Login : Jurnal Teknologi Komputer
ISSN : 2302-9692 (print) | 2723-8695 (online)
Vol. 14, No. 2, December 2020, pp. 113-120
LOGIN http://login.seaninstitute.org/index.php/Login

Application of Data Mining to Predict the Sales of the Best-Selling Shoe Products using the Simple Linear Regression Method

Arista MorinaTinambunan[1], Paska Marto Hasugian[2]
Informatics Engineering Study Program, STMIK Pelita Nusantara, Medan, Indonesia.

Article Info

Article history:

Received Okt 20, 2020
Revised Nov 24, 2020
Accepted Des 31, 2020

Keywords:

Data Mining,
Simple Linear Regression,
Sales,
Prediction.

ABSTRACT

Sales prediction is an estimate of sales at a future time in certain circumstances and is made based on data that has occurred. This prediction is influenced by the sales of products in PT.Sepatu Bata. Simple Linear Regression is a linear relationship between one independent variable and the dependent variable, this analysis is to determine the direction of the relationship between the independent variable and the dependent variable whether it is positive or negative and to predict the value of the dependent if the value of the independent variable has increased or decreased. The author will design a data mining implementation system to predict the sales of the best-selling shoe products to better utilize existing sales transaction data. The design will be implemented using the PHP programming language and MySQL database. This research is expected to produce a datamining implementation system to predict the sales of the best-selling shoe products using the website-based simple linear regression method.

Corresponding Author:

Arista MorinaTinambunan,
Informatics Engineering Study Program,
STMIK Pelita Nusantara, Medan, Indonesia,
Jl. Iskandar Muda No.1, Merdeka, Kec. Medan Baru, Kota Medan, Sumatera Utara 20154.
Email: arista04tinambunan@gmail.com

1. INTRODUCTION

The rapid development in data collection and storage technology has made it easier for organizations to collect large amounts of data resulting in data mountains[1]. The extraction of useful information from data mountains can be quite a challenging task[2]. Often traditional data analysis tools and techniques cannot be used to extract information from large data sizes[3]. Data mining is a series of processes for obtaining useful information from large database warehouses[4][5]. Prediction is a process of systematically estimating something that is most likely to happen in the future based on the past and present information that is owned, so that the difference between something that happened and the forecast results can be minimized[6]. Predictions do not have to provide definite answers to events that will occur, but rather try to find answers as closely as possible to what will happen[7].

Bata Shoe Shop is a footwear manufacturer and is part of the Bata Shoe Organization (BSO) Switzerland. This giant shoe company has many models of shoes and sandals. This company also produces shoes and sandals according to the trend, price and quality are very suitable. Where this company every day must meet consumer needs and are required to be able to make the right decisions in determining sales strategies. The strategy taken is to use sales transaction data.

By utilizing sales transaction data that has been stored in the database, management can find out customer shopping habits and customer behavior regarding which products are frequently purchased. From this data, it will produce knowledge or information in the form of products that are most often purchased by consumers and products that have accumulated because they are not sold[8]. Therefore, it is very important to manage transaction data in order to make it easier for management to plan stock supplies so there is no accumulation of products.

To manage transaction data, data mining and Linear Regression Algorithms are needed using old and large data to find out the best-selling products in the Bata Shoe Store. Linear Regression Algorithm is a statistical method that functions to test the extent of the causal relationship between the causal factor (X) and the consequent variable[9]. The causative factor is generally denoted by X or also known as predictor while the resultant variable is denoted by Y or also known as response[10]. Linear regression is also a statistical method used in production to forecast or predict quality and quantity characteristics[11].

In previous research related to data mining to predict using the Simple Linear Regression Algorithm, namely Murni Marbun, Hengki Tamando Sihotang, Melda Agustina Nababan, entitled Designing a System for Forecasting the Number of Foreign Tourism (Case Study: Statistics Indonesia on the number of foreign tourists in Sumatra) at STMIK Pelita Nusantara said "The Linear Regression Method was successfully applied to forecasting the number of foreign tourists in North Sumatra based on historical data for the last 10 years showing that foreign tourists who will come to visit in January 2016 in North Sumatra are 16,937 people"[12].

2. RESEARCH METHOD

The research method is defined as a scientific way to obtain data for specific purposes and uses. In this study, the authors used quantitative research methods with descriptive and associative problem formulation approaches, because of the variables to be examined for their relationships and the purpose of presenting a structured, factual picture of the facts and the relationship between the variables studied.

A quantitative research model or a simple paradigm in which there are 2 (two) variables, namely the product inventory variable as the independent variable / independent variable as indicated by (X) and the sales variable as the dependent variable as indicated by (Y).

To assist this research, it is necessary to arrange a framework that is clear in stages. This framework represents the steps that will be taken in solving the problems discussed. The framework used is as follows:

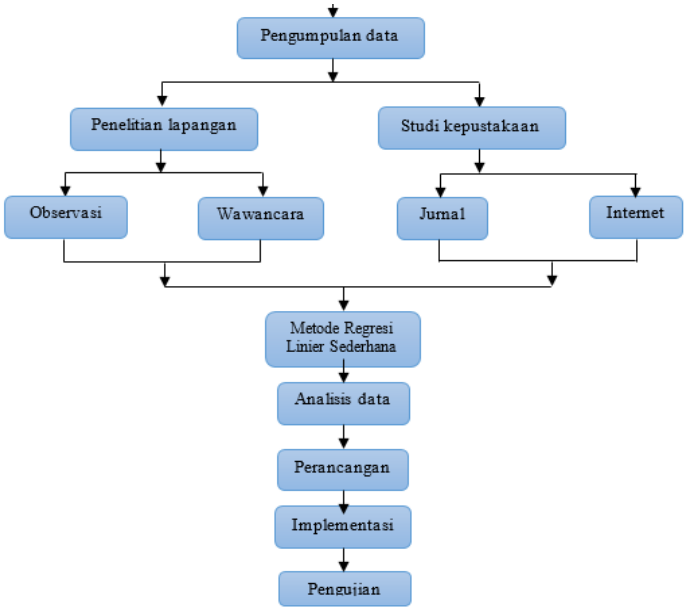

Figure 1. Research framework

3. RESULTS AND DISCUSSION

1. Input Data

The data used is sales data at PT. Sepatu Bata from January 2019 to December 2019. The data to be tested will be taken by sample data on each product, namely Men's Products, Ladies Products, and Children's Products. The data is displayed in table form as follows.

Table 1. Sales of Bata Shoe Products in 2019

	No	Nama Produk	Persediaan	Penjualan
MEN	1	Men Dress	158	224
	2	Men Casual	407	220
	3	Men Weinbrenner	195	122
	4	Men Sandal	962	541
LADIES	5	Drees Flat	344	229
	6	Dress Hell	66	41
	7	Sandal	642	418
	8	Maria Claire	359	174
CHILDREN	9	Infant	209	80
	10	Girls Dress	101	40
	11	Freetime	507	250
	12	Sandal	670	396

Table 2. Sales of Men Dress Products per Month

Kode	Produk	Tahun	Bulan	Penjualan (Y)
M01	Men Dress	2019	1	13
M01	Men Dress	2019	2	24
M01	Men Dress	2019	3	20
M01	Men Dress	2019	4	41
M01	Men Dress	2019	5	16
M01	Men Dress	2019	6	20
M01	Men Dress	2019	7	17
M01	Men Dress	2019	8	15
M01	Men Dress	2019	9	12
M01	Men Dress	2019	10	12
M01	Men Dress	2019	11	14
M01	Men Dress	2019	12	40
Jumlah				**244**

2. Calculating the value of x2 and xy on the dependent and independent variables

After determining the data to be tested, the next step is to calculate the value of x2 where the value of x is raised to the power of 2, and for the xy value by multiplying the variables x and y and the total amount of the data.

Table 3. Result of Summation of Men Dress Product Data

Produk	Tahun	Bulan	Penjualan (Y)	Prediksi (X)	X^2	XY
Men Dress	2019	1	13	-11	121	-143
Men Dress	2019	2	24	-9	81	-216
Men Dress	2019	3	20	-7	49	-140
Men Dress	2019	4	41	-5	25	-205

Men Dress	2019	5	16	-3	9	-48
Men Dress	2019	6	20	-1	1	-20
Men Dress	2019	7	17	1	1	17
Men Dress	2019	8	15	3	9	45
Men Dress	2019	9	12	5	25	60
Men Dress	2019	10	12	7	49	84
Men Dress	2019	11	14	9	81	126
Men Dress	2019	12	40	11	121	440
Jumlah			244	0	572	0

3. Calculating the values of a and b
   After obtaining these values, then calculate the coefficient values a and b:
   a = $\sum Y \ / \ n$
   =244/12
   = 20,33
   b = $(\sum YX) \ / \ (X^2)$
   = 0/572
   = 0
4. Simple Linear Regression Equations
   From these results, the equation for simple linear regression is as follows:
   Y' = a + bX
   Y' = 20,33 + (0)(X)
5. Calculating Sales Predictions (Y)
   From the results of the equation Y = 20.33 + (0) (X), the following results are obtained.
   Prediction Men Dress (Y) = 20.33 + 0 (13)
   Y = 20.33
   From these results, the prediction for Men Dress January 2020 products is 20.33 or 20.33 rounded.

Implementation is the stage of implementing a system to be built where the system or application is ready to operate with the results of the analysis and design carried out, whether the designed system can run well and produce the objectives achieved.

a. Display Login Page
   The login page is the page that is displayed when the user clicks the login button on the dashboard page to open the web application. This is intended to ensure that this application can only be used by users involved in the manufacturing system.

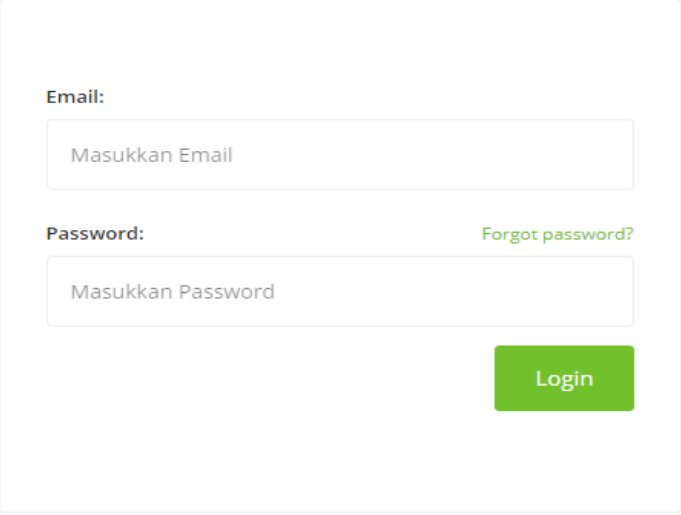

Figure 2. Login page

b. Home Page View

After successfully logging in, the home page will appear. On the home page there is a main menu which consists of several menus, namely Products, Data, and Predictions. The data menu has an inventory and sales sub-menu, the prediction menu has a sales sub-menu.The image of the home page on this system can be seen as follows:

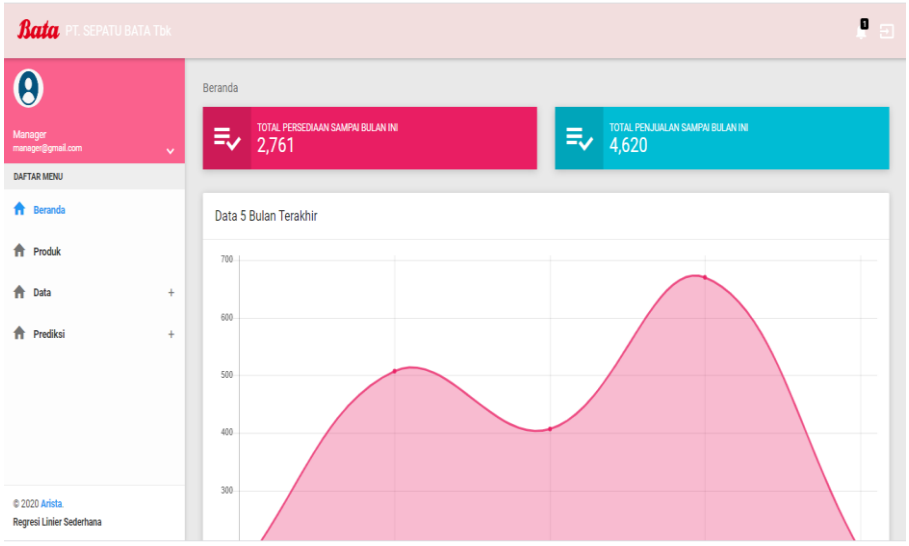

Figure 3. Home page view

c. Display Product Data

The product data display contains data from brick shoes products that will be predicted for each product. On this page, users can add, modify and delete product data. The page view is designed to be easy to use by the user.

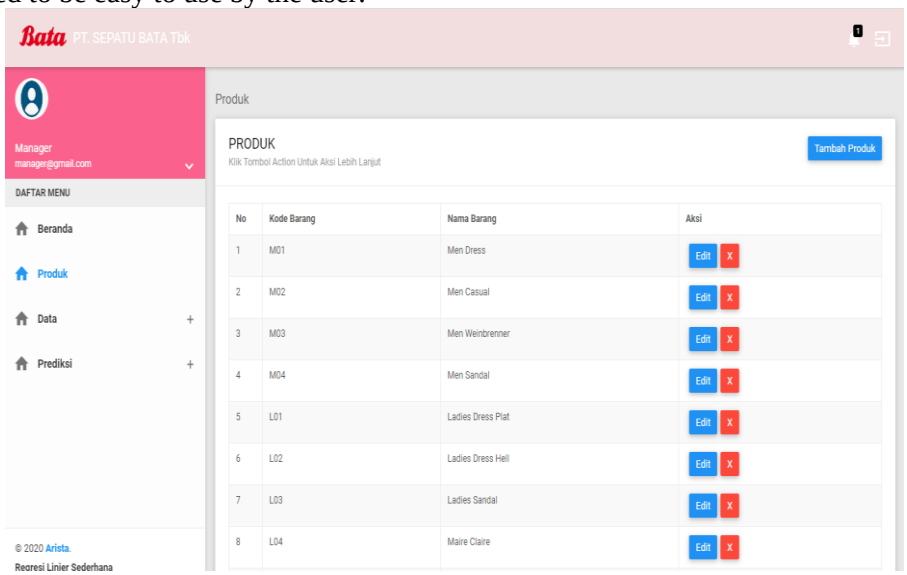

Figure 4. Product data page display

d. Sales Page Views

The sales data display contains data on products sold by month and year. On this page the user can add, change sales data or change dates and delete.

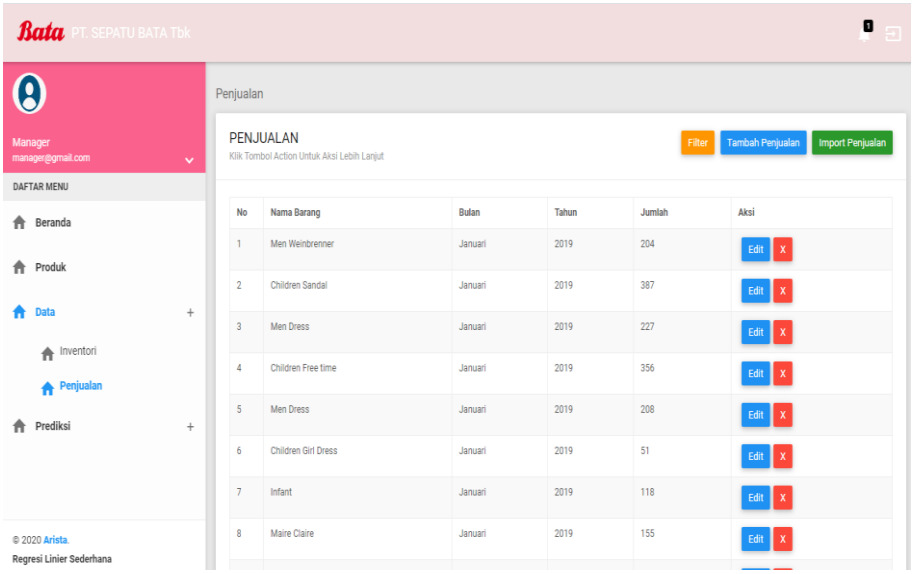

Figure 5. Sales Page Views

e. Sales prediction page display

The prediction page is used to carry out the product prediction process by selecting the product that will be predicted first, then entering the initial month and initial year and then entering the final month and year, then entering the month and year forecast, then clicking the simple linear regression button.

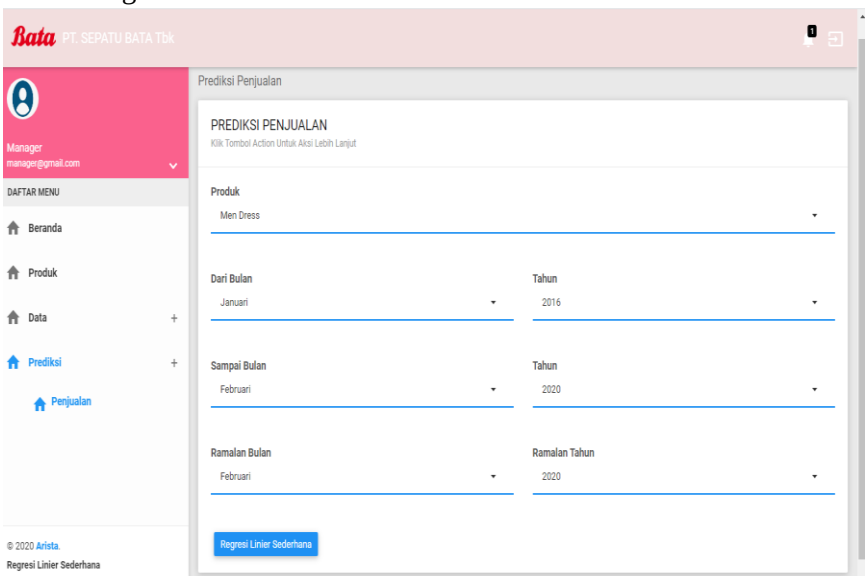

Figure 6. Sales prediction page image

f. Graphical prediction display

Based on the data inputted, the resulting graph is as follows, which indicates an increase in the predicted product purchases.

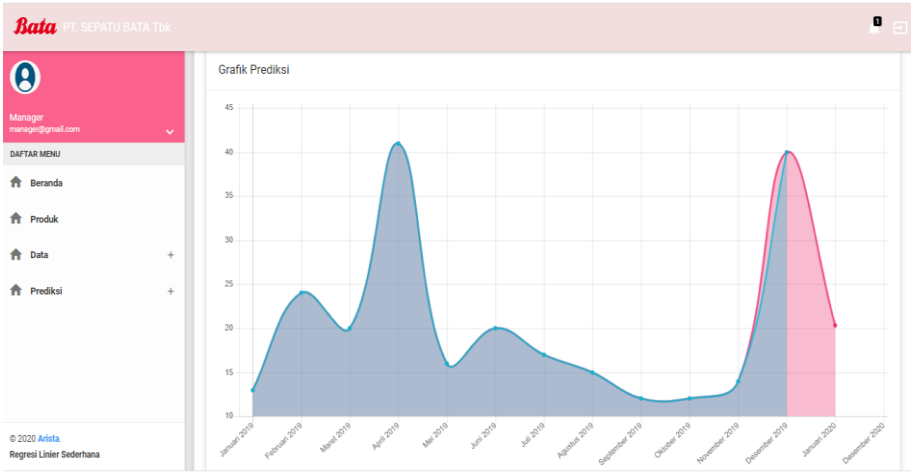

Figure 7. Graphical prediction display

4. CONCLUSION

Predicting the sales of the best-selling shoe products using the simple linear regression method, several conclusions can be drawn, including This research has been able to produce a prediction system for shoe product sales using a simple web-based linear regression method at PT. Bata shoes. The prediction results can be used as very valuable information in making decisions to determine the amount of shoe product inventory that will be sold in the following month.

REFERENCES

[1] P. Marpaung and N. Tarigan, "DATA MINING FOR DETERMINING BOOK LOAN PATTERNS IN-LIBRARY USING APRIORI ALGORITHM," INFOKUM, vol. 8, no. 1, Desembe, pp. 11–15, Nov. 2019, Accessed: Oct. 30, 2020. [Online]. Available: http://infor.seaninstitute.org/index.php/infokum/index.

[2] M. Sagala, "Algorithm Modified K-Nearest Neighbor (M-KNN) for Classification of Attention Deficit Hyperactive Disorder (ADHD) in Children," Login J. Teknol. Komput., vol. 13, no. 1, pp. 11–18, 2019, doi: https://doi.org/10.24224/login.v13i1.21.

[3] P. M. Hasugian, "Pengujian Algoritma C4.5 dengan Aplikasi Weka Dalam Pembentukan Pohon Keputusan," J. Mantik Penusa, 2018.

[4] H. Susanto and S. Sudiyatno, "Data mining untuk memprediksi prestasi siswa berdasarkan sosial ekonomi, motivasi, kedisiplinan dan prestasi masa lalu," J. Pendidik. Vokasi, 2014, doi: 10.21831/jpv.v4i2.2547.

[5] A. Maulana and A. A. Fajrin, "PENERAPAN DATA MINING UNTUK ANALISIS POLA PEMBELIAN KONSUMEN DENGAN ALGORITMA FP-GROWTH PADA DATA TRANSAKSI PENJUALAN SPARE PART MOTOR," KLIK - Kumpul. J. ILMU Komput., 2018, doi: 10.20527/klik.v5i1.100.

[6] S. E. Ginting and A. Rikki, "Prediction of Sparepart Sales Level using Exponential Smoothing Method," Login J. Teknol. Komput., vol. 12, no. 2, pp. 44–52, 2018, doi: https://doi.org/10.24224/login.v12i2.27.

[7] S. Sulistyono and W. Sulistiyowati, "Peramalan Produksi dengan Metode Regresi Linier Berganda," PROZIMA (Productivity, Optim. Manuf. Syst. Eng., 2018, doi: 10.21070/prozima.v1i2.1350.

[8] N. Mayana, B. Tarigan, M. Iqbal, and S. Barus, "Data Mining of Data Processing Items using Apriori Algorithm in Selly Sport & Electronic Shop in Perbaungan," Login J. Teknol. Komput., vol. 13, no. 2, pp. 1–5, 2019, doi: https://doi.org/10.24224/login.v13i2.7.

[9] A. Amrin, "Data Mining Dengan Regresi Linier Berganda Untuk Peramalan Tingkat Inflasi," J. Techno Nusa Mandiri, 2016.

[10] A. Hijriani, K. Muludi, and E. A. Andini, "Implementasi Metode Regresi Linier Sederhana Pada Penyajian Hasil Prediksi Pemakaian Air Bersih Pdam Way Rilau Kota Bandar Lampung Dengan Sistem Informasi Geofrafis," Inform. Mulawarman J. Ilm. Ilmu Komput., 2016, doi:

10.30872/jim.v11i2.212.
[11] S. Ningsih and H. H. Dukalang, “Penerapan Metode Suksesif Interval pada Analsis Regresi Linier Berganda,” Jambura J. Math., 2019, doi: 10.34312/jjom.v1i1.1742.
[12] M. Marbun, H. T. Sihotang, and M. A. Nababan, “Perancangan Sistem Peramalan Jumlah Wisatawan Asing,” J. Mantik Penusa, 2018.