Electronic Integrated Computer Algorithm Journal
Vol. 2, No. 1, October 2024, pp 23~28

ISSN: 3031-0350, DOI: 10.62123/enigma.v2i1.37

Application of Multiple Linear Regression Models in Forming Priority patterns of Village Fund Budget Use

Tia Alfi Sahara Nasution[1], Al-Khowarizmi[1*]

[1] *Information Systems, Universitas Muhammadiyah Sumatera Utara*, Indonesia.

DOI : 10.62123/enigma.v2i1.37

Received : August 31, 2024
Revised : September 21, 2024
Accepted : September 27, 2024

Keywords:
Village fund, Budget utilization, Village development, Population Area, Poverty level

ABSTRACT

This study aims to analyze the factors that influence the level of utilization of the village fund budget in Dolok Maraja Village, Simalungun Regency. Descriptive quantitative method with Multiple Linear Regression analysis was used to achieve the objective. The results showed that there was a significant relationship between the independent variable (village fund budget) and the dependent variable (the utilization rate of the village fund budget). The factors that most influence the utilization rate of the village fund budget are population, area, and poverty level. Implication of The implication of this study is that the government should consider these factors in the process of allocating village fund budgets. Villages with larger populations, larger areas, and higher poverty rates require larger budget allocations to ensure effective and efficient village development. In addition, the importance of good governance of village fund budgets and active community participation in the process of planning and implementing village development is also highlighted in this study. Specifically, this study shows that every year, the utilization rate of the village fund budget increases by an average of 6.5632. Meanwhile, each increase in the number of poor people and the Village Fund Ceiling decreases and increases the utilization rate of the village fund budget by an average of 2.6104 and 3.7433, respectively. Village area did not have a significant effect. Although this regression model has low explanatory power, it is statistically valid and fits the observed data. This research highlights the importance for the government to consider the factors of population, area, and Village Fund Ceiling in allocating the village fund budget, as well as improving governance and community participation in village development.

1. INTRODUCTION

Data mining is a process in which one or more techniques are used in computer learning to analyze and automate knowledge acquisition [1], and the application of data mining has proven effective in solving problems related to a large amount of existing data. Through the use of data mining, we can find, explore, or extract knowledge from the data or information we have [2]. Data Mining is a series of processes to explore added value from large data sets by producing knowledge that was previously unknown manually [3]. Data mining can be completed in two ways, namely regression and classification techniques. Regression is a statistical analysis that describes the relationship between two variables, namely the dependent variable (Y) and the independent variable (X). The types of linear regression consist of simple linear regression and multiple linear regression [4]. Multiple Linear Regression is used to examine the pattern of relationships between dependent variables and two or more independent variables [5]. In this study, Linear Regression uses two types of variables, namely dependent variables and independent variables. The dependent variable is the priority of village fund budget usage, while the independent variables include population, human resources (HR), area, education level, poverty level, health, welfare, and development. Data mining is used to extract complex patterns and trends from village data, while Multiple Linear Regression provides an in-depth understanding of the relationships between relevant variables. By combining the data mining and Multiple Linear Regression approaches, this study can identify the priority of village fund budget usage.

This study was designed with the aim of increasing efficiency, transparency, and fairness in the process of determining the use of village fund budgets. In addition, it is hoped that this application can have a positive impact on community welfare, especially in addressing the urgent needs faced by them. The application of the developed model is expected to be able to increase the efficiency of village fund allocation, strengthen transparency and accountability in decision-making, reduce the risk of inequality in fund distribution, and ensure that village fund allocation decisions are made more accurately according to target.

2. LITERATURE REVIEW

2.1 Data Mining

Data Mining may be a arrangement of processes to physically extricate included esteem within the frame of already obscure data from a database. The coming about data is gotten by extracting and recognizing vital or curiously designs from the information

*Corresponding Author Email: alkhowarizmi@umsu.ac.id

contained within the database [6]. The goal of data mining is to find previously unknown patterns, trends, or relationships from large and complex data sets.

2.2 Multiple Linear Regression

The use of village funds is prioritized with one of the main focuses being community empowerment, aimed at improving village progress [10]. The purpose of this resource allocation is to identify unions that show low performance, identify and prioriti ze specific strategies in the area to improve performance, and provide a platform for review, monitoring, and evaluation of ongoing progress [11]. Village finances are all rights and commitments within the setting of organizing Village administration that c an be esteemed in cash, counting all shapes of riches related to the rights and commitments of the Village. Village accounts come from unique town salary, APBD, and APBN [12]. Attention to the amount of village fund budget received and managed by the Village Government must be the focus of all parties in the village to jointly supervise and manage it in accordance with applicable l egal provisions. Therefore, preventing corruption of village funds is very important by increasing community participation in order to improve the quality of public services in the village. This is the background of this study [13].

3. RESEARCH METHODS

The purpose of this study is to build a Multiple Linear Regression model to determine the priority of village fund allocation by identifying the factors that influence the use of the village fund budget in Dolok Maraja Village.

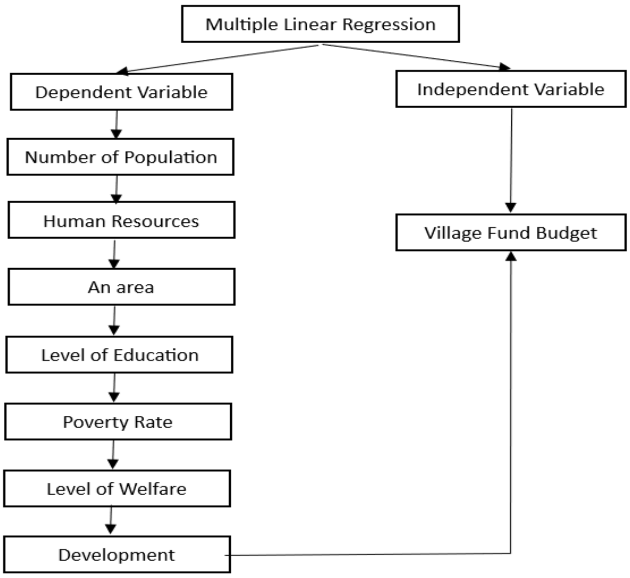


Figure 2. Research Method

4. RESULTS AND DISCUSSIONS

4.1 Programming System Implementation

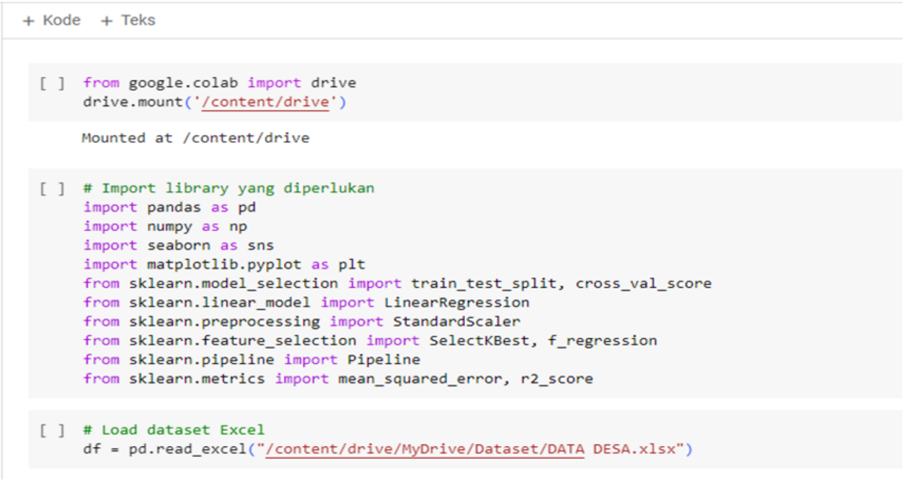
The data analysis program used is Google Collab. The following is a list of programs needed for the Multiple Linear Regression process:

Table 1. Required Programs

Program	Platform
IDE Python	Google Collab
Library	Pandas, Numpy, Seaborn, Matplotlib

4.2 Data Analysis Implementation

The data to be loaded into Python can be in xlsx or csv format. The function to load this data can use the pandas library, with the basic command 'read_excel' or 'read_csv'. Because this regression process requires numeric data, it is important to ensur e that the data is in the integer data type.



```
[ ] from google.colab import drive
    drive.mount('/content/drive')

    Mounted at /content/drive

[ ] # Import library yang diperlukan
    import pandas as pd
    import numpy as np
    import seaborn as sns
    import matplotlib.pyplot as plt
    from sklearn.model_selection import train_test_split, cross_val_score
    from sklearn.linear_model import LinearRegression
    from sklearn.preprocessing import StandardScaler
    from sklearn.feature_selection import SelectKBest, f_regression
    from sklearn.pipeline import Pipeline
    from sklearn.metrics import mean_squared_error, r2_score

[ ] # Load dataset Excel
    df = pd.read_excel("/content/drive/MyDrive/Dataset/DATA DESA.xlsx")
```

Figure 2. Import Data Command

The image above is a basic command in mapping data. The first is to import the library to be used, libraries in data processi ng that are widely used pandas, numpy, seaborn, libraries used in data visualization, namely matplotlib, scipy. To see data visualization in graphical form, it can be concluded by python below:

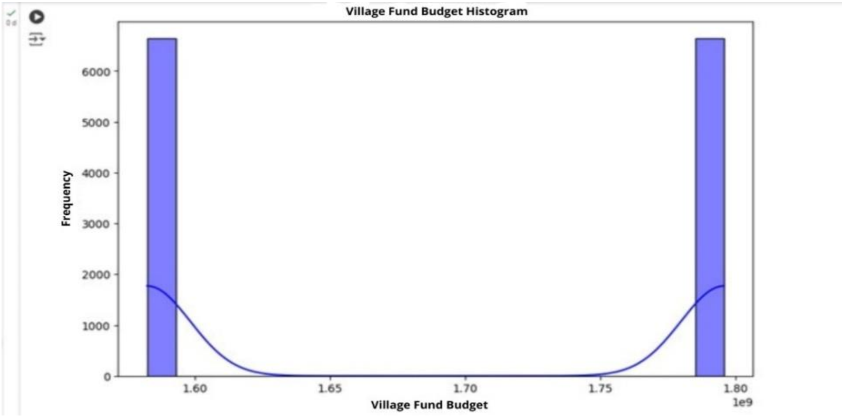
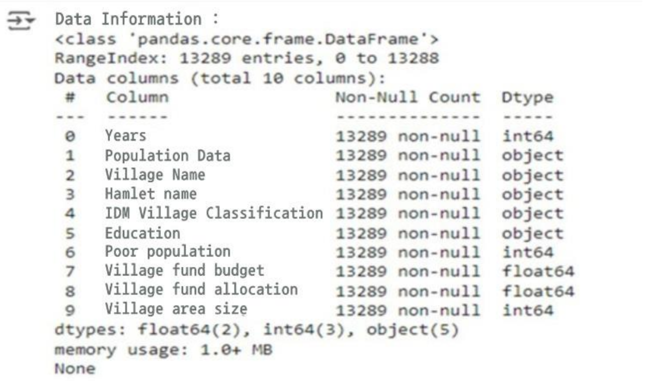


Figure 3. Histogram Graph

As seen in the image above, the histogram graph shows the distribution of the Village Fund budget based on its value. This histogram has an X-axis that shows the value of the Village Fund budget and a Y-axis that shows the frequency of villages with the Village Fund budget value.

4.3 Regression Model Development

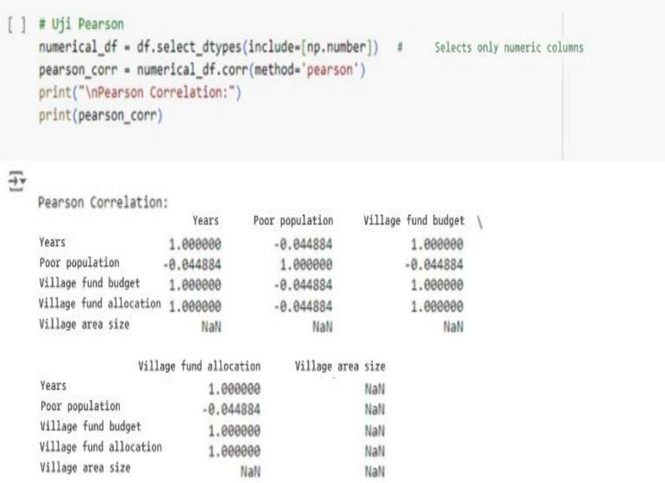
To be able to see the data set that will be used. The data details displayed are as follows:



```
Data Information :
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 13289 entries, 0 to 13288
Data columns (total 10 columns):
 #   Column                    Non-Null Count  Dtype
---  ------                    --------------  -----
 0   Years                     13289 non-null  int64
 1   Population Data           13289 non-null  object
 2   Village Name              13289 non-null  object
 3   Hamlet name               13289 non-null  object
 4   IDM Village Classification 13289 non-null  object
 5   Education                 13289 non-null  object
 6   Poor population           13289 non-null  int64
 7   Village fund budget       13289 non-null  float64
 8   Village fund allocation   13289 non-null  float64
 9   Village area size         13289 non-null  int64
dtypes: float64(2), int64(3), object(5)
memory usage: 1.0+ MB
None
```

Figure 4. Detailed Data Information

The information depicted in the image above is the data set to be tested, consisting of 13289 rows of data and 10 columns of variables. All data has no missing values and is entirely in the form of numeric variables (int64), because Multiple Linear Regression cannot process data in categorical form. The Pearson test is a type of correlation test used to assess the degree of relationship between 2 variables on an interval or ratio scale. The comes about of this test will deliver a relationship coefficient that can have values between -1, 0, and 1. A esteem of -1 demonstrates a idealize negative relationship, demonstrates no relationship, and a esteem of 1 demonstrates a culminate positive relationship. The extend of relationship coefficient values from -1, 0, and 1 can be concluded that the closer the esteem is to 1 or -1, the closer the relationship between factors. Whereas the closer to the esteem of 0, the weaker the relationship between factors.



```
[ ] # Uji Pearson
    numerical_df = df.select_dtypes(include=[np.number])  #    Selects only numeric columns
    pearson_corr = numerical_df.corr(method='pearson')
    print("\nPearson Correlation:")
    print(pearson_corr)
```

```
Pearson Correlation:
                          Years  Poor population  Village fund budget  \
Years                  1.000000        -0.044884             1.000000
Poor population       -0.044884         1.000000            -0.044884
Village fund budget    1.000000        -0.044884             1.000000
Village fund allocation 1.000000       -0.044884             1.000000
Village area size           NaN              NaN                  NaN

                   Village fund allocation  Village area size
Years                             1.000000                NaN
Poor population                  -0.044884                NaN
Village fund budget               1.000000                NaN
Village fund allocation           1.000000                NaN
Village area size                      NaN                NaN
```

Figure 5. Pearson test

To be able to see the correlation graph in Python, do the following:

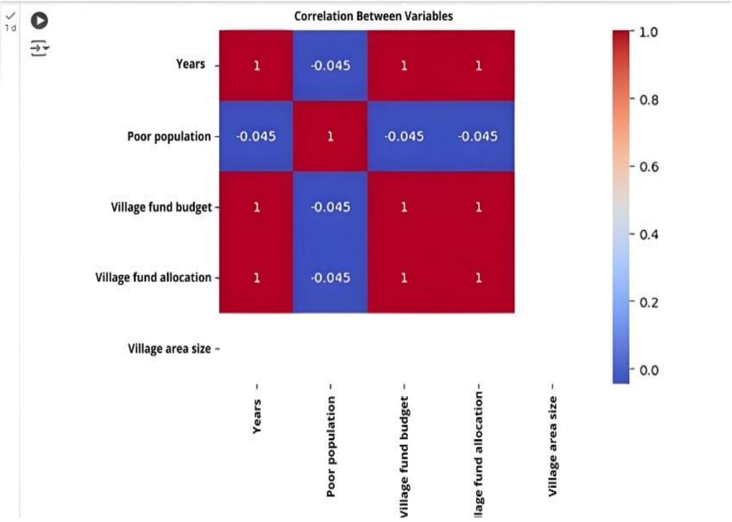


Figure 6. Correlation Test Graph

Judging from the image above, it shows the results of the correlation test between the independent variables (Village Fund Budget) and the dependent variables (Year, Poor Population, Village Fund Ceiling, and Village Area). The correlation coefficient value in the table ranges from -0.045 to 1.0. So with the results that can be seen in the image it can be said that there is no correlation between the independent variables of the village fund budget and the year, poor population, village fund ceiling, and village area. so that the regression process can be carried out, if there is a correlation between independent variables > 0.8, it can be indicated that there is a strong correlation and must be eliminated.

4.4 Data Analysis Results

From the results of the data analysis that has been carried out, there are several parameters of conclusions that can be drawn, including those in the table below:

Table 2. Data Analysis Results

Regression Equation	Conclusion
Village Fund Budget = -1609074668.034 + 6.5632 x Year - 2.6104 x Poor Population + 3.7433 x Village Fund Ceiling + 0.0 x Village Area	• If other variables are constant, the duration will change by itself by a constant value of -1609074668.034. • If other variables are constant, the village fund budget will change by 6.5632 per year. • If other variables are constant, the duration will change by -2.6104 per poor population. • If other variables are constant, the duration will change by 3.7433 per village fund budget. • If other variables are constant, the duration will change by 0.0 per village area.
Multicollinearity	Conclusion
-0.045	According to the correlation frequency table, there is no correlation between the independent variables.
(R^2)	Conclusion
1.0	The large influence of the independent variable on the dependent is 10%. By looking at the correlation table, the current is very low, while the remaining 90% is influenced by other factors.
Adjusted R^2	Conclusion
4,66	The average influence of independent variables on dependent is 50%. By looking at the correlation table, the current is the same, while the remaining 50% is influenced by other factors.
P Test (*P-Value*)	Conclusion
0,000	The P value is lower than the threshold of 0.05 or 5%, then the independent variables of disturbance and billing can be accepted.
Residual Test	Conclusion
Data distribution moves away from the diagonal line	Residual values are not normally distributed
Heteroscedasticity Test	Conclusion
The plot is spread unevenly	There is a difference in the residual value

5. CONCLUSSION

The results of the regression analysis show that there is a significant relationship between the village fund budget and several influencing factors, which are represented in the following regression equation: Village Fund Budget = -1609074668.034 + 6.5632 x Year - 2.6104 x Poor Population + 3.7433 x Village Fund Ceiling + 0.0 x Village Area The independent variables that have the most influence on the level of utilization of the village fund budget are population, area, and village fund ceiling. The study also found that the village fund budget has a significant impact on the level of utilization in Dolok Maraja Village, which is reinforced by the high coefficient of determination (R-squared) value of -0.045. This shows that the greater the allocation of the village fund budget given, the higher the level of utilization for village development. However, it should be noted that there is a multicollinearity problem that needs to be considered, as indicated by the Adjusted R-squared value approaching 1.0. This indicates a correlation between the independent variables, which can affect the interpretation of the regression results. In addition, the P-value test shows the statistical significance of the overall regression model with a very low value (0.000), and the residual test shows that the model has a low residual value, indicating a good level of model fit with the observation data.

REFERENCES

[1] M. Maulita and N. Nurdin, "Pendekatan Data Mining Untuk Analisa Curah Hujan Menggunakan Metode Regresi Linear Berganda (Studi Kasus: Kabupaten Aceh Utara)," *IDEALIS Indones. J. Inf. Syst.*, vol. 6, no. 2, pp. 99–106, 2023, doi: 10.36080/idealis.v6i2.3034.

[2] E. Yanti, M. Yetri, and F. Taufik, "Penerapan Data Mining Untuk Mengestimasi Biaya Pembangunan Di Desa Puang Aja Biaya Pembangunan Di Desa Puang Aja Regresi Linear Berganda," *J. CyberTech*, no. x, 2022, [Online]. Available: https://ojs.trigunadharma.ac.id/

[3] Daniel Kristian Sabar Nadeak, "Jurnal Sistem Informasi," vol. 3, no. 2, pp. 1–2, 2023.

[4] R. A. Prasetyo, "Analisis Regresi Linear Berganda Untuk Melihat Faktor Yang Berpengaruh Terhadap Kemiskinan di Provinsi Sumatera

Barat,” *J. Math. UNP*, vol. 7, no. 2, p. 62, 2022, doi: 10.24036/unpjomath.v7i2.12777.

[5] Tahyani, A. S. Sunge, and M. Wangsadanureja, “Penerapan Data Mining Untuk Mempermudah Produksi Diapers Dengan Menggunakan Algoritma Regresi Linier,” *ISSN 2962-3545 Pros. SAINTEK Sains dan Teknol. Vol.1 No.1 Tahun 2022 Call Pap. dan Semin. Nas. Sains dan Teknol. Ke-1 2022 Fak. Tek. Univ. Pelita Bangsa, Juli 2022*, vol. 1, no. 1, pp. 176–179, 2022, [Online]. Available: https://www.jurnal.pelitabangsa.ac.id/index.php/SAINTEK/article/view/1165/757

[6] A. Fitri Boy, “Implementasi Data Mining Dalam Memprediksi Harga Crude Palm Oil (CPO) Pasar Domestik Menggunakan Algoritma Regresi Linier Berganda (Studi Kasus Dinas Perkebunan Provinsi Sumatera Utara),” *J. Sci. Soc. Res.*, vol. 4307, no. 2, pp. 78–85, 2020, [Online]. Available: http://jurnal.goretanpena.com/index.php/JSSR

[7] B. Subandriyo, “Analisis kolerasi dan regresi,” *Diklat Stat. Tingkat Ahli BPS Angkatan XXI*, p. 31, 2020, [Online]. Available: https://pusdiklat.bps.go.id/diklat/bahan_diklat/BA_Analisis Korelasi dan Regresi_Budi Soebandriyo, SST, M. Stat_2123.pdf

[8] U. P. Duhok *et al.*, “Tinjauan Regresi Linier Komprehensif pada Mesin,” vol. 01, no. 04, pp. 140–147, 2020.

[9] S. Adiguno, Y. Syahra, and M. Yetri, “Prediksi Peningkatan Omset Penjualan Menggunakan Metode Regresi Linier Berganda,” *J. Sist. Inf. Triguna Dharma (JURSI TGD)*, vol. 1, no. 4, p. 275, 2022, doi: 10.53513/jursi.v1i4.5331.

[10] Arima Andhika Ayu, “JEKAWAL KABUPATEN SRAGEN DI ERA PANDEMI COVID-19,” pp. 551–566, 2020.

[11] T. A. Robin *et al.*, “Using spatial analysis and GIS to improve planning and resource allocation in a rural district of Bangladesh,” *BMJ Glob. Heal.*, vol. 4, 2019, doi: 10.1136/bmjgh-2018-000832.

[12] B. R. Sari, “Pengelolaan Keuangan Desa Ditinjau Dari Undang-Undang Desa Menuju Masyarakat Yang Mandiri,” *J. Lex Renaiss.*, vol. 5, no. 2, pp. 488–507, 2020, doi: 10.20885/jlr.vol5.iss2.art15.

[13] R. Zakariya, “Partisipasi Masyarakat dalam Pencegahan Korupsi Dana Desa: Mengenali Modus Operandi,” *INTEGRITAS J. Antikorupsi*, vol. 6, no. 2, pp. 263–282, 2020, doi: 10.32697/integritas.v6i2.670.