

# Utilizing Linear Regression for Predicting Sales of Top-Performing Products

<sup>1</sup>Matthew Avialdo Pratama and <sup>2</sup>Wowon Priatna

<sup>1,2</sup>Faculty of Computer Science, Bhayangkara University Jakarta Raya, Jakarta, INDONESIA

e-mail : <sup>1</sup>matthewavialdo@gmail.com, <sup>2</sup>wowon.priatna@dsn.ubharajaya.ac.id

**Publisher's Note:** JPPM stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2023 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Corresponding Author:** Wowon Priatna

## Abstract

PT Ajidarma Delta Medika is a company engaged in the sale of medical devices in the city of Bekasi. This company markets a variety of medical device products. Judging from the large number of consumer requests for medical device products based on sales data for the last 3 years, predictions are needed for the best-selling product sales, in order to facilitate the company in planning the supply of stock. To find out the best-selling medical device product sales, data prediction techniques are used with the Linear Regression algorithm. By using the Linear Regression algorithm, the results are obtained to predict the best-selling sales of several products sold at PT Ajidarma Delta Medika. This research produces an accuracy value with the MAPE formula for predicting the best-selling product sales of 14.2%. This shows that the linear regression method is good at predicting sales of medical devices in the following year.

**Keywords** — Prediction, Linear Regression, MAPE, Medical Devices

## 1 Introduction

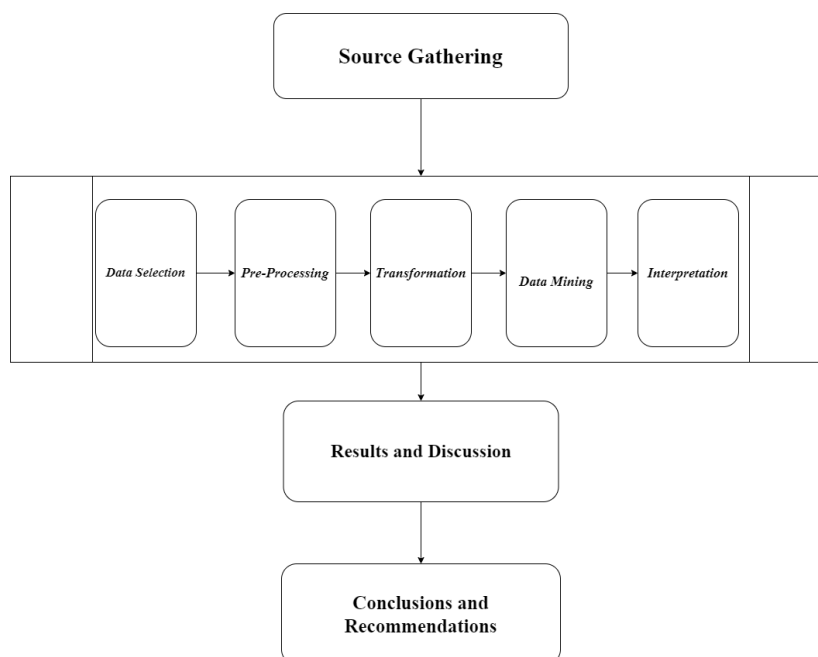
Prediction is a method of systematically estimating future events based on past and present information designed to minimize error (the difference between what happened in the past and what is expected)[1]. With historical data as a guide, forecasting can be done. Predictions are commonly used by company owners to facilitate decision making in determining the number of goods that must be provided by the company. Prediction can be used to determine the best-selling products in the company."[2]

PT.Ajidarma Delta Medika is a company in the city of Bekasi that is engaged in providing medical devices for hospitals using a sales system that supports company performance and sales services. This company does not yet have a calculation method to calculate consumer demand each year. Judging from this, the company needs a calculation method for prediction for the sale of medical devices that are most desired by consumers. Prediction can be used to find information from large amounts of data so that data mining is needed. "Data mining is the science that solves the problem of retrieving information from large databases by combining techniques from statistics, machine learning, data visualization, pattern recognition, and databases." Data mining is also used to extract crowd-sourced information so that this information can be used to measure markets[3].

In data mining, methods included in prediction include the linear regression method, this method is a suitable solution for companies that produce different products because by comparing the number of different products, companies can increase profits while still considering the right amount of production. However, applying linear regression requires knowledge of the right product mix, target market, and the amount of product demand[4].

In addition, the linear regression method also has the advantage of producing the most accurate prediction model for linear data. and Easy to evaluate and has various methods or matrices that are easy to apply[5]. Therefore the purpose of this research is to predict the sales of the best-selling products at PT Ajidarma Delta Medika.

## 2 Research methods



**Figure 1** Research method

In this process section, several stages will be explained during the implementation of the research. These processes are described in the form of Figure.1

- 1) Data collection. This process is carried out by interviewing the accounting department, observation and documents.
- 2) Data mining processing. The data that has been collected will be processed in accordance with the stages of Knowledge Discovery in Database (KDD). KDD is an interdisciplinary program that refers to the process of obtaining knowledge from data[6]. to produce research results in accordance with a predetermined sequence, the following stages are Knowledge Discovery in Database (KDD)[7].
- 3) Data Selection Data selection from operational data must be done before the start of the information analysis stage with KDD. Data that has undergone the selection process will be processed using data mining, stored in a file, separate from the operational database. The data used in this research comes from PT Ajidarma Delta Medika, namely sales data for medical device products in 2020, 2021 and 2022. All attributes in the sales data include, Item name, Quantity and Month. From all existing data, 3 fields will be selected and used for the knowledge discovery in database (KDD) process. These fields are:
  - 1) Item Name is an attribute contained in the sales data table which contains information about the name of the goods marketed at PT Ajidarma Delta Medika.
  - 2) Amount is an attribute contained in the sales table that contains information on the amount of product sales.
  - 3) Month is an attribute contained in the sales data table that contains product sales month information.
- 4) Preprocessing Before the data mining process can be carried out, a preprocessing stage must be carried out, at this stage a data integration process will be carried out to combine data from different databases[8], then data cleaning is carried out to produce a clean dataset so that it can be used for the next stage, namely mining. The following is an explanation of the two processes:
  - a) Data Integration. This step is the process of combining data from different databases so that the data is interconnected. Data integration is done on attributes that are identified with specific entities [9].
  - b) Data Cleaning. At this stage, unnecessary data, empty values, and raudants will be removed. This is because relevant data, not missing values, and not radudant are the initial requirements in doing data mining. A data is said to be missing value in a dataset there are attributes that do not have a value or null, while data is said to be radudant if in one dataset it has the same value [18], after cleaning the data that meets more requirements based on sales data
- 5) Transformation  
The Transformation stage is a step to modify the data that has been selected so that the data is suitable for the data mining process[10].

The transformation process in KDD is a creative process and depends heavily on the type or pattern of information to be sought in the database. From this stage, all operational data is obtained, which has been grouped into attributes that will be used for the data mining transformation process, namely the attribute of the number of sales per year which is the reference in the data mining process.

#### 6) Data Mining

This stage is the process of finding interesting patterns or information on selected data using certain algorithms based on the overall KDD process. the algorithm used in this research is the Linear Regression algorithm. Here is the formula for Linear Regression:

$$a = \frac{(\sum y)(\sum x^2) - (\sum x)(\sum yx)}{n(\sum x^2) - (\sum x)^2} \quad (1)$$

$$b = \frac{n(\sum xy) - (\sum x)(\sum y)}{n(\sum x^2) - (\sum x)^2} \quad (2)$$

$$y = a + bx \quad (3)$$

Description

Y: Dependent Variable

x: Free

Variabel

a: Constant that shows the value of y if x = 0

b: magnitude of change in y value

The steps that need to be taken to perform linear regression calculations are as follows:

- a) Calculating  $X^2$ ,  $XY$  and the total of each of them.
- b) Calculate a and b using the predetermined formula
- c) Modeling the Regression Line Equation
- d) Make predictions on predictor or response variables

#### 7) Interpretation/Evaluation

The information patterns generated from the data mining process need to be presented in a format that is easily understood by interested parties. this step includes checking whether the patterns or information found conflict with existing facts or assumptions. At this stage, the sales pattern of medical devices is obtained from the data mining process using the linear regression method, the information generated from the data mining process is in the form of rules obtained from linear regression calculations.

The evaluation stage uses accuracy measurement with the Mean Absolute Percentage Error formula. (MAPE)[11]. Where MAPE indicates how much error in forecasting compared to the real value in regression. The MAPE method can be calculated using the following formula:

$$MAPE = \frac{1}{n} \sum_{t=1}^n \frac{|Y_t - Y_t^1|}{Y_t} \quad (4)$$

Where:

$x_t$  = actual value

$y_t$  = predicted value

n = amount of data

#### 8) Results and Discussion.

The discussion at this stage explains the results of the data mining process carried out using the linear regression method.

#### 9) Conclusions and suggestions.

Draw conclusions from the research results and provide suggestions for the company so that it can be even better.

### 3 Results and Discussion

In this section, the results obtained after carrying out the stages described in the methodology will be displayed. From the data collection stage, a sales report for medical devices at PT Ajidarma Delta Medika was obtained during the period January 2020 to December 2022 or within 3 years consisting of product names and the number sold each month.

Sales data that has been collected and selected into 4 attributes, namely item name, quantity and month at the data selection stage.

The next stage is preprocessing where the data that has previously passed the data selection stage will be processed data cleaning and data integration. In the data cleaning process, data that does not have attributes is eliminated. And in the data integration process, the annual sales data is merged to simplify the calculation process. After the preprocessing stage, the transformation stage is carried out. Where the transformation stage is carried out the data reduction process. So that the data is easier to process and the data produces a clear picture, the reduction carried out in this study is in the form of changing the quantity data from per month to per year. To find out the best-selling products, twenty-five medical devices marketed by PT.Ajidarma Delta Medika were tested. Of the twenty-five products, the prediction calculations will be loaded from those marketed by PT.Ajidarma Delta Medika specifically in the data mining stage with the linear regression algorithm.

a. ALT/GPT

In ALT/GPT products, the data that has undergone the preprocessing stage is then determined X, Y, XY and XX with values that can be seen in table 1.

In Table.1, the values in column X are taken from the numeric "Period". Column Y is taken from the sales quantity of ALT/GPT products. Column XY is the product of the contents of columns X and Y and column XX contains the square of the contents of column X.

The X, Y, XY, and XX values that have been obtained are then used to calculate the values of a and b. These values are the coefficients used in forming the regression equation model which will then be used to carry out the prediction stage.

**Table 1.** ALT/GPT Product Calculation Data

| ALT/GPT      | X | Y   | XY  | XX |
|--------------|---|-----|-----|----|
| 2020         | 1 | 44  | 44  | 1  |
| 2021         | 2 | 32  | 64  | 4  |
| 2022         | 3 | 38  | 114 | 9  |
| <b>TOTAL</b> | 6 | 114 | 222 | 14 |

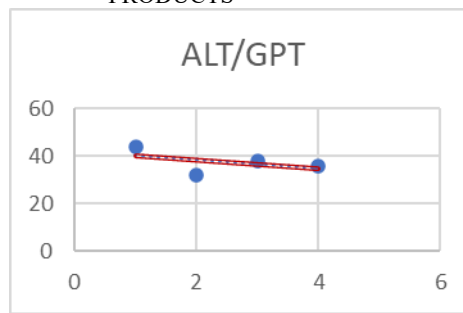
In table.2 in the calculation of finding the constant value b with formula (2) the result is 0.069 and the constant value a is 38.14 which is calculated using formula (1). After obtaining the results of constants a and b, the calculation is carried out to determine the regression or Y equation which is done by formula (3). In calculating the regression on ALT/GPT products, the equation  $Y = 38.13953 + 0.069(X)$  is obtained.

To test the accuracy of prediction results on ALT / GPT products, it is necessary to calculate the Mean Absolute Percentage Error (MAPE). Where the smaller the deviation between the prediction results and the actual conditions shows that the prediction method used is good. MAPE calculation on ALT / GPT products with formula (4) obtained a result of 7.9%. This shows that linear regression prediction for ALT/GPT products is a very good prediction.

**Table 2.** ALT/GPT product calculation results

| Konstanta                                   | Nilai |
|---------------------------------------------|-------|
| <i>a</i>                                    | 38.14 |
| <i>b</i>                                    | 0.069 |
| <b>Persamaan Y</b>                          |       |
| <b><math>Y = 38.13953 + 0.069(X)</math></b> |       |

Figure 2 shows the original value of ALT/GPT product sales and the prediction results from calculations using the linear regression method. The difference between the original value and the prediction which is not far away is one indication if the prediction using linear regression for ALT/GPT product sales data is considered accurate.



**Figure 2** ALT/GPT Product Graphics

Where the Y equation obtained can be used to calculate the estimated sales of ALT/GPT products in the future, for example to find out the sales of ALT/GPT products in 2023 using the Y equation obtained, will get a result of 35 pieces.

The results of the prediction of sales of medical devices at PT Ajidarma Delta Medika in 2023 can be seen in the following table 3

**Table 3** Prediction Results and MAPE of Medical Device Products PT. Ajidarma Delta Medika

| Nama Barang                 | Prediksi | MAPE |
|-----------------------------|----------|------|
| ALT/GPT                     | 36       | 7,9% |
| AST/GOT                     | 35       | 6,3% |
| Clean Kit                   | 176      | 5,4% |
| Diluent Dymind              | 40       | 0,0% |
| Diluent M30                 | 54       | 21%  |
| Gas O2                      | 71       | 18,% |
| HbsAg Strip                 | 250      | 4,6% |
| I-chroma Hba1c Isi 25 Test  | 162      | 8,0% |
| Lyse                        | 29       | 45%  |
| Reagen Cholesterol          | 212      | 8,7% |
| Reagen Gula                 | 261      | 1,9% |
| Reagen Wright               | 117      | 19%  |
| Roller Mixer Oregon KJMR-II | 15       | 25%  |
| Sputum Terumo 3CC           | 17       | 32%  |
| Strip Gula darah            | 268      | 4,3% |
| Tabung Clot Activator       | 35       | 0%   |
| Tabung EDTA 3MI             | 20       | 20%  |
| Tabung LED/ESR              | 17       | 41%  |
| Triglyceride                | 16       | 6,7% |
| Tubex TF                    | 24       | 20%  |
| Urea Uv                     | 32       | 5,9% |
| Urid Acid                   | 38       | 5,0% |
| Wash Pack M348              | 79       | 1,3% |
| Widal                       | 43       | 2,3% |
| Yellow tipe                 | 5        | 44%  |

Where the calculation to calculate the overall accuracy of the prediction uses the MAPE formula (4) as follows:

$$MAPE = \sum_{t=1}^n \frac{354,1120\%}{25}$$

$$MAPE = 14,1645\%$$

Which can be interpreted as showing good prediction results because it is in the range of 10%-20%. The smaller the percentage error in MAPE, the more accurate the forecasting results will be. Some analysis states that the variation in the Mean Absolute Percentage Error value has different meanings.

- If the MAPE value is less than 10%, the forecasting model's ability is very good.
- If the MAPE value is between 10% - 20% then the ability of the forecasting model is good
- If the MAPE value is in the range of 20% - 50%, the ability of the forecasting model is feasible.
- If the MAPE value ranges more than 50%, the ability of the forecasting model is poor.

From this value, we can understand that the MAPE value can still be used if it does not exceed 50%. When the MAPE value is above 50%, the forecasting model can no longer be used[12].

#### 4 Conclusion

Based on research that has been conducted at PT Ajidarma Delta Medika. It can be concluded that the results of the calculation of sales predictions of the best-selling products at PT.Ajidarma Delra Medika are blood sugar strips with predicted sales of 268 pieces, sugar reagents with predicted sales of 261 pieces, and HbsAg strips with sales of 250 pieces. Meanwhile, the product with the least sales prediction is Yellow Type with predicted sales of only 5 pieces.

Predictions using the linear regression algorithm can be said to be in a good category. This is based on the accuracy test with the MAPE formula which gets a total accuracy value of 14.1645%. Where it shows that the prediction results are good because they are in the range of 10% - 20%.

#### 5 Suggestion

Some things that can be developed for further research with the same scope are as follows:

- For further research, other forecasting methods can be used to compare the reliability of forecasting results for each method.
- In future research, it can be developed with the support of a system that can assist in calculating predictions. So that it is easier for companies to know the highest and lowest sales levels to be utilized in supporting the provision of medical device products in the company.

#### 6 Acknowledgments

Thank you to Mr. Wowon Priatna as the supervisor of this writing and research. So that it can be completed properly and on time.

#### BIBLIOGRAPHY

- [1] F. Petropoulos *et al.*, "Forecasting: theory and practice," *Int. J. Forecast.*, vol. 38, no. 3, pp. 705–871, 2022, doi: 10.1016/j.ijforecast.2021.11.001.
- [2] I. Yolanda and H. Fahmi, "Penerapan Data Mining Untuk Prediksi Penjualan Produk Roti Terlaris Pada PT . Nippon Indosari Corpindo Tbk Menggunakan Metode K-Nearest Neighbor," vol. 3, no. 3, pp. 9–15, 2021.
- [3] J. Suntoro, *22-DATA MINING Algoritma dan Implementasi Menggunakan Bahasa Pemrograman PHP*, vol. 9, no. 9. 2019.
- [4] M. Galih, P. D. Atika, and Mukhlis, "Prediksi Penjualan Menggunakan Algoritma Regresi Linear Di Koperasi Karyawan 'Usaha Bersama,'" *J. Inform. Inf. Secur.*, vol. 3, no. 2, pp. 193–202, 2023, doi: 10.31599/jiforty.v3i2.1354.
- [5] F. H. Hamdanah and D. Fitrianah, "Analisis Performansi Algoritma Linear Regression dengan Generalized Linear Model untuk Prediksi Penjualan pada Usaha Mikro, Kecil, dan Menengah," *J. Nas. Pendidik. Tek. Inform.*, vol. 10, no. 1, p. 23, 2021, doi: 10.23887/janapati.v10i1.31035.
- [6] I. A. Nikmatun and I. Waspada, "Implementasi Data Mining untuk Klasifikasi Masa Studi Mahasiswa Menggunakan Algoritma K-Nearest Neighbor," *J. SIMETRIS*, vol. 10, no. 2, pp. 421–432, 2019.
- [7] P. S. T. Informatika and U. I. Majapahit, "Pemanfaatan Knowledge Data Discovery(KDD) Pada Pola Permainan Atlet Bulutangkis," *Explor. IT J. Keilmuan dan Apl. Tek. Inform.*, vol. 11, no. 1, pp. 1–6, 2019, doi: 10.35891/explorit.v11i1.1467.
- [8] M. Sholeh, E. K. Nurnawati, and U. Lestari, "Penerapan Data Mining dengan Metode Regresi Linear untuk Memprediksi Data Nilai Hasil Ujian Menggunakan RapidMiner," *JISKA (Jurnal Inform. Sunan Kalijaga)*, vol. 8, no. 1, pp. 10–21, 2023, doi: 10.14421/jiska.2023.8.1.10-21.

- [9] U. Ependi and A. Putra, “Solusi Prediksi Persediaan Barang dengan Menggunakan Algoritma Apriori (Studi Kasus: Regional Part Depo Auto 2000 Palembang),” *J. Edukasi dan Penelit. Inform.*, vol. 5, no. 2, p. 139, 2019, doi: 10.26418/jp.v5i2.32648.
- [10] R. K. Rainer, B. Prince, and C. Cegielski, *Introduction to Information Systems, 5th Edition: Fifth Edition*, 5th ed., vol. 12. 2014.
- [11] N. Nafi’iyah, “Prediksi Jumlah Penjualan pada Toko Makmur Jaya Elektronik dengan Regresi Linier,” *Res. Comput. Inf. Syst. Technol. Manag.*, vol. 2, no. 2, p. 47, 2019, doi: 10.25273/research.v2i02.5143.
- [12] I. Nabillah and I. Ranggadara, “Mean Absolute Percentage Error untuk Evaluasi Hasil Prediksi Komoditas Laut,” *JOINS (Journal Inf. Syst.*, vol. 5, no. 2, pp. 250–255, 2020, doi: 10.33633/joins.v5i2.3900.