



LINEAR REGRESSION METHOD APPLICATION TO PREDICT CIMORY MILK SALES

Hana Athhifa Ryantika¹, Merry Parida², Rustam³, Herman Afandi⁴, Sani Hanika Lubis⁵, Uswandi⁶, Asep Afandi⁷, Dwi Febry Nugroho⁸

*Corresponding author

hanaatthifaryantika@gmail.com¹,

Rustam@dcc.ac.id³

Informations System, Institut Teknologi Bisnis Dan Bahasa
Dian Cipta Cendikia, Lampung, Indonesia
Jl. Negara No.3 Candimas, Kotabumi, Lampung Utara
(34581), Indonesia

Article history:

Received Dec 13, 2022

Revised March 19, 2023

Accepted April 26, 2023

Keywords:

Data Mining;

Linier Regression;

RapidMiner.

Abstract

PT. Rasa Prima Sejati Wall's is a company engaged in the field of ice cream food and Cimory drinks. This company has various types of ice products and Cimory drinks to offer. Every year the company can create new products, especially in various flavors, not only that, the company guarantees the quality of the products it produces. The problem faced by the company is that the sale of available goods does not match consumer demand. The company also has not used predictors or plans for the sale of goods so that there is a buildup of goods which results in losses in the company. In this case the research will make predictions by looking at past sales data. The data taken is only sales data for Cimory products for the last three years from 2019-2021. Research using the Liner Regression method is one of the methods in the function of predicting sales. Linear regression is a statistical method used to construct a model or relationship between one or more independent variables X and response variable Y. The software used to support data processing is Rapid miner. The purpose of this study is to use data mining to identify the most popular, least popular, and desirable products or items at Pt. Rasa Prima so that companies can use this information as a guideline for managing their inventory and attending to customer disappointment properly available using a simple linear regression method, the prediction results for 2022 and Rapid Miner are 770000. This prediction can help companies make decisions and reduce inventory shortages.

1.0 INTRODUCTION

PT. Rasa Prima Sejati Wall's is a company engaged in the field of ice cream food and Cimory drinks. This company has various types of ice products and Cimory drinks to offer. Every year the company can create new products, especially in various flavors, not only that, the company guarantees the quality of the products it produces. Companies faced by companies Sometimes the sales of available goods do not match consumer demand. The company also has not used a predictor or plan for the sale of goods so there is a buildup of goods which results in losses for the company. In this case, the research will make predictions by looking at past sales data. Predictions are very important in helping the company determine the sales of

goods that must be provided by the company from 2019 to 2021. If this forecasting or prediction is applied to part of the product design process, then the company will be more assisted in selling production results, because predictions can provide the best output so it is hoped that the risk of errors caused by design errors can be determined as much as possible.

This study proposes using the Simple Linear Regression method[1] this method is one of the methods in the function of predicting sales. Linear regression is a statistical method that is used to form a model or relationship between one or more independent variables X and the response variable Y. Regression analysis with one variable X or also called a predictor while the effect variable is denoted by Y or also called a response[2]. The software used to support data processing is Rapidminer[3]. Based on the description of the background, the writing draws the title "Implementation of the Liner Regression Method for Predicting Sales of Cimory Milk".

2.0 THEORETICAL

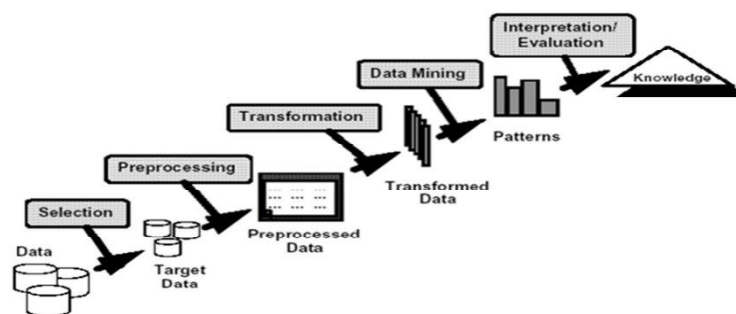
Data mining is often used in large data to find important information in data[4]. The application of data mining in research has been widely applied, one of which is predicting sales using Simple Linear Regression. One of them is by applying the science of forecasting[5]. Research related to predictive model analysis in the property sector has previously been carried out using e-commerce data and the Decision Tree and K-NN classification methods with results of accuracy of 75% and 71% for each method. Several other studies have implemented this forecasting technique by using various other methods. One of these prediction methods is the cascading series method. This method is a predictive method that estimates sales/demand in future periods using historical data. Some of the methods included in the time series method include : *Simple Moving Average*[6] *Weighted Moving Average*[7] *Exponential Smoothing*[8] and *Linear Regression*.

Linear regression as a prediction method has been used to predict book sales and inventory for the coming period. In this study, a simple linear regression method will be implemented to predict property sales in the future period using sales data in the previous period. The choice of the linear regression method as a prediction method in this study is based on its advantages in estimating simple model parameters and data based on time series. In addition, this method can perform analysis using several independent variables (X) so that the prediction results can be more accurate. Linear regression is one of the methods used in production to predict or predict quality and quantity characteristics. This is because by estimating various product combinations, companies can maximize profits and estimate the right amount of production.

3.0 METHODOLOGY

3.1 Data Mining

Data mining is a combination of several disciplines that brings together techniques from machine learning, pattern recognition, statistics, databases, and visualization to deal with the problem of retrieving information from large databases. The process of data mining or commonly known as Knowledge Discovery Database (KDD). KDD is often used interchangeably to describe the process of extracting hidden information in large databases. The stages of KDD in this study can be seen in Figure 1 below.



Picture 1. Knowledge Discovery Database Process[9].

1. *Data Selection*, is the stage of selecting data from a series of operational data that must be carried out before entering the information search stage in the Knowledge Discovery Database[10]. The data used in this study are from 2019-2021.

2. *Preprocessing Data Cleaning*, at the data cleaning stage which includes removing duplicate data, and selecting consistent data and noise[11].
3. *Data Mining*, The method used in this study is the K-Means Clustering method, where this method partitions the existing data into one or more clusters/groups.
4. *Evaluation*, At this stage, relevant data is obtained from the data mining process using the K-Means Clustering method. patterns or information generated from the data mining process in the form of rules obtained from K-Means Clustering calculations.
5. *Knowledge*, At this stage, namely testing the data that has been calculated to determine the accuracy of the research results, and whether the data or information obtained contradicts the facts or hypotheses that existed before[12].

3.2 Regresi Linier

Linear regression is a statistical tool used to determine the effect of one or several variables on one variable. Variables that affect are often called independent variables, independent variables, or explanatory variables. The affected variable is often referred to as the dependent variable or the dependent variable.

Linear regression can only be used on interval and ratio scales. Simple Linear Regression Simple linear regression analysis is used to determine the effect of one independent variable on one dependent variable. The general equation is : $Y = a + b X$. Where Y is the dependent variable and X is the independent variable. The coefficient a is a constant (intercept) which is the point of intersection of the regression line and the Y axis in Cartesian coordinates.

$$Y = a + bX \quad [13]$$

$$b = \frac{n(\sum xy) - (\sum x)(\sum y)}{n(\sum x^2) - (\sum x)^2} \quad [14]$$

$$a = \frac{\sum y - b(\sum x)}{n} \quad [15]$$

4.0 RESULTANTS

Sales information from PT Rasa Prima Sejati was gathered for this study from 2019 to 2021. The data is used as the starting point for computations using a straightforward linear regression approach to examine sales data. The processing of the sales data will then result in special data that is prepared to be shaped into new information. where Table 1 has the sales information.

Table 1. Sales Data

NUMBER	Mont	PRIOD 2019		PRIOD 2020		PRIOD 2021	
		SALE PRICE / PCS	TOTAL SALES/ PCS	SALE PRICE / PCS	TOTAL SALES/ PCS	SALE PRICE / PCS	TOTAL SALES/ PCS
1	JANUARY	Rp3.000	200000	Rp3.300	200000	Rp3.500	250000
2	FEBRUARY	Rp3.200	250000	Rp3.200	250000	Rp3.550	250000
3	MARCH	Rp3.250	200000	Rp3.300	200000	Rp3.450	220000
4	APRIL	Rp3.200	350000	Rp3.400	350000	Rp3.450	350000
5	MAY	Rp3.100	250000	Rp3.300	250000	Rp3.550	250000
6	JUNE	Rp3.300	350000	Rp3.300	350000	Rp3.500	380000
7	JULY	Rp3.400	300000	Rp3.400	300000	Rp3.450	390000
8	AUGUST	Rp3.350	150000	Rp3.350	150000	Rp3.450	300000
9	SEPTEMBER	Rp3.200	300000	Rp3.300	300000	Rp3.450	320000
10	OCTOBER	Rp3.300	150000	Rp3.300	150000	Rp3.550	350000
11	NOVEMBER	Rp3.300	150000	Rp3.300	150000	Rp3.450	360000
12	DECEMBER	Rp3.400	100000	Rp3.400	100000	Rp3.500	320000

4.1 Analysis of data calculation

The data analysis method is a method used to manage data results to obtain a conclusion. By looking at this frame of mind, the data analysis technique that can be used in this study is a quantitative analysis using a simple linear regression method. At this stage the author's group data based on annual sales to simplify the prediction calculation process. Then after the data is grouped, the data can be added up so that it can become sales data for all Cimory beverage products. The following is sales data that has been grouped in tables 4 :

Table 2. Group Sales Data 2019-2021

NUMBER	MONTH	SALE 2019	SALE 2020	SALE 2021
1	January	200000	200000	250000
2	February	250000	250000	250000
3	March	200000	200000	220000
4	April	350000	350000	350000
5	May	250000	250000	250000
6	June	350000	350000	380000
7	July	300000	300000	390000
8	August	150000	150000	300000
9	September	300000	300000	320000
10	October	150000	150000	350000
11	November	150000	150000	360000
12	December	100000	100000	320000

4.2 Data processing

In this study, data processing was performed using regression analysis to measure the strength of the linear relationship between two or more variables. This design is intended to test how much influence the X variable has on the Y variable.

A. Add up the previous total sales

Add up the previous total at this stage, the author groups the number of sales of Cimory products from 2019-2021 along with the sales data that the author has grouped, the formula is : =SUM(C4:E4), and so on.

Tabel 3. Total Sale 2019-2021

NUMBER	MONTH	Sales Amount
1	January	650000
2	February	750000
3	March	620000
4	April	1050000
5	May	750000
6	June	1080000
7	July	990000
8	August	600000
9	September	920000
10	October	650000
11	November	660000
12	December	520000

B. Determine the X and Y variables.

The next step is to create variables (X) and variables (Y) from the data contained in the following table :

Tabel 4. Variable X dan Y

X	Y
1	650000
2	750000
3	620000
4	1050000
5	750000
6	1080000
7	990000
8	600000
9	920000

10	650000
11	660000
12	520000

In the table above, the Y value is the sales value for the last 3 years from 2019 to 2021, while the X value in this case is the month of the year.

C. Determine the variables X^2 and $X*Y$

The next step is to create variables (X^2) and variables ($X*Y$) from the data contained in the following table:

Tabel 5. Variable X^2 dan $X*Y$

X	Y	X^2	$X * Y$
1	650000	1	650000
2	750000	4	1500000
3	620000	9	1860000
4	1050000	16	4200000
5	750000	25	3750000
6	1080000	36	6480000
7	990000	49	6930000
8	600000	64	4800000
9	920000	81	8280000
10	650000	100	6500000
11	660000	121	7260000
12	520000	144	6240000

D. Determining the Total Value for Each Attribute

The next step is to determine the total value of the variables X, Y X^2 , and $X*Y$, while the total value of each variable is as follows:

Table 6. Total Value

Total X	Total Y	Total X^2	Total $X * Y$
78	9240000	650	58450000

4.3 Application of the Linear Regression Method

Table 7. Result

Looking for value (a)	Looking for value (b)	Y
843181.8182	-11258.74126	770000.

4.5 Implementation Results

The process for calculating profit predictions from selling Cimory products uses simple linear regression and uses Microsoft Excel 2013. The information obtained from drug sales in 2019-2021 is as follows:

1. Prediction of sales of Cimory Products
2. Total sales results
 - a. Cimory Sales in 2019 = 2750000
 - b. Cimory Sales in 2020 = 2750000
 - c. Cimory Sales in 2021 = 3740000

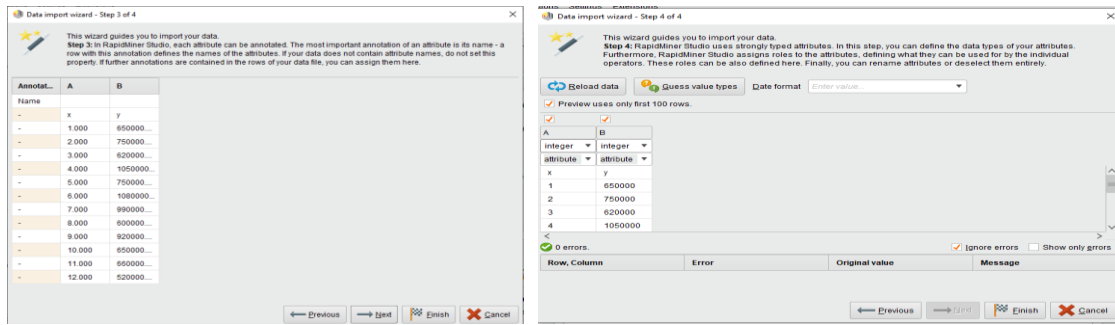


Figure 3. Display of selected data and Display of data that has changed its attributes

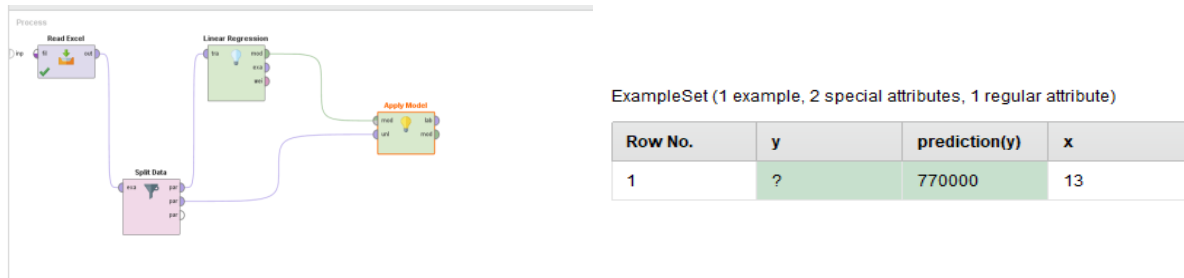


Figure 4. Connecting Split Data and Linear Regression to the Apply Model and Prediction result display

5.0 CONCLUSION

Simple linear regression was used in this study's modeling, which was done manually. For the prediction of 2022, which was 770000, the findings from the simple linear regression approach utilizing the RapidMiner software were the same. The study's findings should help businesses estimate stock sales and deal with product shortages. A prediction using the simple linear regression approach that has an error value of less than 20% is considered to have an accuracy of 80% from 90% training data and 10% test data, according to the MAPE error value requirement. The researcher aims to use a variety of data mining techniques in further studies, such as classification and clustering methods, in order to obtain more accurate results. This research is still far from flawless because there are still not enough methods being used.

REFERENCES

- [1] C. E. Simbolon, "Penerapan Algoritma Regresi Linier Sederhana Dalam Memprediksi Keuntungan dan Kerugian Kelapa Sawit Pt . Sri Ulina," *J. Inf. Syst. Res.*, vol. 2, no. 2, pp. 169–172, 2021.
- [2] F. Ginting, E. Buulolo, and E. R. Siagian, "Implementasi Algoritma Regresi Linear Sederhana Dalam Memprediksi Besaran Pendapatan Daerah (Studi Kasus: Dinas Pendapatan Kab. Deli Serdang)," *KOMIK (Konferensi Nas. Teknol. Inf. dan Komputer)*, vol. 3, no. 1, pp. 274–279, 2019, Doi: 10.30865/komik.v3i1.1602.
- [3] S. Kasus, P. Pt, S. Adimitra, and J. Cibitung, "Penerapan Data Mining Untuk Mempermudah Produksi Diapers Dengan Menggunakan Algoritma Regresi Linier," vol. 1, no. 1, pp. 176–179, 2022.
- [4] Z. Niqotaini *et al.*, "Analisis Perbandingan Kinerja Algoritma Klasifikasi Data Mining Penelusuran Minat Calon Mahasiswa," vol. 11, pp. 271–290, 2022.
- [5] D. Harini and L. S. Wahyuniar, "Estimasi Jumlah Murid Baru Menggunakan Metode Forecasting," *J. Instr. Math.*, vol. 2, no. 2, pp. 64–70, 2021, Doi: 10.37640/jim.v2i2.1025.
- [6] Y. N. Lubis, H. Winata, and S. Sobirin, "Data Mining Untuk Memprediksi Data Pengunjung dengan Menggunakan Algoritma Simple Moving Average," *J. SAINTIKOM (Jurnal Sains Manaj. Inform. dan Komputer)*, vol. 21, no. 2, p. 50, 2022, doi: 10.53513/jis.v21i2.5958.
- [7] A. R. Samsudin and A. S. Purnomo, "Implementasi Data Mining Untuk Menentukan Produk Laris Dengan K-Means Clustering dan Weighted Moving Average Implementation Data Mining To Determine The Best Selling Products With K-Means Clustering And Weighted Moving Average," no. December 2019, pp. 29–36, 2021.
- [8] N. Hafiza, "Model Prediksi Penjualan Barang Menggunakan Metode Single Exponential

- Smoothing," vol. 2, no. 9, pp. 1–11, 2022, [Online]. Available: <http://103.140.189.167/index.php/cyberarea/article/view/242>
- [9] S. Widaningsih, "Penerapan Data Mining untuk Memprediksi Siswa Berprestasi dengan Menggunakan Algoritma K Nearest Neighbor," *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 9, no. 3, pp. 2598–2611, 2022, doi: 10.35957/jatisi.v9i3.859.
- [10] J. Cybertech et al., "PENERAPAN DATA MINING UNTUK MENGESTIMASI," no. x, 1978.
- [11] D. Novianty, N. D. Palasara, and M. Qomaruddin, "Algoritma Regresi Linear pada Prediksi Permohonan Paten yang Terdaftar di Indonesia," *J. Sist. dan Teknol. Inf.*, vol. 9, no. 2, p. 81, 2021, Doi: 10.26418/justin.v9i2.43664.
- [12] J. Ilmiah and W. Pendidikan, "Implementasi Algoritma K-Means Clustering Status Gizi Balita Nurul Rizki Octaviyani 1 , Rini Mayasari 2 , Susilawati 3 Universitas Singaperbangsa Karawang," vol. 8, no. 13, pp. 370–381, 2022.
- [13] G. N. Ayuni and D. Fitriannah, "Penerapan metode Regresi Linear untuk prediksi penjualan properti pada PT XYZ," *J. Telemat.*, vol. 14, no. 2, pp. 79–86, 2019, [Online]. Available: <https://journal.ithb.ac.id/telematika/article/view/321>
- [14] W. M. Baihaqi, M. Dianingrum, and K. A. N. Ramadhan, "Regresi Linier Sederhana Untuk Memprediksi Kunjungan Pasien Di Rumah Sakit Berdasarkan Jenis Layanan Dan Umur Pasien," *J. Simetris*, vol. 10, no. 2, pp. 671–680, 2019, [Online]. Available: <https://www.jurnal.umk.ac.id/index.php/simet/article/view/3484>
- [15] 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 Geografis," *Inform. Mulawarman J. Ilm. Ilmu Komput.*, vol. 11, no. 2, p. 37, 2016, Doi: 10.30872/jim.v11i2.212.