Simple Linear Regression Method to Predict Cooking Oil Prices in the Time of Covid-19

Lilis Harianti Hasibuan[*1]; Darvi Mailisa Putri[2]; Miftahul Jannah[3]
Matematika, Fakultas Sains dan Teknologi, Universitas Islam Negeri Imam Bonjol Padang[1,2,3]
Email: lilisharianti@uinib.ac.id[1], darvimailisa@uinib.ac.id[2], miftahuljannah@uinib.ac.id[3]

Abstract

The background of this research is the soaring price of cooking oil during the Covid-19 period which continues to increase in the city of Padang. The research method used is a case study of data on cooking oil prices in the city of Padang. The purpose of this study is to obtain predictions of cooking oil prices. Linear regression is used as a prediction method for cooking oil prices in the next X(t) period. The research method used is a case study using simple linear regression. In this study, the actual cooking oil price Y(t) is the effect variable and the time period is the causal variable. The linear regression equation obtained is Y'=25239+124.56X. Testing the accuracy of the prediction results using RMSE with a value of 0.1913. The prediction of cooking oil prices using the linear regression method can be said to be in the very good category, it can be seen that the RMSE value is very small in the test and meets the standards.

Keywords: Prediction; Regresi Linear; RMSE.

Abstrak

Latar belakang dari penelitian ini adalah melonjaknya harga minyak goreng pada masa Covid-19 yang terus meningkat di kota Padang. Metode penelitian yang digunakan adalah studi kasus kasus terhadap data harga minyak goreng di kota Padang Tujuan dari penelitian ini adalah untuk mendapatkan prediksi harga minyak goreng. Regresi linear digunakan sebagai metode prediksi harga minyak goreng pada periode X(t) ke depan. Metode penelitian yang digunakan adalah studi kasus dengan menggunakan regresi linear sederhana. Pada penelitian ini harga minyak goreng aktual Y(t) merupakan variable akibat dan periode waktu sebagai variable penyebab. Persamaan regresi linear yang diperoleh adalah $Y^{'}$=25239+124,56X. pengujian keakurasian terhadap hasil prediksi dilakukan menggunakan RMSE dengan nilai 0,1913. Prediksi harga minyak goreng menggunakan metode regresi linear dapat dikatakan tergolong pada kategori yang sangat baik terlihat nilai RMSE yang sangat kecil pada pengujian tersebut dan memenuhi standar.

Kata Kunci: *Prediksi; Regresi Linear; RMSE.*

---

*Correspondence:
  Email: lilisharianti@uinib.ac.id

INTRODUCTION

The linear regression method is one of the methods in mathematical sciences with a family of statistical sciences that are widely used in research in the field of mathematical and applied sciences. The most commonly heard regression method is a simple linear regression method, in which this method wants to see the influence of one variable to another. This regression method can also be used in statistical science, time series problems, namely about forecasting or prediction. The prediction in the study is to predict the price of cooking oil, which at this time is very troubling to the public. One of the basic needs for everyone who lives in Indonesia is cooking oil. Commodity goods such as cooking oil are primary needs. Even though the price of cooking oil rises or falls, people will definitely continue to buy and consume it because cooking oil cannot be replaced with others. The price of cooking oil in the past three years has been caused by the existence of the Covid 19 virus which greatly affects the price of basic necessities. Unstable market prices make the prices of commodities such as cooking oil tend to fluctuate. Forecasting methods are very diverse in the field of statistics including (Putri & Hasibuan, 2020) predicting stock prices using geometric Brown motion. Price forecasting using linear regression has been done by many researchers, including (Hasibuan & Musthofa, 2022) predicting rice prices using a simple linear regression method, in this study researchers used simple linear regression. (Muhartini et al., 2021) forecasting the number of admissions of new students using linear regression, this research is very useful for some campuses to predict the number of applicants on a campus. Linear regression can also be done to predict property sales prices (Ayuni & Fitrianah, 2019), meaning that linear regeresi can make a prediction.

In this case, researchers are interested in predicting the price of basic necessities in West Sumatra Province in 2021 and 2022 by using actual data in 2020. Because cooking oil is a basic need that can be sold at retail or in per kg. For small communities with low economic resilience, they often buy cooking oil in the market or supermarket at a count per kg. For this reason, of course, traders or investors must have a calculation, if this cooking oil is sold per kg / in retail. It

should be that the investrors or suppliers of cooking oil should benefit from retail sales. To be able to estimate the number of sales, profits, application of forecasting science or forecasting can be done by making predictions or forecasting about what will happen in the future by looking at the situation in the past (Khotimah & Nyndiasari, 2017)

Forecasting or prediction has been used as a tool or one of the considerations of decision making, especially in the business or economic sector where losses are minimized as much as possible and profits or profits are achieved as much as possible (Mulyani, 2015). In this study, a simple linear regression method will be implemented to predict the price of Cooking oil in the future period using cooking oil price data in the previous period. The selection of this simple linear regression method as a prediction method in this study is based on its advantages in estimating simple model parameters and time-based data. In addition, this method can perform analysis using analysis using several free variables (X) so that the prediction results can be more accurate. The hope of this research is that it can become a reference material and knowledge for investers or swordsmen to make sales at retail.

RESEARCH METHODS

A. Prediction or Forecasting

Prediction or forecasting can be assumed as an activity to predict what will happen in the future (Prasetyowati, 2018). This activity is carried out by paying attention to past or current data or information either mathematically or statistically (Hakimah et al., 2020). Prediction aims to know, see and forecast the economic prospects or business activities and others.

A prediction can be qualitative (not in the form of numbers) or quantitative (in the form of numbers) (Syafruddin et al., 2014). Qualitative predictions tend to be difficult to obtain good results because the variables have very relative properties (Syafruddin et al., 2014). Meanwhile, in quantitative predictions, the results of predictions made are very dependent on the methods used (Hakimah et al., 2015). Different methods will produce different

predictive values (Hakimah et al., 2015). The purpose of this prediction is to find out and estimate the economic prospects or business activities (Gustriansyah, 2017).

In designing a prediction or forecasting method, there are three stages that must be passed, namely (Indarwati et al., 2019):

1. Analyzing past data. This step aims to get an idea of the pattern of the data in question.

2. Select the method to use. There are various methods available with their requirements. The choice of method can affect the outcome of the forecast. Forecast results are measured using the smallest errors or errors. Therefore, there is no forecasting method that is definitely good for any type of data.

3. The process of transformation of past data using the chosen method. If necessary, changes will be made according to their needs.

B. Linear Regression Method

The regression method is a statistical method that performs predictions using the development of mathematical relationships between variables, namely dependent variables (Y) and independent variables (X). Dependent variables are cause variables or cause variables or variables that are influenced, while independent variables are cause variables or variables that affect (Kusumawati et al., 2017). Predictions of the value of dependent variables can be made if the independent variables are known (Gustriansyah, 2017). Basically the price of commodities such as cooking oil can be modeled using simple linear regression.

Regression is a curve matching technique for low-fidelity data. Examples of low-precision data are observational data, experiments in the laboratory, and statistical data. Such data is called observational data from the objects we observe. Errors contained in bberaerasal data from the inaccuracy of the measuring instrument used, errors in reading the measuring instrument (parallax) or due to the behavior of the measured system. (Kavitha et al., 2016). In regression, curve matching is carried out by creating curves by creating a

function that approaches data points. The formula for linear regression by the least or simple squares method is:

$$a = \frac{\sum Y_i - b \sum X_i}{n} \ldots \ldots \ldots \ldots \ldots (1)$$

$$b = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{n \sum x_i^2 - (\sum x_i)^2} \ldots \ldots \ldots \ldots (2)$$

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

Where:

Y= Variable

a = Constant

b = Regression Coefficient

The method steps for the linear regression method are as follows: Data generation set The formation of the linear model of regression The steps of forming the model are as follows:

Step 1: Calculate X^2,Y^2,XY and the totals of each of them.

Step 2: Calculate a by using equations (1) and b using equations (2)

Step 3: Create a simple regression linear equation model.

Step 4: Make predictions or forecasts of causal variable factors or causal variables.

C. **Testing the Accuracy**

Accuracy of a prediction is determined by how much deviation or error this is, which occurs between the predicted data and the actual data or actual data (Izzah & Widyastuti, 2017). Errors in the formulation of a prediction are not only caused by the element of error but also the inability of a forecasting model to recognize other elements in the data series that affect the magnitude of deviations in predictions (Pranata et al., 2018). The magnitude of this error or deviation can be caused by the magnitude of unexpected factors (outliers) where there is no accurate prediction method or it can also be caused by the prediction method used cannot predict exactly the trend component, seasonal component or cycle component that may be contained in the data series (Pranata et al., 2018).

Among the various ways to calculate the magnitude of the error are some of them are mean squre error (MSE), root mean squre error (RMSE), and mean absolute percentage error (MAPE). MSE is the average of the squared difference between predicted and observed values, RMSE is the root of MSE, and MAPE is the absolute differential mean between predicted and actual values (Izzah & Widyastuti, 2017). The prediction results are declared good if the MAPE value is less than 10% (Indarwati et al., 2019). As for MSE and RMSE which use a gradient-based method, the lower the value, the better the prediction made.

The formula for the magnitude of the error can be seen below:

$$\text{MSE} = \frac{1}{n}\sum (Y_t - Y_t')^2$$

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum (Y_t - Y_t')^2}$$

$$\text{MAPE} = \frac{1}{n}\sum_{t=1}^{n} \frac{|Y_t - Y_t'|}{Y_t}$$

Where:

Y_t = actual value

Y_t' = peride forecasting value t

n = number of periods

D. Data Analysis Techniques

In this section of data analysis techniques will be explained several stages that were carried out during the implementation of this study. The stages of this research methodology are as follows:

a. Data collection The first stage in the implementation of this research is to collect data. The data collected is data on the price of cooking oil per kg in West Sumatra during the period January -December 2020 or for a period of one year.

b. Pre-procesing After data collection is carried out, the next step is to pre-procesing the data that has been obtained. This stage includes the implementation of data filtering where eliminating the required

variables and grouping cooking oil prices according to the arrangement per month, starting from January to December.

c. Calculations of X, Y, XY, and XX At this stage, calculations are carried out on X, Y, XY and XX. With X is the time period (month) while Y is the price of cooking oil.

d. Calculations a and b After obtaining the values of X, Y, XY, and XX, calculations are carried out to get the values of a and b. These values of a and b are the coefficients used in forming the regression equation model which will then be used to perform the prediction stages.

e. Getting the Linear Regression Equation At this stage, the coefficients a and b that have been obtained in the previous stage are then used to obtain the linear regression equation.

f. Predicting The linear regression equation model that has been obtained in the previous stage is then used to make predictions by applying the time series that will be predicted to it.

g. RMSE test The prediction results that have been obtained are then carried out accuracy testing. This accuracy test is carried out using only one accuracy test, namely by using RMSE.

RESULTS AND DISCUSSION

The following results and discussions will explain the steps that have been presented in the methodology section. The following are the results and discussion of this study as follows:

1. Data Collection

In the table below, the cooking oil price data will be presented that has been sorted starting from January to December 2020.

Table 1. Cooking Oil Price Data for 2022

Month	Price
January	25326.00
February	25528.00
March	25783.00
April	26112.00
May	25812.00
June	25798.00
July	25798.00
August	25819.00
September	26296.00
October	26616.00
November	26716.00
December	26977.00

2. Pre-Procesing

From Table 1 above, it can be seen that our data has been sorted from January and December of 2020. Outlier data is also not found on this data.

3. Calculation of the value of X,Y, XY,XX

Tabel 2. Calculation of the Value of X,Y, XY, XX

Month and Year	Cooking Oil Price (kg)	X(t)	Y(t)	XY	XX
Januari 2020	25326.00	1	25326.00	25326	1
February 2020	25528.00	2	25528.00	51056	4
March 2020	25783.00	3	25783.00	77349	9
April 2020	26112.00	4	26112.00	104448	16
May 2020	25812.00	5	25812.00	129060	25
June 2020	25798.00	6	25798.00	154788	36
July 2020	25798.00	7	25798.00	180586	49
August 2020	25819.00	8	25819.00	206552	64
September 2020	26296.00	9	26296.00	236664	81
October 2020	26616.00	10	26616.00	266160	100
November 2020	26716.00	11	26716.00	293876	121
December 2020	26977.00	12	26977.00	323724	144
		$\sum X_t$ = 78	$\sum Y_t$ = 312581	$\sum XY$ = 2049589	$\sum XX$ = 650

4. Calculation a and b

First we do the calculation of the value of b, which is as follows:

$$b = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{n \sum x_i^2 - (\sum x_i)^2}$$

The value of b is obtained:

$$b = \frac{12 \times 2049589 - 78 \times 312581)}{12 \times 650 - (78 \times 78)} = 124{,}56$$

From the above formula obtained b = 124.56. Next after we get the value of b will kit acari the value of a with the formula below:

$$a = \frac{\sum Y_i - b \sum X_i}{n}$$

Value of a obtained:

$$a = \frac{312581 - 124{,}56 \times 78}{12} = 25239$$

From the above calculations obtained a=25239 and b=124.56.

5. Linear Regression Equation

The coefficients a =25239 and b =124.56 that have been obtained are used to obtain the linear regression equation y=a+bx. The coefficient obtained produces a simple linear regression equation, namely: Y=25239+124.56X, the value of X is the time period running from January to December 2020. The Linear Regression Equation can be described using Ms.exel obtained by simulation as described above:

Table 3. Table of Regression Equations Obtained from Coefficients a and b

Month and Year	X(t)	Y'=25239+124,56X
January 2020	1	25363.32051
February 2020	2	25487.88345
March 2020	3	25612.44639
April 2020	4	25737.00932
May 2020	5	25861.57226
June 2020	6	25986.1352
July 2020	7	26110.69814
August 2020	8	26235.26107
September 2020	9	26359.82401
October 2020	10	26484.38695
November 2020	11	26608.94988
December 2020	12	26733.51282

The 2020 cooking oil price movement above will be modeled using a simple linear regression as shown below:
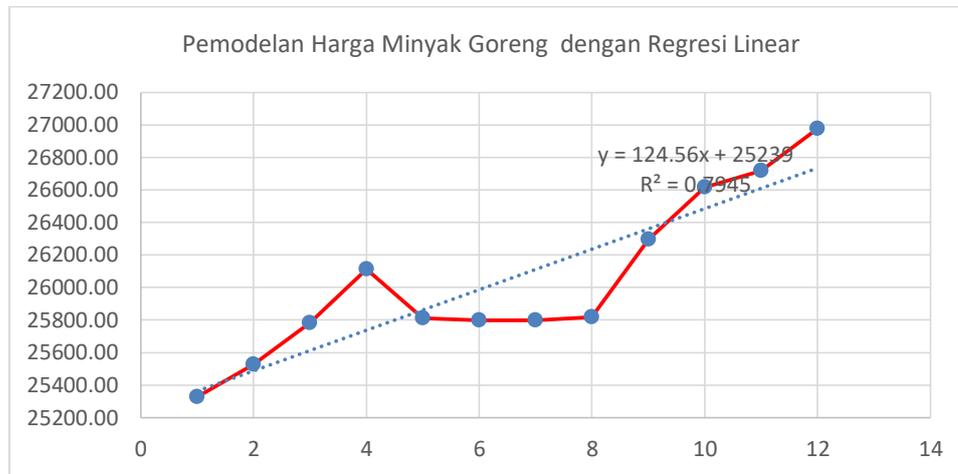

Figure 1. Modeling the Price of Cooking Oil with Simple Linear Regression

It can be seen that the value of the coefficient calculated manually is the same as that produced by ms.exel software with the equation as in figure 2.

6. Making Predictions

To make predictions, what will be used is the linear regression equation that has been obtained. In this study, cooking oil price predictions will be predicted in 2021 and 2022. The real data we have with January X(t) is labeled the 1st month and until December is labeled the 12th month. We will predict for the 13th month to the 36th month (in 2021 and in 2022).

Table 4. Predictions for the 13th Month to the 24th Month of 2021

Month and Year 2021	X(t)	$Y'=25239+124,56X$
January 2021	13	26858.07576
February 2021	14	26982.63869
March 2021	15	27107.20163
April 2021	16	27231.76457
May 2021	17	27356.32751
June 2021	18	27480.89044
July 2021	19	27605.45338
August 2021	20	27730.01632
September 2021	21	27854.57925
October 2021	22	27979.14219
November 2021	23	28103.70513
December 2021	24	28228.26807

Table 5. Predictions for the 25th Month to the 36th Month of 2022

Month and Year 2022	X(t)	Y'=25239+124,56X
January 2022	25	28352.831
February 2022	26	28477.39394
March 2022	27	28601.95688
April 2022	28	28726.51981
May 2022	29	28851.08275
June 2022	30	28975.64569
July 2022	31	29100.20862
August 2022	32	29224.77156
September 2022	33	29349.3345
October 2021	34	29473.89744
November 2021	35	29598.46037
December 2021	36	29723.02331

7. RMSE Test

To test the linear regression equation obtained how accurately predicting the price of cooking oil in 2021 and 2022 as the table above, the following formula is used:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum(Y_t - Y_t')^2}$$

The comparison between the Y(t) actual value and the Y'(t) predicted value is as follows is presented in the table below:

Table 6. Comparison of Y(t) values and Y'(t) Values

Month and Year	Cooking Oil Price (kg)	X(t)	Y(t)	Y(t)'=25239 +124,56X	Deviasi	(Deviasi)²
January 2020	25326.00	6.76	25326.00	25363.32051	2.6	6.77334
February 2020	25528.00	0.16	25528.00	25487.88345	-0.4	0.126365
March 2020	25783.00	0.49	25783.00	25612.44639	0.7	0.471255
April 2020	26112.00	94.09	26112.00	25737.00932	9.7	94.64251
May 2020	25812.00	77.44	25812.00	25861.57226	8.8	76.91985
June 2020	25798.00	144	25798.00	25986.1352	12	139.5317
July 2020	25798.00	795.24	25798.00	26110.69814	-28.2	792.1797
August 2020	25819.00	361	25819.00	26235.26107	-19	364.9525
September 2020	26296.00	100	26296.00	26359.82401	-10	101.2392

October 2020	26616.00	1	26616.00	26484.38695	-1	1.04002
November 2020	26716.00	64	26716.00	26608.94988	8	64.3548
December 2020	26977.00	289	26977.00	26733.51282	17	291.1836
Total						$\sum$ 1933,18

So the RMS error is obtained:

$$RMSE = \sqrt{\frac{1}{n} \sum (Y_t - Y'_t)^2}$$

$$RMSE = \left(\frac{1933,18}{12}\right)^{1/2} = 0,1913$$

The RMS error value is very accurate, which is 0.1913. It can be seen from the following figure that the predicted value is very close to the linear timeline.

CONCLUSION

As a conclusion from the results of research on cooking oil price prediction using a simple linear regression, the result of the cooking oil price prediction received an error or RMS error of 0.1913 or an accuracy rate of 96.56%. It can be known with this research that traders or investors to predict the price of cooking oil. Perhaps subsequent researchers could use multiple linear regression to predict the price of another commodity.

REFERENCES

Ayuni, G. N., & Fitrianah, D. (2019). Penerapan Metode Regresi Linear Untuk Prediksi Penjualan Properti pada PT XYZ. *Jurnal Telematika*, *14*(2), 79–86. https://journal.ithb.ac.id/telematika/article/view/321

Gustriansyah, R. (2017). Analisis Metode Single Exponential Smoothing Dengan Brown Exponential Smoothing Pada Studi Kasus Memprediksi Kuantiti Penjualan Produk Farmasidi Apotek. *Seminar Nasional Teknologi Informasi Dan Multimedia 2017*, *3*, 5–12. https://ojs.amikom.ac.id/index.php/semnasteknomedia/article/view/16 53

Hakimah, M., Muhima, R. R., & Yustina, A. (2015). RANCANG BANGUN APLIKASI PERAMALAN PERSEDIAAN BARANG DENGAN

METODE TREND PROJECTION. *Jurnal Simantec*, *5*(1). https://doi.org/10.21107/SIMANTEC.V5I1.1023

Hakimah, M., Rahmawati, W. M., & Afandi, A. Y. (2020). Pengukuran Kinerja Metode Peramalan Tipe Exponential Smoothing Dalam Parameter Terbaiknya. *Network Engineering Research Operation*, *5*(1), 44. https://doi.org/10.21107/nero.v5i1.150

Hasibuan, L. H., & Musthofa, S. (2022). *Penerapan Metode Regresi Linear Sederhana Untuk Prediksi Harga Beras di Kota Padang | Hasibuan | JOSTECH: Journal of Science and Technology*. https://www.ejournal.uinib.ac.id/jurnal/index.php/jostech/article/view/3802/pdf

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

Izzah, A., & Widyastuti, R. (2017). Prediksi Harga Saham Menggunakan Improved Multiple Linear Regression untuk Pencegahan Data Outlier. *Juni*, *KINETIK*(3), 141–150. https://doi.org/10.22219/kinetik.v2i3.268

Kavitha, S., Varuna, S., & Ramya, R. (2016). A comparative analysis on linear regression and support vector regression. *2016 Online International Conference on Green Engineering and Technologies (IC-GET)*, 1–5. https://doi.org/10.1109/GET.2016.7916627

Khotimah, T., & Nyndiasari, R. (2017). *View of FORECASTING DENGAN METODE REGRESI LINIER PADA SISTEM PENUNJANG KEPUTUSAN UNTUK MEMPREDIKSI JUMLAH PENJUALAN BATIK (STUDI KASUS KUB SARWO ENDAH BATIK TULIS LASEM)*. https://e-jurnal.pelitanusantara.ac.id/index.php/mantik/article/view/541/327

Kusumawati, N., Kusumawati, N., Marisa, F., & Wijaya, I. D. (2017). PREDIKSI KURS RUPIAH TERHADAP DOLLAR AMERIKA DENGAN MENGGUNAKAN METODE REGRESI LINEAR. *JIMP (Jurnal Informatika Merdeka Pasuruan)*, *2*(3). https://doi.org/10.37438/jimp.v2i3.79

Muhartini, A. A., Sahroni, O., Rahmawati, S. D., Febrianti, T., & Mahuda, I. (2021). *View of ANALISIS PERAMALAN JUMLAH PENERIMAAN MAHASISWA BARU DENGAN MENGGUNAKAN METODE REGRESI LINEAR SEDERHANA*. 17–23. https://bayesian.lppmbinabangsa.id/index.php/home/article/view/2/10

Mulyani, D. (2015). Prediction of New Student Numbers using Least Square Method. *IJARAI) International Journal of Advanced Research in Artificial Intelligence*, *4*(11). www.ijarai.thesai.org

Pranata, A., Hsb, M. A., Akhdansyah, T., Anwar, S., & Statistika, P. S. (2018).

Penerapan Metode Pemulusan Eksponensial Ganda dan Tripel Untuk Meramalkan Kunjungan Wisatawan Mancanegara ke Indonesia. *Journal of Data Analysis*, *1*(1), 32–41. https://doi.org/10.24815/JDA.V1I1.11873

Prasetyowati, E. (2018). Aplikasi Penentuan Harga Pokok Produksi Batik Madura Dengan Metode Activity Based Costing Dan Analisis Regresi Linier. *JUTI: Jurnal Ilmiah Teknologi Informasi*, *16*(1), 48. https://doi.org/10.12962/j24068535.v16i1.a690

Putri, D. M., & Hasibuan, L. H. (2020). Penerapan Gerak Brown Geometrik pada Data Saham PT. ANTM. *MAp (Mathematics and Applications) Journal*, *2*(2), 1–10.

Syafruddin, M., Hakim, L., & Despa, D. (2014). Metode Regresi Linier Untuk Prediksi Kebutuhan Energi Listrik Jangka Panjang (Studi Kasus Provinsi Lampung). *Jurnal Informatika Dan Teknik Elektro Terapan*, *2*(2). https://doi.org/10.23960/jitet.v2i2.237