Published online on the page: https://journal.orange-institute.org/index.php/cister/index

International Journal CISTER

| ISSN (Online) 2963-2994 |

Orange Institute

Prediction System of Poor Population In Lubuklinggau City Using Multiple Linear Regression Method

Andri Anto Tri Susilo [a,*], *Budi Santoso* [b]

[a] *Bina Insan University, Lubuklinggau, Indonesia*

ARTICLE INFORMATION

Article History:
Received: 11 June 2023
Final Revision: 11 July 2023
Published *Online*: 11 July 2023

KEYWORDS

Poverty
Prediction
Multiple Linear Regression

CORRESPONDENCE

E-mail: andri_anto_tri_susilo@univbinainsan.ac.id*

A B S T R A C T

There are several factors that greatly affect poverty, namely the Human Development Index (IPM), economic growth, and unemployment. There are various forecasting or prediction methods that can be used to predict the poverty rate in the future. One method that is commonly used is the multiple linear regression method. Multiple Linear Regression is a linear regression model involving more than one independent variable or predictor. By forecasting using multiple linear regression models, it is hoped that the government will be able to obtain accurate information about the number of poor people in the future. From the results of the study, it was found that the data error rate obtained MSE (Mean Squere Error) = 114672.764, RMSE (Root Mean Squere Error) = 338.633673146356 and MAPE (Mean Absolute Percentage Error) = 1.227198678. With a mape value of 1.227198678 or 1.227198678%, the value obtained is in the Very Good category

I. Introduction

Poverty is by far one of the most difficult socio-economic problems to solve. As an archipelagic country with geographical, economic, socio-cultural and government regulations, Indonesia faces a very complex and varied problem of poverty in each region. There are several factors that greatly influence poverty, namely the Human Development Index (IPM), economic growth, and unemployment [1]. Based on data from the Central Statistics Agency for Lubuklinggau City, Lubuklinggau City occupies the second position in South Sumatra Province for the largest number of poor people. BPS data records the number of poor people in Lubuklinggau City as follows [2].

Tabel 1. Table 1. Data on the Poor Population in Lubuklinggau City

Year	Poor Population
2012	29220
2013	30730
2014	30180
2015	33210
2016	31050
2017	29540
2018	29740
2019	29980
2020	29800
2021	31610

To overcome the existing problems, Lubuklinggau City needs to make various breakthroughs and actions in an effort to reduce the number of poor people in the future. In order for poverty alleviation efforts in Lubuklinggau City to run well, the Lubuklinggau City government needs accurate information about future data on the poor population and a method is needed so that the government can predict accurately the poverty rate in the future.

II. Method

The method uses a qualitative approach, namely a mixture of various data sources and various methods. This research is applied research, where researchers use research data, namely data on the number, poor population, human development index, the number of unemployed and economic growth figures from 2012 to 2021. The prediction method used is multiple linear regression

Linear regression is a statistical method that makes predictions using the development of mathematical relationships between variables, namely the dependent

https://doi.org/10. 56481/cister.v2i01.160

variable (Y) and the independent variable (X). The dependent variable is the effect variable or the variable that is affected. Predictions of the value of the dependent variable can be made if the independent variable is known

The steps to be taken can be seen below:

1. Research dataset creation/processing
2. Establishment of a Linear Regression model (model created based on training data).
3. Create a Calculation Auxiliary Table to calculate the total value of X1Y, X2Y, X3Y, X1X2, X1X3, X2X3, X12, X32,X32
4. Make matrices A, A1, A2, A3 and A4
5. Look for the determinant value in each matrix. To find the determinant value, the author uses the Sarrus method. And from the calculation results obtained the determinant value in each matrix is
   a. Determinant value A = 30291678933.7127
   b. Determinant Value A1 = -461020273658920
   c. Determinant value A2 = 18014137441455.8
   d. Determinant value A3 = 14723741251.8222
   e. Determinant value A4 = -6779239890619.62
6. Recognize the values b1,b2,b3 and b4

   b1 value $= \frac{\text{determinant value of A1}}{\text{determinant value}}$

   Value b1 $= \frac{-461020273658920}{30291678933{,}7127}$

   $= -15219{,}37013$

   b2 value $= \frac{\text{determinant value of A2}}{\text{determinant value}}$

   b2 value $= \frac{18014137441455.8}{30291678933.7127}$

   $= 594.689303319111$

   b3 value $= \frac{\text{determinant value of A3}}{\text{determinant value}}$

   b3 value $= \frac{14723741251.8222}{30291678933.7127}$

   $= 0.486065539121889$

   b4 value $= \frac{\text{determinant value of A4}}{\text{determinant value}}$

   b4 value $= \frac{-6779239890619.62}{30291678933.7127}$

   $= -223.798750325283$

7. Determine the equation or multiple linear regression model. After getting the values b1.b2, b3 and b4, then we can determine the multiple linear regression model, namely

   Y = -15219.37013+ + 594.689 X1 + 0.486065 X2

   + (-223,7987) X3

8. Determine MSE, RMSE and MAPE Values:
   a. MSE (Mean Squere Error) = 114672.764
   b. RMSE (Root Mean Squere Error) = 338.633673146356
   c. MAPE (Mean Absolute Percentage Error) = 1.227198678
   d. Based on the results of calculating the mape value of 1.227198678 or 1.227198678%, the value obtained is in the very good category.

III. Results and Discussion

3.1. Read_CSV

The read_csv command in the Python programming language is used to access data sets in CSV (Comma Separated Values) format. The read_excel command can be seen in the following image:

```
df = pd.read_csv('Data_penduduk.csv')
df.head()
```

Figure 1. Display of the CSV Read Command

	ipm	pengangguran	pertumbuhan_ekonomi	penduduk_miskin
0	72.04	6850	6.35	29220
1	72.55	7170	3.37	30730
2	72.84	6800	6.30	30180
3	73.17	12310	6.00	33210
4	73.57	8970	6.33	31050

Figure 2. Display Results

3.2. Matplotlib

Matplotlib is the most popular python library for making data visualization more interesting and easy to understand so that matplotlib will feel more natural to learn. The Matplotlib command can be seen in the following figure:

```
%matplotlib inline
plt.scatter(df.ipm,df.pengangguran,df.pertumbuhan_ekonomi,df.penduduk_miskin)
```

Figure 3. Display of the matplotlib command

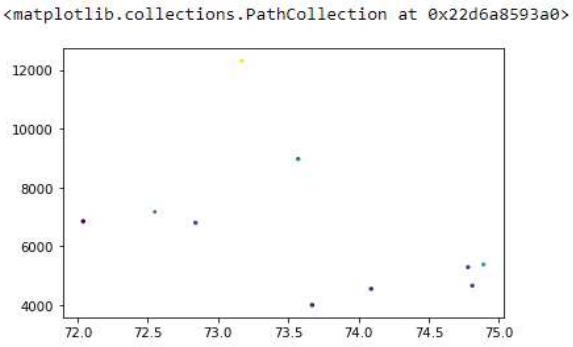


Figure 4. Display results matplotlib

https://doi.org/10.56481/cister.v2i01.160

3.3. Linear Regression Equations

Multiple linear regression equations can be created using the following command:

```
#model persamaan regresi berganda
reg = LinearRegression()
reg.fit(df[['ipm','pengangguran','pertumbuhan_ekonomi']], df.penduduk_miskin)
```

```
LinearRegression()
```

Figure 5. Display of Linear Regression Commands

```
#mencari nilai a
reg.intercept_
```

```
-15219.370133517303
```

```
#mencari nilai b1,b2 dan b3
reg.coef_
```

```
array([ 5.94689303e+02,  4.86065539e-01, -2.23798750e+02])
```

Figure 6. Results of values a, B1, B2 and B3

3.4. Regression Model Equation Test

We can test the regression model formed with new data as shown in the following figure:

```
#tes prediksi
reg.predict([[69.5,19000,7.98]])
```

```
array([33560.8676629])
```

Figure 7. Display of the Regression Model Test

3.5. Implementation With Orange(View)

The view command in the orange application is used to display the research data used. The view command can be seen in the following image:

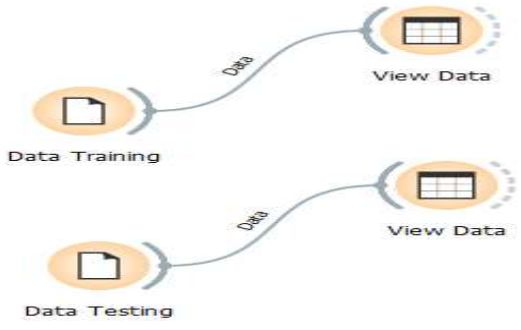


Figure 8. Work Flow View Data

	Penduduk_miskin	IPM	Pengangguran	Pertumbuhan_ekonomi
1	29220	72.04	6850	6.35
2	30730	72.55	7170	3.37
3	30180	72.84	6800	6.30
4	33210	73.17	12310	6.00
5	31050	73.57	8970	6.33
6	29540	73.67	4000	6.31
7	29740	74.09	4550	6.01
8	29980	74.81	4660	5.69
9	29800	74.78	5290	5.20
10	31610	74.89	5380	5.42

Figure 9. View Data Training Results

	Pertumbuhan_ekonomi	Tahun	IPM	Pengangguran
1	7.18	2022	69.50	19000
2	7.41	2023	68.42	17650
3	5.90	2024	68.15	17900
4	6.28	2025	68.21	17820
5	6.98	2026	70.20	18190
6	7.41	2027	70.65	16540
7	7.32	2028	71.85	11660

Figure 10. View Data Testing Results

3.6. Regression Equation Simulation

The simulation for finding a model or regression equation in the orange application can be seen in the following figure:

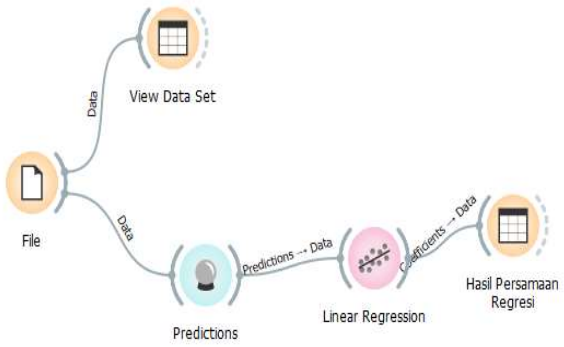


Figure 11. Work Flow View Data

	name	coef
1	intercept	-15219.4
2	IPM	594.689
3	Pengangguran	0.486066
4	Pertumbuhan_...	-223.799

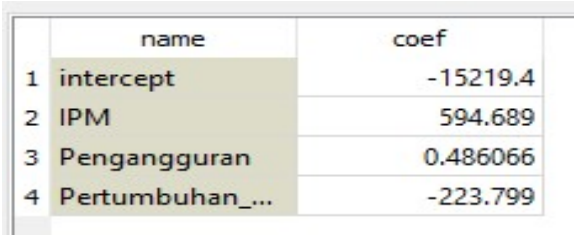
Figure 12. View Data Training Results

https://doi.org/10.56481/cister.v2i01.160

E-ISSN : 2963-2994

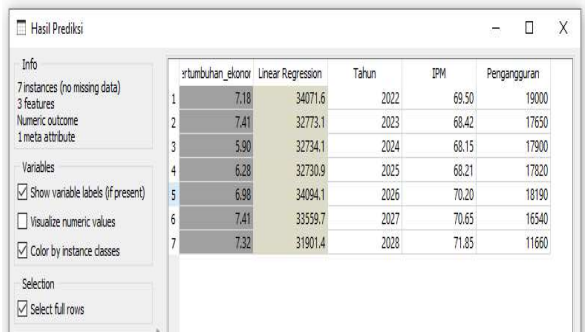
3.7. Simulation of Prediction Results

The simulation of prediction results in the orange application can be seen in the following figure:



	name	coef
1	intercept	-15219.4
2	IPM	594.689
3	Pengangguran	0.486066
4	Pertumbuhan_...	-223.799

Figure 13. Work Flow View Data



	rtumbuhan_ekonor	Linear Regression	Tahun	IPM	Pengangguran
1	7.18	34071.6	2022	69.50	19000
2	7.41	32773.1	2023	68.42	17650
3	5.90	32734.1	2024	68.15	17900
4	6.28	32730.9	2025	68.21	17820
5	6.98	34094.1	2026	70.20	18190
6	7.41	33559.7	2027	70.65	16540
7	7.32	31901.4	2028	71.85	11660

Figure 14. View Data Training Results

3.8. Graph Simulation of Predicted Results

he predicted graph simulation using the orange application can be seen in the following figure:

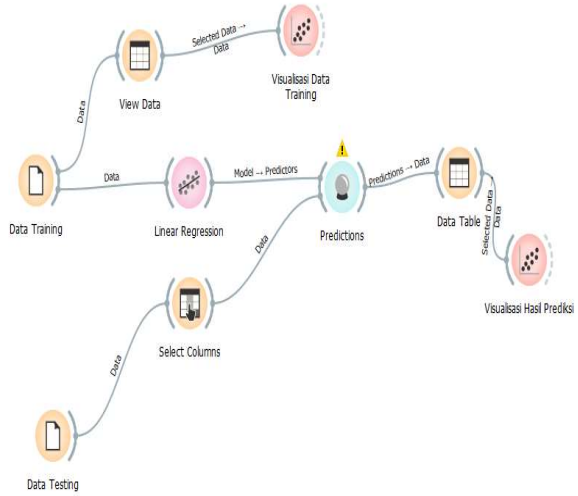


Figure 15. Work Flow View Data

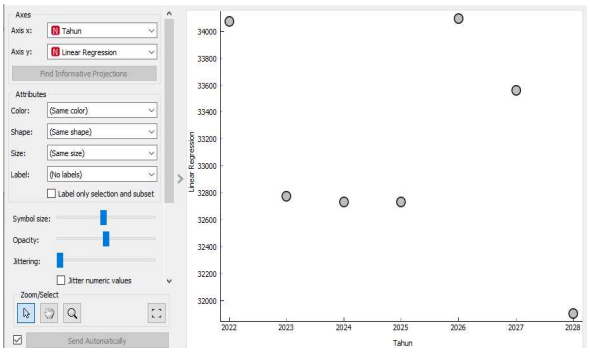


Figure 16. Figure 17. View Data Training Results

IV. Conclusion

With a system for predicting the number of poor people, the City Government of Lubuklinggau will no longer have difficulty predicting the number of poor people in the future, and this will affect the policies that will be taken to reduce the number of poor people. And also with the existence of a prediction system for the number of poor people, the City Government of Lubuklinggau has a special method that can be used to predict the number of poor people in the future. From the calculation of the data error level, the MSE (Mean Squere Error) value is 1893491.772, the RMSE (Root Mean Squere Error) value is 1376.042068 and the MAPE (Mean Absolute Percentage Error) value is 0.021490768. Based on the results of calculating the mape value of 0.021490768 or 0.021490768%, the value obtained is in the Very Good category.

References

[1] Y. C. Pratama, "Analysis of Factors Affecting Poverty in Indonesia," Essence, vol. 4, no. 2. 2015, doi: 10.15408/ess.v4i2.1966.

[2] Central Bureau of Statistics, "Total Poor Population in Lubuklinggau City," 2020. https://lubuklinggaukota.bps.go.id/indicator/23/139/7/nomor-penduduk-miskin.html.

[3] F. Andalia and E. B. Setiawan, "Development of a Job Seeker Data Processing Information System at the Padang City Social and Labor Service," Komputa: Scientific Journal of Computers and Informatics, vol. 4, no. 2.pp. 93–97, 2015, doi: 10.34010/komputa.v4i2.2431.

[4] S. Wardah and Iskandar, "WRAPPING PACKAGING (Case Study: Home Industry Arwana Food Tembilahan)," Journal of Industrial Engineering, vol. 9, no. 3.pp. 135–142, 2016.

[5] E. P. K. Orpa, E. F. Ripanti, and T. Tursina, "Prediction Model of Early Student Study Period Using the C4.5 Decision Tree Algorithm," Journal of Systems and Information Technology (JUSTIN), vol.

https://doi.org/10.56481/cister.v2i01.160

7, no. 4.p. 272, 2019, doi: 10.26418/justin.v7i4.33163.

[6] Initu.id, "Definition and Types of Prediction, Forecasting and Implementation," 2018. https://initu.id/amp/pengertian-dan-tipe-prediksi-ramalan-forecasting-dan-implementasi/.

[7] D. Kurniawan, "Poverty in Indonesia and its Management," Gema Eksos, vol. 5, no. 1.pp. 1–18, 2011.

[8] P. A. Rahman and R. Firman, "POVERTY IN THE PERSPECTIVE OF SOCIOLOGY," J. Educator. Tambusai, vol. 3, no. 6, pp. 1542–1548, 2019.

[9] et all Solikatun ., "POVERTY IN DEVELOPMENT," Journal of Chemical Information and Modeling, vol. 53, no. 9. p. 287, 2018.

[10] S. 'Diyah and E. Adawiyah, "Poverty and its Causing Factors," Journal of Social Work and Social Service, vol. 1, no. 1.p. 1, 2020.

https://doi.org/10.56481/cister.v2i01.160

E-ISSN : 2963-2994