTROL PRODUCTS IN PT. SARANA INSTRUMENT

Master Maruahal Sidabutar¹

Gerry Firmansyah²

^{1,2}Master of Computer Science Study Program, Faculty of Computer Science, Esa Unggul University Jakarta

*e-mail: masterstar88@student.esaunggul.ac.id, gerry@esaunggul.ac.id

*Correspondence: masterstar88@student.esaunggul.ac.id

Submitted: March 07th 2023, *Revised:* March 23th 2023, *Accepted :* March 30th 2023

Abstract: PT. Sarana Instrument is a national private company that is an authorized sales agent for several instrumentation and control products originating from European countries and the United States for sales in Indonesia. Every company certainly targets sales to be achieved every year, for that the company certainly needs sales forecasting. PT. Sarana Instrument does not currently have a prediction system so for making annual sales targets, it still uses manual estimates by looking at sales data from the previous year's sales. So that PT. Sarana Instrument cannot get accurate sales predictions and the company cannot prepare human resources and financial resources according to the company's needs. Therefore, a forecasting system is needed to help make forecasts. The purpose of this study was to analyze the error rate of forecasting sales data for 2013-2021 at PT. Sarana Instrument uses a forecasting algorithm, namely the Linear Regression Algorithm, Neural Net and Arima, to obtain a sales forecast method with the smallest error rate and can be implemented at PT. Sarana Instrument. The results showed that (1) forecasting using the Linear Regression method resulted in an RMSE value of 12,268 for Total Sales, a RMSE value of 10,771,126 for GM-MSA Product Sales, an RMSE value of 775,499 for SOR Product Sales. (2) Forecasting with the Neural Net method produces a RMSE value of 5,976 for Total Sales, an RMSE value of 5045,773 for GM-MSA Product Sales, a RMSE value of 10,830,255 for SOR Product Sales. (3) Forecasting with the Arima method produces an RMSE value of 12,154 for Total Sales, a RMSE value of 5045,773 for GM-MSA Product Sales, and a RMSE value of 445,273 for SOR Product Sales. So, a comparison of forecasting is more appropriate using the Arima method, because it produces the smallest error value.

Keywords: Forecasting; Sales; Linear Regression; Neural Net; Arima; RMSE.

INTRODUCTION

Every company certainly targets sales to be achieved every year, and for that, the company certainly needs sales forecasting. Forecasting future sales means determining the estimated size of sales volume, even determining the sales potential and market area controlled in the future (R. M. Sari, 2022). Such forecasting greatly influences the manager's decision to determine the amount of production of goods that the company must provide (Danil & Sukanta, 2022). The activity of predicting the future is one of the company's efforts in making effective and efficient decisions in business continuity.

PT. Sarana Instrument is a national private company that is the official sales agent of several instrumentation and control products from European countries and the United States for sales areas in Indonesia. PT. Sarana Instrument currently does not have a prediction system (forecasting) so for making sales targets every year still uses manual estimates by looking at sales data from the previous year's sales. So that PT. Sarana Instrument cannot get accurate sales forecasts and the company cannot prepare human resources and financial resources that are by the needs of the company. Therefore, a forecast system is needed that is easy to use and helps make sales forecasts.

The purpose of this study is to analyze the error value in forecasting sales data at PT. Sarana Instrument uses the sales forecasting algorithm in the machine learning application, namely Rapid Miner. The methodology used in this study used 3 forecasting algorithms, namely the Linear

Regression, Neural Net, and Arima Algorithms. The analysis uses sales data for 2013 - 2021. From the results of the comparison of the accuracy of the three prediction methods (forecast), this selling will get a sales forecast method with the smallest error value and can be implemented in PT. Sarana Instrument.

MATERIALS AND METHODS

Sales

Sales is an integrated effort to develop strategic plans that are directed at satisfying the needs and desires of buyers to get profit-making sales. Sales are the lifeblood of a company because the company can get profit and an effort to lure consumers who are seeking to know their attractiveness so that they can know the results of the products produced. Sales are the amount that the Buyer charges for the merchandise submitted constituting the income of the company (Indarwati et al., 2019).

Forecasting

Forecasting is the art and science of predicting future events by involving taking historical data and projecting it into the future with a systematic approach model (Indarwati et al., 2019). As for the stages of forecasting in summary, three stages must be passed in the design of a forecasting method, namely:

1. Analyze past data. This step aims to get a picture of the pattern from the data in question.
 2. Choose the method to use. There are a variety of methods available for its use. The choice of method can affect the result of the forecast. The forecast result is measured
-

by calculating the smallest error or error. Therefore, no forecasting method is good for any type of data.

3. The process of transformation of past data using the chosen method. If necessary, changes are made as needed.

Data Mining

Data mining is data that is stored for a long time then the data is collected, it turns out that after analyzing the data has a certain pattern. Data mining usually contains statistics, economics, weather forecasts as well as various other types of work-related data. Data mining is a lesson in what has happened in the past and then will be applied in the future to obtain better results (Wibowo, 2018). Rapidminer is a tool or software for managing data mining. This application is one of the solutions for carrying out data mining analysis, forecasting analysis, and text mining. This application uses a variety of predictive and descriptive techniques to convey knowledge to the user until it can produce decisions that match what is expected. (Persediaan et al., 2022).

Linear Regression Algorithm

Linear Regression is a statistical analysis that studies the relationship between two or more quantitative variables so that one variable can be predicted from another. The relationship between two variables can be distinguished into two, namely the functional relationship and the statistical relationship (Tengah, 2022). A simple linear regression equation can be formulated as follows (Dewi Putri et al., 2022) :

$$Y = a + bx$$

Information:

y = Variable prediction or non-free

(dependent variable)

x = Variable predictor or free (independent variable)

a = constant (intercept)

b = parameter of free variable regression coefficient

Neural Net Algorithm

The use of information technology that is increasingly advanced in today's era Artificial Neural networks or artificial neural networks can be used to predict sales. Neural networks are of artificial representations of the human brain that always try to simulate the learning process in the human brain. The term artificial here is used because this neural network is implemented using a computer program that can complete several calculation processes (Tiara, 2018).

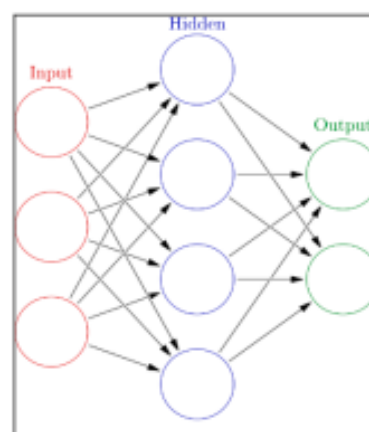


Figure 1. Neural Network Model

Arima algorithm

The ARIMA (Autoregressive Integrated Moving Average) forecasting method is a very appropriate method to overcome the complexity of time series and other forecasting situations. The ARIMA method or commonly referred to as the Box-Jenkins method was first developed by George Box and Gwilym Jenkins in 1975 is a forecasting

method that is different from the existing forecasting method, because:

- This method does not assume a specific data pattern to forecast.
 - This method identifies a model that allows it to be used, and then the existing data is matched to the predefined model.
- The model in the ARIMA method consists of 3 models, namely autoregressive, moving average, and autoregressive and moving average. Where the autoregressive (AR) model becomes the p-order, the moving average (MA) becomes the q-order, and the autoregressive and moving average (ARMA) is a mixture of the first two models (Sutanto et al., 2017).

Rapid Miner

RapidMiner is a software that aims to analyze data mining, text mining, and prediction analysis that is open source and uses various descriptive and predictive techniques in providing insights to users so that they can make the best decisions. RapidMiner is written using the java language so that it can work on all operating systems. RapidMiner has approximately 500 data mining operators, including operators for input, output, data preprocessing, and visualization. RapidMiner is a stand-alone software for data analysis as a data mining engine that can be integrated into its products (Karsito & Monika Sari, 2018)

Testing of Forecasting Methods

RMSE (*Root Mean Square Error*) Is the value of the magnitude of an error rate in the prediction. The smaller the RMSE value, the more accurate the resulting prediction will be (Uang, 2022)

$$RMSE = \sqrt{\frac{\sum (Y' - Y)^2}{n}}$$

Y ' = Predicted Value

Y = True Value

n = Total Data

MAE (*Mean Absolute Error*)

Shows the average predicted error value of its actual value. MAE is often used to calculate predictions that are errors (Uang, 2022)

$$MAE = \sum \frac{|y' - y|}{n}$$

Where:

Y ' = Predicted value

Y = True Value

n = Total Data

MAPE (*Mean Absolute Percentage Error*)

MAPE is calculated as the mean of absolute differentiation between the forecasted and actual values, expressed as a percentage of the actual value. If it has a forecasted and actual value for n periods. The smaller the MAPE value, the more accurate the results or forecasting techniques performed and vice versa (Hernadewita et al., 2020).

$$MAPE = \left(\frac{100}{n}\right) \sum |A_t - \frac{F_t}{A_t}|$$

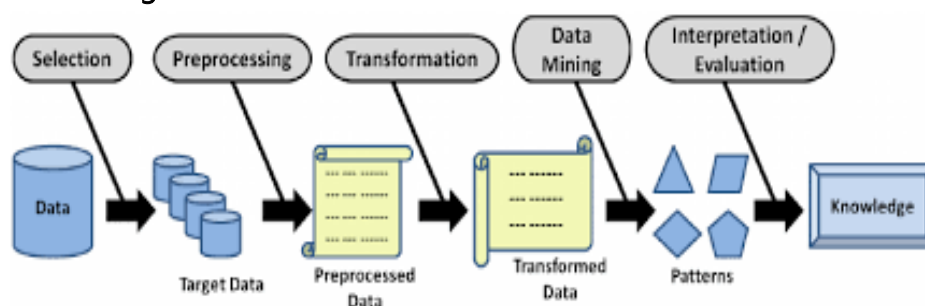
Where:

F_t= Forecast value for the t-th period

A_t= Actual value in the t-th period

n= The amount of forecasting result data

Stages of Forecasting



(Source: Data Mining as a Business Solution, Veronika S.Moertini, 2017)

Here are the Stages of Forecasting in Data Mining (Moertini, 2017):

a. Selection

In most cases, only a portion of the data can be extracted from the database used for analysis. This is because the data taken from the database must comply with the analysis standard of the database to be collected

b. Preprocessing

One of the stages in mining data. Before heading to the processing stage. The raw data will be processed first. Data Preprocessing or preprocessing data is usually done through the elimination of inappropriate data.

c. Transformation

The data mining process is initiated by entering data or entering it into the appropriate format for the process. Sebelum can be used, many data mining techniques create specific data formats.

d. Data Mining

Data mining is the process of finding interesting patterns or information in selected data using certain techniques or methods. Techniques, methods, or algorithms in data mining vary widely. The selection of the right method or algorithm depends largely on the overall

KDD goals and process

e. Interpretation

The pattern of information generated from the data mining process needs to be displayed in a form that is easily understood by interested parties. This stage is part of the KDD process called interpretation. This stage includes the examination of whether the patterns or information found contradict previously existing facts or hypotheses.

2.10 Data Collection

In this study, the data used was secondary data obtained, namely sales data from PT. Sarana Instrument from 2013 – 2021. The data taken was 3,204 data consisting of 20 attributes.

RESULTS AND DISCUSSION

Data selection

From this data, the data used was 108 data consisting of 2 attributes, namely Date, Total Sales and Total Sales based on Product Type (GM-MSA and SOR), where the average total sales were taken each year. The data that has been selected attributes will be used as dataset.

Preprocessing

The dataset above to be tested is divided into two parts, namely Data Training and Data Testing then analyzed

using Rapidminer software. Sales dataset from PT. Sarana Instrument from 2013 - 2021 has 108 records, for training data has

84 data records (in 2013-2019) and testing data has 24 data records (in 2020-2021).

Transformation

The details of the attributes that will be used in the mining process in the table and the results are as follows:

Table 1. Attributes in the mining process

No	Attribute Name	Data Type	Types of Role Data
1.	Number	Integer	id
2.	Date (2013-2019)	Date	regular
	Date (2020-2021)	Date	regular
3.	Total Sales (Billion)	Real/Integer	label
	Total Sales Based on GM-MSA Products (Million)	Real /Integer	label
	Total Sales By SOR Product (Million)	Real/ Integer	label

Data Mining

a. Data Mining Implementation of Linear Regression Algorithm

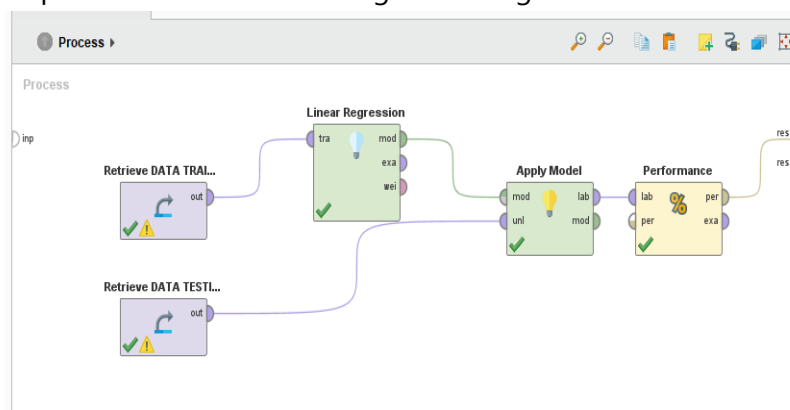


Figure 2. Data Mining Implementation for Regression Linaer Algorithm

b. Data Mining Implementation of Neural Net Algorithm

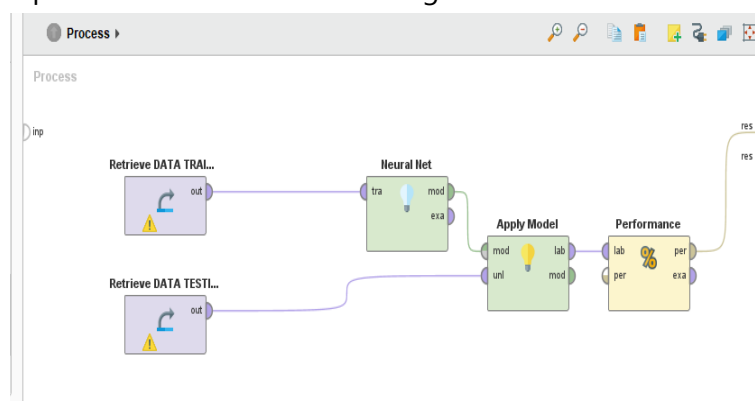


Figure 3 Data Mining Implementation for Neural Net Algorithms

c. Arima Algorithm Data Mining Implementation

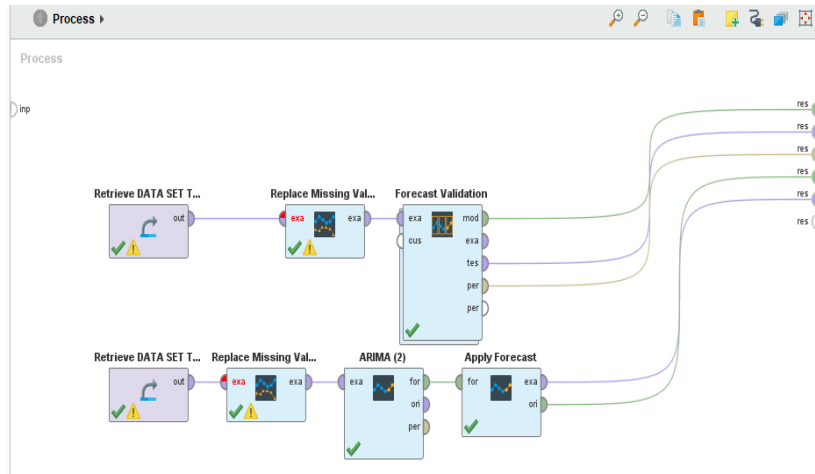


Figure 4 Data Mining Implementation for Arima Algorithm

Forecasting Results

a. Linear Regression Algorithm

Forecasting Results

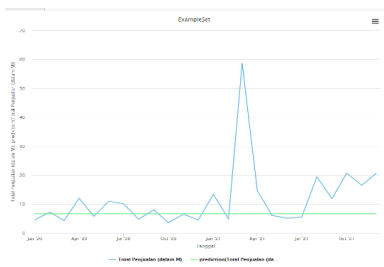


Figure 5. Graph of Total Sales Forecasting Results Using Regression Algorithms

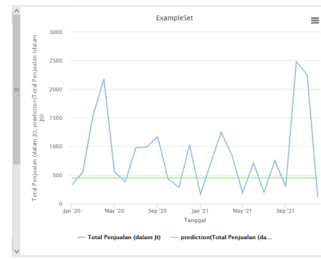


Figure 7. Graph of Forecasting Total Sales of SOR Products Using Regression Algorithms

b. Neural Net Algorithm Forecasting Result

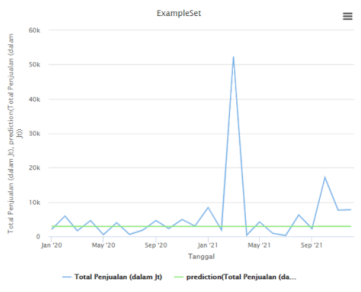


Figure 6. Graph of Forecasting Total GM-MSA Product Sales Results Using Regression Algorithms

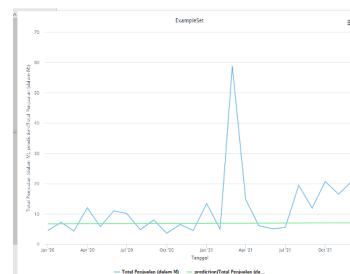


Figure 8. Graph of Forecasting Total GM-MSA Product Sales Results Using Neural Net Algorithms

c. Arima Algorithm Forecasting Results

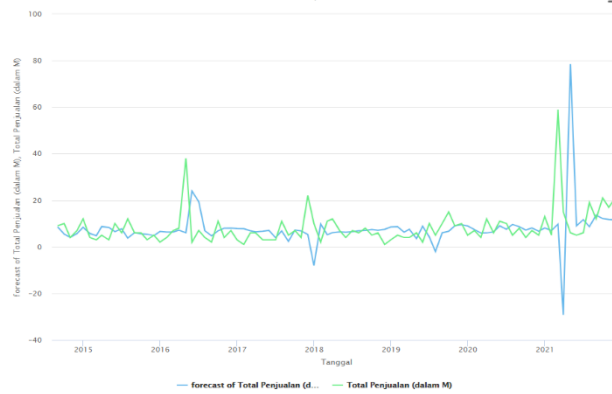


Figure 9. Graph the results of forecasting total sales using the Arima Algorithm

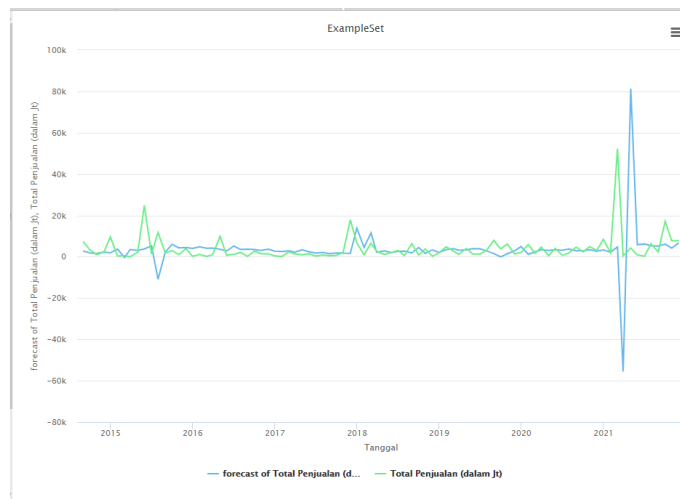


Figure 10. Graph of Forecasting Total GM-MSA Product Sales Results Using the Arima Algorithm



Figure 11. Graph of Forecasting Total Sales of SOR Products Using the Arima Algorithm

Table 2. Performance test results of RMSE value on data Total Sales 2013-2021

Model arima	RMSE Results
ARIMA(1.0.1)	206055875257343550
ARIMA(1.0.2)	86879788578815540
ARIMA(2.0.1)	5.976

Table 3. Performance test results of RMSE value on data GM-MSA product sales 2013-2021

Model arima	RMSE Results
ARIMA(1.0.1)	206055875257343550
ARIMA(1.0.2)	86879788578815540
ARIMA(2.0.1)	5045.773

Table 4. Performance test results of RMSE value on data SOR product sales 2013-2021

Model arima	RMSE Results
ARIMA(1.0.1)	450.352
ARIMA(1.0.2)	103215241640196920
ARIMA(2.0.1)	445.273

From these results, it can be concluded that ARIMA (2.0.1) with a value of RMSE 5,976 for total sales data, RMSE 5045,773 for GM-MSA product sales data and RMSE

445,273 for SOR product sales data, has a smaller value compared to other arima models. so that the model will be used until the stage of the forecasting process.

Testing and Evaluation

a. Results of Testing and Evaluation Using Linear Regression Algorithms

Table 5. Testing Results & Evaluation of Total Sales Using Linear Regression Algorithms

Test Results	Value
RMSE	12.268
MAE	6.480

Table 6. Testing Results & Evaluation of Total Sales GM-MS Products Use Linear Regression Algorithms

Test Results	Value
RMSE	10771.126
MAE	4600.829

Table 7. Testing Results & Evaluation of Total Sales SORM Products Use Linear Regression Algorithms

Test Results	Value
RMSE	775.499

MAE	6.440
-----	-------

b. Testing and Evaluation Results Using Neural Net Algorithms

Table 8. Testing Results & Evaluation of Total Sales Using Neural Net Algorithms

Test Results	Value
RMSE	12.154
MAE	6.440

Table 9. Testing Results & Evaluation of Total Sales GM-MS Products Use Neural Net Algorithm

Test Results	Value
RMSE	10830.255
MAE	4612.229

Table 10. Testing Results & Evaluation of Total Sales SOR Products Using Neural Net Algorithms

Test Results	Value
RMSE	782.176
MAE	542.175

c. Arima Algorithm Testing and Evaluation Results

Table 11. Arima algorithm testing and evaluation results for data Total Sales for 2013-2021

Test Results	Value
RMSE	5.976
MAE	150.310

Table 12. Arima algorithm testing and evaluation results for data GM-MSA product sales in 2013-2021

Test Results	Value
RMSE	5045.773
MAE	152171318.538

Table 13. Arima algorithm testing and evaluation results for data SOR product sales in 2013-2021

Test Results	Value
RMSE	445.273
MAE	404502.051

Comparison Results of Error to Forecasting 3 Forecasting algorithms (Linear Regression, Neural Net and Arima)

Table 14. Results of comparison accuracy of 3 forecasting algorithms for Total Sales of PT. SI Year 2013-2021

Algorithm	Value
Linear Regression	RMSE = 12.268
Neural net	RMSE = 12.154
Arima	RMSE = 5.976

Table 15. Results of comparison accuracy of 3 forecasting algorithms for GM-MSA PT. SI Year 2013-2021

Algorithm	Value
Linear Regression	RMSE = 10771.126
Neural net	RMSE = 10830.255
Arima	RMSE = 5045.773

Table 16. Results of comparison accuracy of 3 forecasting algorithms for the Sale of SOR PT. SI Year 2013-2021

Algorithm	Value
Linear Regression	RMSE = 775.499
Neural net	RMSE = 782.176
Arima	RMSE = 445.273

CONCLUSIONS

Based on the results of the analysis, it can be concluded that: 1). Application of data mining to forecast *PT. Saran Instrument* helps in making decisions on the sale and inventory of products so that companies can take important steps in dealing with them. 2). The Arima forecast algorithm is more suitable for forecasting (predicting) the total sales of *PT. Arima Sarana Instrument*, *GM-MSA product sales*, and *SOR product sellers* are compared with linear regression and neural net algorithms. because the RMSE value of the Arima forecast algorithm is smaller compared to linear regression and neural net algorithms.

REFERENCES

- Danil, J. A., & Sukanta, S. (2022). Penerapan Metode Moving Average dan Exponential Smoothing pada Peramalan Produksi di PT. Mada Wikri Tunggal. *Jurnal Ilmiah Wahana Pendidikan*, 8(7), 37–40.
- Dewi Putri, R., Ilmu Komputer, F., Bina Darma, U., Jenderal Ahmad Yani No, J., Seberang Ulu, K. I., Palembang, K., & Selatan, S. (2022). 2022) 2254-2263 Accredited. *Jurnal Mantik*, 6(2), 2254–2263.
- Hernadewita, Hadi, Y. K., Syaputra, M. J., & Setiawan, D. (2020). Peramalan Penjualan Obat Generik Melalui Time
-

- Series Forecasting Model Pada Perusahaan Farmasi di Tangerang: Studi Kasus. *Journal Industrial Engineering & Management Research (Jiemar)*, 1(2), 35–49.
- Indarwati, T., Irawati, T., & Rimawati, E. (2019). Penggunaan Metode Linear Regression Untuk Prediksi Penjualan Smartphone. *Jurnal Teknologi Informasi Dan Komunikasi (TIKOMSiN)*, 6(2), 2–7. <https://doi.org/10.30646/tikomsin.v6i2.369>
- Karsito, & Monika Sari, W. (2018). Prediksi Potensi Penjualan Produk Delifrance Dengan Metode Naive Bayes Di Pt. Pangan Lestari. *Jurnal Teknologi Pelita Bangsa*, 9(1), 67–78.
- Moertini, V. (2017). Data Mining Sebagai Solusi Bisnis. *Data Mining Sebagai Solusi Bisnis*, 7(1), 44–56.
- Persediaan, D. A. N., Jumputan, K., Toko, P., & Colet, B. (2022). *Jurnal Mantik Penerapan Naive Bayes Dalam Memprediksi Penjualan Tuan Kentang Palembang*. 6(36), 2502–2507.
- R. M. Sari. (2022). "Implementation of Data Mining to Predict Sales Using Preliminary Research Methods of Data Collection Analysis and Design of Test Implementation. *Senashtek*, 74–82.
- Sutanto, P., Setiawan, A., & Setiabudi, D. H. (2017). Perancangan Sistem Forecasting di Perusahaan Kayu UD. 3G dengan Metode ARIMA. *Jurnal Infra*, 05(01), 325–330.
- Tengah, L. (2022). *Implementasi Forecasting Pada Perencanaan Sistem Pemesanan Buku Lks (Lembar Kerja Siswa)* Menggunakan Algoritma Regresi Linear. (Studi Kasus : Toko Buku Darul Ulum ,. 4(1), 1–11.
- Tiara, B. (2018). Prediksi Penjualan Menggunakan Algoritma Neural Network: Studi Kasus di PT. Balaraja Food Makmur Abadi. *Ipsikom*, 6(1).
- Uang, M. (2022). *Komparasi Algoritma Linear Regression dan Neural Network Untuk Memprediksi Nilai KURS*. 10(2), 64–71.
- Wibowo, D. A. (2018). Prediksi Penjualan Obat Herbal Hp Pro Menggunakan Algoritma Neural Network. *Technologia: Jurnal Ilmiah*, 9 (1), 33. <https://doi.org/10.31602/tji.v9i1.1100>



© 2023 by the authors. Submitted for possible open-access publication

under the terms and conditions of the Creative Commons Attribution (CC BY SA) license (<https://creativecommons.org/licenses/by-sa/4.0/>).
