Journal of Artificial Intelligence and Engineering Applications

Website: https://ioinformatic.org/

15th October 2024. Vol. 4. No. 1; e-ISSN: 2808-4519

Analysis of Village Residents Receiving Social Assistance Using Linear Regression Method

Rafli Fitriawan[1]*, Rusmin Saragih[2], Indah Ambarita[3]

[1, 2, 3] STMIK Kaputama

raflifitriawan64@gmail.com[1], evitha12014@gmail.com[2], yesnovada@gmail.com[3]*

Abstract

This study aims to analyze the recipients of social assistance in Banyumas Village using the simple linear regression method. The research examines how household income affects the amount of social assistance received. Data was collected from the Banyumas Village Office, including information on income and the amount of social assistance received by residents. The results show a negative relationship between household income and the amount of assistance received, where higher income leads to smaller assistance. The model also demonstrates good accuracy with an average prediction error (MAPE) of 9.38%. Additionally, an R^2 value of 0.999972 indicates that the model can explain almost all variations in the data. This study provides valuable insights into the effectiveness of the social assistance program in Banyumas Village and to help improve the program in the future.

Keywords: *Receiving Sosial Assistence, Village Residents Receiving, Linear Regression*

1. Introduction

The issues of poverty and economic inequality in underdeveloped regions are inevitable challenges in the dynamics of modern society, including in Indonesia. Despite significant efforts to address these problems, the challenges of reaching and providing assistance to the most vulnerable groups remain a major task for the government. The government strives to respond to these challenges by implementing various social assistance programs (Bansos) as a tangible expression of their commitment to combating poverty and economic inequality.

2. Literature Review

2.1. Data Analysis

Data analysis is a method used to understand how to describe data, data relationships, data semantics, and data constraints within an information system. It transforms raw data into useful information. This process involves investigating, cleaning, transforming, and modeling data to uncover patterns, trends, and relationships that are beneficial for decision-making.

2.2. Population

Population refers to a group of individuals or people, either individually or in groups, who occupy a specific area or country for at least one year at the time of data collection or a population census. These are people who are legally entitled to reside in a certain area, meaning they have official documents to live there, such as proof of citizenship, but may choose to live elsewhere. In sociology, the population is a group of people occupying a specific geographic area and space.

2.3. Village

A village is a type of human settlement typically located in rural areas or outside large cities. Villages usually have a smaller population than cities and consist of several houses and other structures, such as churches, schools, and small markets. Villages often have distinct social and cultural characteristics, such as close-knit community relationships and a more relaxed lifestyle compared to cities.

2.4. Social assistance

According to the Ministry of Social Affairs, social assistance is temporary aid provided to the poor, with the aim of helping them improve their living conditions in a reasonable manner. Social assistance is one component of the Social Security program, representing the government's commitment to addressing the needs of the poor and marginalized. Social assistance

refers to the provision of aid in the form of donations or goods, either from the government or organizations, to individuals, families, communities, and the general public. This assistance is not permanent and is selective, with the goal of protecting individuals from potential social threats.

2.5. Application

According to the Indonesian Dictionary (KBBI), 'application' is defined as the act of applying something. This definition refers to the concrete action of putting something into practice. Meanwhile, according to several experts, 'application' refers to the act of practicing a theory, method, or concept to achieve specific goals and meet the interests of a particular group or organization that has been planned and organized in advance. Application is not just a routine activity but involves a systematic process.

2.6. Linear Regression

Linear regression is a statistical method used to create a model that describes the relationship between a dependent variable (response: Y) and one or more independent variables (predictors: X). If there is only one independent variable, it is called simple linear regression, whereas if there are multiple independent variables, it is referred to as multiple linear regression.

$$Y = a + bX$$

Description:
Y: predicted value (dependent variable)
a : intersept
b : regression coefficient of variable *x*
x : independent variable
To find the intercept value (a) and coefficient (b), the following formula is used

$$a = \frac{(\sum y)(\sum x2) - (\sum x)\sum xy}{n\sum x2 - (\sum x)2}$$

$$b = \frac{n(\sum xy) - (\sum x)(\sum y)}{n(\sum x2) - (\sum x)2\ (2)}$$

2.7. Python

Python is a high-level programming language with simple syntax that works well for both new and seasoned developers. Compared to other programming languages that use sophisticated syntax, Python has a straightforward structure that makes it easier to write and maintain code.

2.10. MAPE (Mean Absolute Persentage Error)

MAPE is a predictive model-based test of results performed using historical data. It teaches how to utilize the MAPE approach to determine the level of accuracy by calculating the difference between actual and predicted data. The smaller the MAPE value, the higher the level of accuracy.

3. Analysis and Design

3.1. Research Methodology

In solving a problem in research, researchers must have a method or approach to address the issue so that the research can be completed effectively and yield the desired results. Research methodology is conducted to systematically investigate a matter using scientific methods and applicable sources. Based on the research methodology used in this study, a workflow or sequence of research activities can be developed.

1. Preparation
   This stage marks the beginning of the research process. The preparations include:
   a. Determining the background of the problem, which involves identifying issues and constraints by obtaining information directly from the Banyumas Village Office.
   b. Formulating the problems and the process of solving them.
   c. Setting boundaries to define the scope of the research, including the data used, variables, software or systems employed, and the expected output, which is to analyze social assistance recipients.
   d. Defining the objectives, specifying what results are to be achieved from the research process.
   e. Research benefits, providing a basis to evaluate whether the social assistance program is functioning well and how it can be improved in the future.
2. Theoretical Review
   In this stage, a theoretical review will be conducted on the existing problems. This review is aimed at determining the concepts to be used in the research, particularly regarding data mining, simple linear regression methods, and Python to be used in the analysis.
3. Data Collection

This stage involves gathering supporting data obtained from the Banyumas Village Office by visiting the office directly.

4. Data Analysis
   This stage will involve analyzing the supporting data, specifically the social assistance data obtained from the previous stage at the Banyumas Village Office, using data mining techniques with simple linear regression algorithms as the problem solving method. Data analysis is necessary to find solutions to the research problems being addressed.
5. Testing and Implementation
   This stage will involve testing data variables and implementing data, as well as developing the system program by:
   a. Preparing the data to be analyzed, specifically the social assistance data.
   b. Determining the variables to be used, including household income and received social assistance.

3.2. Research Supporting Data

To analyze data in a research study, supporting data is required to ensure that the research proceeds as expected. Simple linear regression is used to model the relationship between household income and the amount of assistance received, in relation to the needs of social assistance recipients in Banyumas Village. The data provided includes information about social assistance recipients. Based on the research conducted at the Banyumas Village Office.

Table 1: Social Assistance Data

No	Population	Income	Amount of Assistance
1.	A	2000000	200000
2.	B	2000000	200000
3.	C	2500000	150000
4.	D	2000000	200000
5.	E	2500000	150000
6.	F	2000000	200000
7.	G	2000000	200000
8.	H	2000000	200000
9.	I	1900000	275000
10.	J	1900000	275000

3.3. Analysis of the Application of Linear Regression

In the context of this research, simple linear regression is used to analyze the population receiving social assistance, where the recipients of social assistance are considered the dependent variable, and their values are influenced by one independent variable. This regression method was chosen because it can predict changes in the dependent variable based on changes in the independent variable. For example, if there is a change in the independent variable such as income, this method can be used to predict whether this change will affect the amount of social assistance received by the population. Additionally, this method allows for understanding the direction and strength of the relationship between the independent and dependent variables, helping to measure the extent to which these variables contribute to determining the recipients of social assistance.

Table 2: Analysis of the Application of Linear Regression

Population	Income	Amount of Assistance	X^2	Y^2	XY
A	2000000	200000	4000000000000	40000000000	400000000000
B	2000000	200000	4000000000000	40000000000	400000000000
C	2500000	150000	6250000000000	22500000000	375000000000
D	2000000	200000	4000000000000	40000000000	400000000000
E	2500000	150000	6250000000000	22500000000	375000000000
F	2000000	200000	4000000000000	40000000000	400000000000
G	2000000	200000	4000000000000	40000000000	400000000000
H	2000000	200000	4000000000000	40000000000	400000000000
I	1900000	275000	3610000000000	75625000000	522500000000
J	1900000	275000	3610000000000	75625000000	522500000000
Total	**20800000**	**2050000**	**43720000000000**	**436250000000**	**4195000000000**

a. Calculate the constant/intercept (a) and the regression coefficient (b)

$$a = \frac{(\sum Y)(\sum X^2) - (\sum X)(\sum XY)}{(\sum X^2) - (\sum X)^2}$$

$$a = \frac{(2.050.000_1)(43.720.000.000_9) - (2.080.000)(4.195.000.000.000)}{437.200.000.000.000 - 4.326.400.000.00}$$

$$a = \frac{(3536764809) - (3491535630)}{(5916796) - (5313025)} = 519.736$$

$$b = \frac{n \sum XY - (\sum X)(\sum Y)}{n(\sum X^2) - (\sum X)^2}$$

$$b = 10\,\frac{(4.195.000.000.000) - (2.080.000)(2.050.000)}{10(43.720.000.000.000) - (2.080.000)^2}$$

$$b = \frac{(41.950.000.000.000) - (4.264.000.000.000)}{} = 0{,}15131$$

(437.200.000.000.000) – (4.326.400.000.0005)

b. Calculating the simple linear regression equation Linear Regression

Y = 519,736.84+−0.1513 (X)

c. Performance testing based on the prediction model that has been created using testing data with MAPE output

Table 3. Calculation of the MAPE

Population	x	Y	$(\hat{Y})prediksi$	$\lvert Y-\hat{Y}\rvert(selisih)$	$\frac{\lvert Y_i-\hat{Y}_i\rvert}{Y_i}$
A	2000000	200000	217105,2641	-17105,2641	0,08552632
B	2000000	200000	217105,2641	-17105,2641	0,08552632
C	2500000	150000	141447,3696	8552,6304	0,057017536
D	2000000	200000	217105,2641	-17105,2641	0,08552632
E	2500000	150000	141447,3696	8552,6304	0,057017536
F	2000000	200000	217105,2641	-17105,2641	0,08552632
G	2000000	200000	217105,2641	-17105,2641	0,08552632
H	2000000	200000	217105,2641	-17105,2641	0,08552632
I	1900000	275000	232236,843	42763,157	0,155502389
J	1900000	275000	232236,843	42763,157	0,155502389
$\text{MAPE} = \sum \frac{\lvert Y_i-\hat{Y}_i\rvert}{Y_i}$					0,938197773
$\text{MAPE} = \sum_{}^{n} \frac{\lvert Y_i-\hat{Y}_i\rvert}{Y_i} \times 100\%$					9,38%

MAPE Value Calculation:

MAPE $= \frac{0{,}938197773 \text{ x } 100}{10} = 9{,}38\%$

Based on the MAPE value with a result of 9,38 % with an accuracy rate of 90.62 %.

3.4. Flowchart

A flowchart is a visual representation of a series of steps or processes in the form of a diagram. The steps in a workflow or algorithm are depicted using graphical symbols. To make complex processes easier to understand, both for expert and non-expert users, flowcharts can be used. The following is a flowchart of the linear regression algorithm.

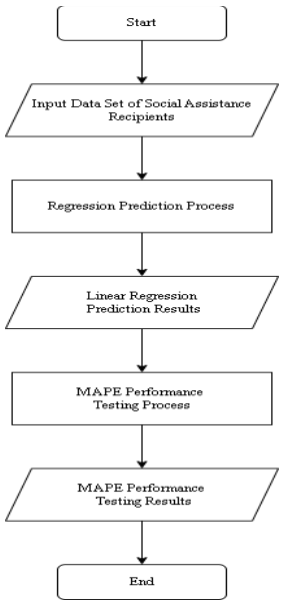

Fig. 1: Flowchart

4. Discussion and Implementation

4.1 Discussion

The steps taken for analyzing the data of village residents receiving social assistance using linear regression methods aim to predict the amount of social assistance received in the following year. This knowledge provides conclusions that can be used as further strategies to optimize efforts in identifying residents in Banyumas Village, improving the accuracy of data on social assistance recipients, ensuring a rapid response in the distribution of assistance, validating data, and creating a more responsive distribution of social assistance in the village.

4.2 Implementation

In this chapter, the results of the program trial predicting the amount of assistance received in the following year using linear regression methods will be explained.

1. Linear Regression Plot Display

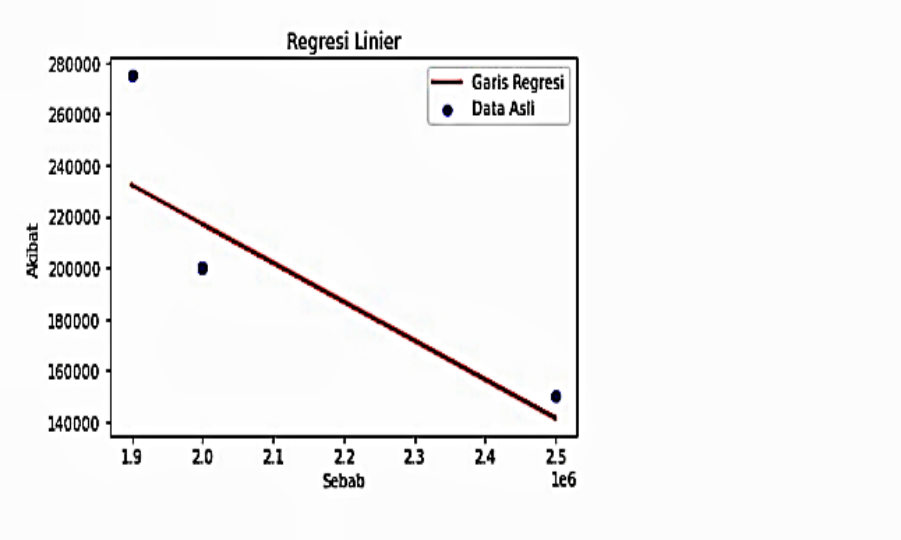

Fig. 2: Result Linear Regression Plot Display

2. Ordinary Least Squares (OLS)

OLS Regression Results

Dep. Variable:	y	R-squared:	0.653
Model:	OLS	Adj. R-squared:	0.609
Method:	Least Squares	F-statistic:	15.02
Date:	Fri, 06 Sep 2024	Prob (F-statistic):	0.00470
Time:	14:38:35	Log-Likelihood:	-114.87
No. Observations:	10	AIC:	233.7
Df Residuals:	8	BIC:	234.3
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	5.197e+05	8.16e+04	6.367	0.000	3.32e+05	7.08e+05
x1	-0.1513	0.039	-3.876	0.005	-0.241	-0.061

Omnibus:	2.840	Durbin-Watson:	1.118
Prob(Omnibus):	0.242	Jarque-Bera (JB):	1.765
Skew:	0.974	Prob(JB):	0.414
Kurtosis:	2.334	Cond. No.	2.05e+07

Fig. 3: Results of OLS Regression

5. Conclusion

Based on the analysis conducted by the author, the following conclusions can be drawn:

1. The variables used include data X (Income) and data Y (Amount of Assistance), with data obtained from the Banyumas Village Office. This can assist the village government in formulating more effective social assistance policies and programs.
2. The system design uses income data and the amount of assistance received, with results showing projections for the following year.

3. This research indicates that factors such as the level of poverty, the number of family members, and access to information are key determinants in the distribution of social assistance.

References

[1] Agus, S., Purwanto, Sumartono, & Makmur, M. (2013). Implementation of the Family Hope Program (PKH) in Breaking the Poverty Chain. Wacana, 16(2), 79-96.

[2] Alfarizi, M. R. S., Al-Farish, M. Z., Taufiqurrahman, M., Ardiansah, G., & Elgar, M. (2023). The Use of Python as a Programming Language for Machine Learning and Deep Learning. Student Scientific Work of Monotheism (KARIMAH TAUHID), 2(1), 1-6.

[3] Amalia, L., & Samputra, P. L. (2020). Economic Resilience Strategies for Poor Families Receiving Social Assistance Funds inTanah Tinggi, Central Jakarta. Sosio Konsepsia, 9(2), 113-131.

[4] Belakang, A. L. (1945). Immanuel Hutabarat.

[5] Faraag, A. H., & EI-Batal, A. I. E. (2017). [Effectiveness of Social Assistance in Combating Poverty Amidst Economic Slowdown in Indonesia Using Non-Parametric Approach]. Natural Science, 15(5), 55-68. https://doi.org/10.1360/N972017-00542

[6] Goleman, D., Boyatzis, R., & McKee, A. (2018). The Concept of Character, Capacity, Capital, Condition of Economy, and Collateral. Journal of Chemical Information and Modeling, 53(9), 1689-1699.

[7] Kus, I. L., Yunaini, F., & Hamidi, N. (2023). Sustainable Innovation in Business: Leveraging Flowcharts to Optimize Company Waste Value. Journal of Community Service Information, 1(4), 10012. https://doi.org/10.47861/jipmnalanda.v1i4.552

[8] Maharadja, A. N., Maulana, I., & Dermawan, B. A. (2021). Application of Multiple Linear Regression Method to Predict State Losses Based on Corruption Cases. Journal of Informatics and Computer Applications, 5(1), 95-102. https://doi.org/10.30871/jaic.v5i1.3184

[9] Noerkaisar, N. (2021). Effectiveness of Government Social Assistance Distribution to Address the Impact of COVID-19 in Indonesia. Journal of Treasury Management, 2(1), 83-104. https://doi.org/10.33105/jmp.v2i1.363

[10] Pelham, I. E. (2023). Secret Pathways (5th ed., p. 135135). https://doi.org/10.1093/oso/9780198599425.003.005

[11] Rahmat, C. A., Kurniabudi, & Novianto, Y. (2023). Application of Multiple Linear Regression Method to Estimate Population Growth Rate in Musi Banyuasin Regency. Journal of Informatics and Computer Engineering (JAKAKOM), 3(1), 359-369. https://doi.org/10.33998/jakakom.2023.3.1.732

[12] Rosaly, R., & Prasetyo, A. (2019). Understanding Flowcharts and the Most Commonly Used Flowchart Symbols.

[13] Saraswati, M. (2018). Be Smart Social Sciences. Graf Media Pratama.

[14] Suharto, E., & Thamrin, D. (2012). Family Hope Program (PKH): Cutting the Poverty Chain of the Nation's Children. Journal of Bandung Aspirations, 3(1), 1-20.

[15] UU Republik Indonesia. (2014). Indonesian Law on Villages. JDIH Ministry of Finance, April 18, 2(1), 45-54.

[16] Wijayanto, A. (2008). Simple Linear Regression Analysis. Undip, 4(1), 1-4.