The 2nd International Conference on Multidisciplinary Engineering
(2nd ICOMDEN 2024)

The 2nd International Conference on Multidisciplinary Engineering: 00022 (2024) https://doi.org/10.29103/icomden.v2.xxxx
Volume 2, 2024
eISSN: xxxx-xxxx
Research Original Article

Application Of Data Mining For Classification Of Blt-Dd Recipients Using The Support Vector Machine Method

Meliza Putri[1] **Bustami**[2] **Ar Razi**[3]
[1]Department Of Informatics, Universitas Malikussaleh, Bukit Indah, Lhokseumawe, 24353, Indonesia, meliza.200170070@mhs.unimal.ac.id
[2]Department Of Informatics, Universitas Malikussaleh, Bukit Indah, Lhokseumawe, 24353, Indonesia, bustami@unimal.ac.id
[3]Department Of Informatics, Universitas Malikussaleh, Bukit Indah, Lhokseumawe, 24353, Indonesia, arrazi@unimal.ac.id
Corresponding Author: **Meliza Putri | Phone: +6282272037915**

Abstract
This study focuses on the application of the Support Vector Machine (SVM) algorithm for classifying recipients who are eligible to receive Direct Cash Assistance from the Village Fund (BLT-DD) in Nurussalam District, East Aceh Regency. The background of this research is the difficulty in identifying households eligible for BLT-DD due to increasing poverty and economic inequality exacerbated by the COVID-19 pandemic. This study aims to address this issue by utilizing the SVM algorithm, which can separate household data into two categories: "Eligible" and "Not Eligible." A total of 550 data points from Nurussalam District were used in this study, with 400 data points for training and 150 data points for testing. In the training data, 322 households (80.5%) were classified as "Eligible," while 78 households (19.5%) were categorized as "Not Eligible." The data collected includes variables such as household income, type of employment, education level, history of chronic disease, and home ownership status. After preprocessing the data, such as normalization and encoding, the SVM model was trained to classify BLT-DD recipients. In the testing data, 128 data points (85.33%) were classified as "Eligible," while 22 data points (14.67%) were classified as "Not Eligible." Further analysis of village distribution in Nurussalam District shows that some villages have a high percentage of eligible recipients, such as Blang Rambong and Alue Jagat, with 100% of recipients classified as "Eligible." Other villages, such as Arul Pinang and Alue Dua Muka O, show more varied eligibility rates, with 71.43% and 72.73% classified as "Eligible," respectively. In conclusion, the SVM algorithm provides an effective approach in determining the eligibility of BLT-DD recipients, helping the government to distribute assistance more accurately and efficiently in Nurussalam District.
Keywords: Support Vector Machine, BLT-DD, classification, socio-economic data, data mining

Introduction

overty is a fundamental issue that consistently captures global attention. Many factors can trigger poverty, both directly and indirectly. These factors include limited job opportunities, lack of education, restricted access to resources, unequal economic conditions, low wages, and the impact of the Covid-19 pandemic, which has led to an increase in unemployment and poverty. The low quality of life among the poor can result in lower levels of education and health, negatively affecting productivity. The existence of the BLT-DD assistance program has a significant impact on maintaining the welfare of the community, which is why the researcher is interested in exploring the implementation of BLT-DD recipients more deeply [1].

Several previous studies have focused on recipients of assistance using various classification methods. The use of Support Vector Machine (SVM) for classification is particularly suitable due to its accurate results. Other studies have shown that SVM is rooted in statistical learning theory, yielding promising results that can outperform other methods. Additionally, SVM works well with high-dimensional datasets, as it uses kernel techniques to map original data from its original dimensions to relatively higher dimensions n [2].

This research is essential due to the increase in unemployment caused by the pandemic, which has led to a rise in poverty. Several criteria are considered: monthly income, employment, chronic illness, home ownership, loss of livelihood, vulnerable families, receipt of assistance, single elderly individuals, and female heads of households. The goal is to create technology capable of classifying eligible and ineligible recipients. Furthermore, this method is based on statistical learning concepts and has been proven effective in various applications, from recognizing handwritten digits to classification tasks [3].

Literature Review

Direct Cash Assistance (BLT)

The origin of BLT-DD (Direct Cash Assistance from Village Funds) can be traced back to Brazil in the 1990s, initiated by Luiz Inácio Lula da Silva, the 35th president of Brazil. This program was initially called Bolsa Escola, but was later renamed Bolsa Família because the original name was deemed less relevant for a conditional cash assistance program. Indonesia first recognized the term BLT in 2005, which continued in 2009, and in 2013 it was renamed "Temporary Direct Cash Assistance for Communities" (BLSM). The BLT program was established in response to the increase in Fuel Oil (BBM) at that time, with the aim of assisting the poor [4].

Village Fund

The Village Fund is a government program in Indonesia, sourced from the State Budget (APBN), which is transferred through the Regional Revenue and Expenditure Budget of the district/city (APBD). This fund is used to finance village government, develop villages, and empower all levels of village communities. The Village Fund is money received by villages each year from the APBN, specifically allocated for villages by transferring it directly through the district/city APBD, which is used to finance all processes related to government affairs or village development and to empower all village communities [5].

Support Vector Machine (SVM) Algorithm

Support Vector Machine (SVM) was first introduced by Vapnik in 1992 as a machine learning method. The goal of this method is to find the best hyperplane that separates two classes of data or predicts target values in the input space. By using the kernel trick, SVM can work with high-dimensional datasets. This method is rooted in statistical learning theory, where its results are very promising for achieving better outcomes than other methods. Additionally, SVM performs well on high-dimensional datasets; even SVM using kernel techniques must map the original data from its initial dimensions to another relatively higher dimension [6].

The basic mathematical formula for the hyperplane in SVM is as follows:

$$\mathbf{W \cdot X + b = 0} \qquad (1)$$

Where:
W is the weight vector.
X is the input vector.
b is the bias.
The decision function for classification using linear SVM is:

$$\mathbf{f(X) = sign(W \cdot X + b)} \qquad (2)$$

For non-linear SVM, we use the kernel trick, and the decision function becomes:

$$\mathbf{f(X) = sign(\sum i = 1N\alpha iyiK(Xi, X) + b)} \qquad (3)$$

Where:
K(Xi,X) is the kernel function.
αi is the Lagrange coefficient obtained from the optimization process.
yi is the class label of the iii-th input vector.
N is the number of support vectors.

Research Methodology

This research utilizes the SVM Method. This research methodology consists of several stages, including data preprocessing, hyperparameter optimization, model validation, and evaluation metrics. Overall, the methods used can be explained as follows:

Problem and Needs Analysis

At this stage, an analysis is conducted on the system requirements to collect data, train and test the SVM model, as well as store and display the results. Non-functional requirements focus on aspects of security, performance, scalability, reliability, and user experience. The dataset used in this study consists of the names of the BLT recipients. This dataset will be processed and divided into training and testing data.

System Architecture Design

A system is a network of interrelated procedures that come together to carry out activities or achieve specific goals. A system is a series of data or more components that are interconnected and interact to reach objectives. The system is an organized and procedurally connected unit of data. This system functions as a web-based application that allows the admin to train and test the SVM model using the training and testing data stored in the database. All essential processes, including data retrieval, data preprocessing, model training, model evaluation, model testing, and result storage, are managed through interactions between the web module, database, and SVM model, with a user-friendly interface for the admin to view the results.

System Schema

The following is the system schema for this research:

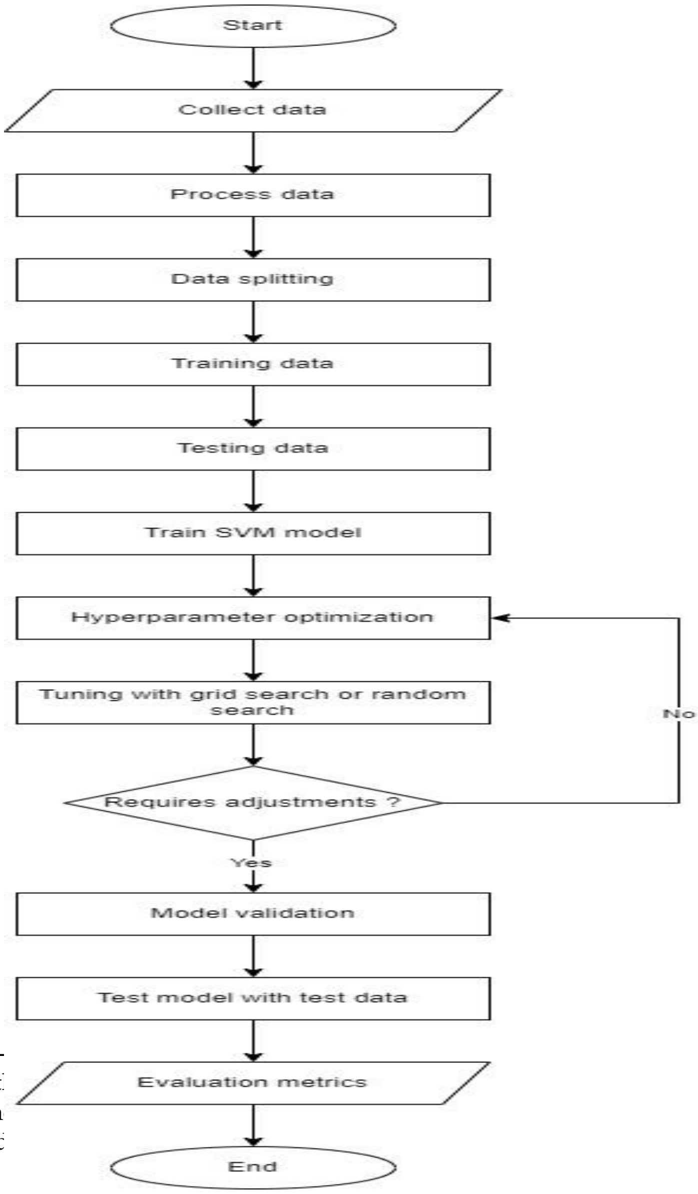

Figure 1. System Schema

Materials & Methods

This system functions as a web- ... SVM model using training and testing data stored in t ... ata preprocessing, model training, model evaluation, m ... ction between the web module, database, and SVM moc

Calculation of the Support Vector Machine (SVM) Algorithm

The steps involved in applying the Support Vector Machine (SVM) for classifying recipients of BLT-DD generally include several phases, starting from data collection to model evaluation. The data used is as follows:

1. Testing Test Data

To calculate predictions using the test data based on the trained model, we will follow these steps one by one. We will process the test data, compute the decision function $f(x_{uji})$, and determine the class based on these calculations.

Test Data (Before Encoding)

Table 1. Test Data

ID	Name	Income	Occupation	Chronic Illness History	Education Level	Home Ownership Status
25	Muhammad Zikri	1,300,000	Farmer	No	Junior High	Rent
26	Rahmatullah Husaini	2,300,000	Laborer	No	Junior High	Own
27	Cut Aisyah	2,500,000	Private Employee	Yes	High School	Own
28	Zulfikar Darmawan	1,200,000	Laborer	No	Elementary	Rent
29	Siti Khadijah	1,500,000	Farmer	No	Elementary	Own
30	Teuku Fadhil	2,100,000	Farmer	Yes	High School	Own

Table 2. Test Data (Continued)

ID	Name	Job Loss	Vulnerable Member	Not Receiving Assistance	Single Elderly Household	Head of Family Female
25	Muhammad Zikri	Yes	No	Yes	Yes	No
26	Rahmatullah Husaini	No	No	Yes	Yes	No
27	Cut Aisyah	Yes	Yes	No	Yes	No
28	Zulfikar Darmawan	No	Yes	No	Yes	No
29	Siti Khadijah	Yes	Yes	Yes	Yes	No
30	Teuku Fadhil	No	No	No	Yes	No

2. Normalization and Feature Encoding

Table 3. Normalization and Encoding

ID	Name	Income (normalized)	Type of Job	History of Chronic Illness	Education Level	Home Ownership Status
25	Muhammad Zikri	0.07	0	0	0	0
26	Rahmatullah Husaini	0.84	1	0	0	1
27	Cut Aisyah	1.0	0	1	0	1
28	Zulfikar Darmawan	0.76	1	0	0	0
29	Siti Khadijah	0.23	0	0	0	1
30	Teuku Fadhil	0.69	0	1	0	1

Tabel 4. Normalisasi Dan Encoding (Lanjutan)

ID	Nama	Kehilangan Pekerjaan	Anggota Rentan	Tidak Menerima Bantuan	Lansia Tunggal	Perempuan Kepala Keluarga
25	Muhammad Zikri	0	0	1	1	0
26	Rahmatullah Husaini	0	0	1	1	0
27	Cut Aisyah	1	1	0	1	0
28	Zulfikar Darmawan	0	1	0	1	0
29	Siti Khadijah	1	1	1	1	0
30	Teuku Fadhil	0	0	0	1	0

Weights (w) and Bias (b) Used:
Weights (w):

$$\mathrm{w} = (-2.205, 1.0, 1.0, -2.0, -0.5, -1.0, -1.0, 0, 0, -0.5)$$

Bias (b):

$$b = 1.5824$$

3. Manual Calculation of the Decision Function
Data 25 (Muhammad Zikri):

$$\mathbf{x}_{\mathrm{uji},25} = [0.07,0,0,0,0,0,0,0,1,1,0]$$

$$\mathbf{w}^T\mathbf{x}_{\mathrm{uji},25} = (-2.205 \times 0.07) + (1 \times 0) + (1 \times 0) + (-2 \times 0) + (-0.5 \times 0) + (-1 \times 0) + (-1 \times 0) + (0 \times 0) + (0 \times 1) + (-0.5 \times 0)$$

$$\mathbf{w}^T\mathbf{x}_{\mathrm{uji},25} = -0.15435 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 = -0.15435$$

$$f(\mathbf{x}_{\mathrm{uji},25}) = -0.15435 + 1.5824 = 1.42805$$

Data 26 (Rahmatullah Husaini):

$$\mathbf{x}_{\mathrm{uji},26} = [0.84,1,0,0,1,0,0,1,1,0]$$

$$\mathbf{w}^T\mathbf{x}_{\mathrm{uji},26} = (-2.205 \times 0.84) + (1 \times 1) + (1 \times 0) + (-2 \times 0) + (-0.5 \times 1) + (-1 \times 0) + (-1 \times 0) + (0 \times 1) + (0 \times 1) + (-0.5 \times 0)$$

$$\mathbf{w}^T\mathbf{x}_{\mathrm{uji},26} = -1.8522 + 1 + 0 + 0 - 0.5 + 0 + 0 + 0 + 0 + 0 = -1.3522$$

$$f(\mathbf{x}_{\mathrm{uji},26}) = -1.3522 + 1.5824 = 0.2302$$

Data 27 (Cut Aisyah):

$$\mathbf{x}_{\mathrm{uji},27} = [1.0,0,1,0,1,1,1,0,1,0]$$

$$\mathbf{w}^T\mathbf{x}_{\text{uji},27} = (-2.205 \times 1.0) + (1 \times 0) + (1 \times 1) + (-2 \times 0) + (-0.5 \times 1) + (-1 \times 1) + (-1 \times 1) + (0 \times 0) + (0 \times 1) + (-0.5 \times 0)$$
$$\mathbf{w}^T\mathbf{x}_{\text{uji},27} = -2.205 + 0 + 1 + 0 - 0.5 - 1 - 1 + 0 + 0 + 0 = -3.705$$
$$f(\mathbf{x}_{\text{uji},27}) = -3.705 + 1.5824 = -2.1226$$

Data 28 (Zulfikar Darmawan):

$$\mathbf{x}_{\text{uji},28} = [0.76,1,0,0,0,0,1,0,1,0]$$
$$\mathbf{w}^T\mathbf{x}_{\text{uji},28} = (-2.205 \times 0.76) + (1 \times 1) + (1 \times 0) + (-2 \times 0) + (-0.5 \times 0) + (-1 \times 0) + (-1 \times 1) + (0 \times 0) + (0 \times 1) + (-0.5 \times 0)$$
$$\mathbf{w}^T\mathbf{x}_{\text{uji},28} = -1.6758 + 1 + 0 + 0 + 0 + 0 - 1 + 0 + 0 + 0 = -1.6758$$
$$f(\mathbf{x}_{\text{uji},28}) = -1.6758 + 1.5824 = -0.0934$$

Data 29 (Siti Khadijah):

$$\mathbf{x}_{\text{uji},29} = [0.23,0,0,0,1,1,1,1,1,0]$$
$$\mathbf{w}^T\mathbf{x}_{\text{uji},29} = (-2.205 \times 0.23) + (1 \times 0) + (1 \times 0) + (-2 \times 0) + (-0.5 \times 1) + (-1 \times 1) + (-1 \times 1) + (0 \times 1) + (0 \times 1) + (-0.5 \times 0)$$
$$\mathbf{w}^T\mathbf{x}_{\text{uji},29} = -0.50715 + 0 + 0 + 0 - 0.5 - 1 - 1 + 0 + 0 + 0 = -3.00715$$
$$f(\mathbf{x}_{\text{uji},29}) = -3.00715 + 1.5824 = -1.42475$$

Data 30 (Teuku Fadhil):

$$\mathbf{x}_{\text{uji},30} = [0.69,0,1,0,1,0,0,0,1,0]$$
$$\mathbf{w}^T\mathbf{x}_{\text{uji},30} = (-2.205 \times 0.69) + (1 \times 0) + (1 \times 1) + (-2 \times 0) + (-0.5 \times 1) + (-1 \times 0) + (-1 \times 0) + (0 \times 0) + (0 \times 1) + (-0.5 \times 0)$$
$$\mathbf{w}^T\mathbf{x}_{\text{uji},30} = -1.52145 + 0 + 1 + 0 - 0.5 + 0 + 0 + 0 + 0 + 0 = -1.02145$$
$$f(\mathbf{x}_{\text{uji},30}) = -1.02145 + 1.5824 = 0.56095$$

The following are the classification results based on the decision function f(x) :

$$\text{Category} = \begin{cases} \text{"Eligible"} & \text{if } f(\mathbf{x}) \geq 0 \\ \text{"Not Eligible"} & \text{if } f(\mathbf{x}) < 0 \end{cases}$$

Explanation:
Eligible: If the decision function value ($f(\mathrm{x})$) is greater than or equal to $0(f(\mathrm{x}) \geq 0)$ the data is labeled as "Eligible."
Not Eligible: If the decision function value is less than $0(f(\mathrm{x}) < 0$, the data is labeled as "Not Eligible."
Here are the results:

Table 5. Classification Results of Decision Function

ID	Name	f(x)	Category
1	Muhammad Zikri	1.42805	Eligible
2	Rahmatullah Husaini	0.2302	Eligible
3	Cut Aisyah	-2.1226	Not Eligible
4	Zulfikar Darmawan	-0.0934	Not Eligible
5	Siti Khadijah	-142475	Not Eligible
6	Teuku Fadhil	0.56095	Eligible

Conclusion:
Based on the calculated decision function results, the final categories for each individual are as follows:
Eligible: Muhammad Zikri, Rahmatullah Husaini, Teuku Fadhil
Not Eligible: Cut Aisyah, Zulfikar Darmawan, Siti Khadijah

4. Comparison of Actual Data with Predictions:

Table 6. Comparison Results

ID	Name	Actual Category	Predicted Category
1	Muhammad Zikri	Layak	Layak
2	Rahmatullah Husaini	Tidak Layak	Layak
3	Cut Aisyah	Tidak Layak	Tidak Layak
4	Zulfikar Darmawan	Layak	Tidak Layak
5	Siti Khadijah	Tidak Layak	Tidak Layak
6	Teuku Fadhil	Tidak Layak	Layak

Comparison Analysis:

1. Correct Predictions (Predictions matching Actual Categories):
   a. ID 25 (Muhammad Zikri): Actual = Eligible, Predicted = Eligible
   b. ID 27 (Cut Aisyah): Actual = Not Eligible, Predicted = Not Eligible
   c. ID 29 (Siti Khadijah): Actual = Not Eligible, Predicted = Not Eligible
2. Not Matched (Predicted differs from Actual Category):
   a. ID 26 (Rahmatullah H): Actual = Not Eligible, Predicted = Eligible
   b. ID 29 (Zulfikar Darmawan): Actual = Eligible, Predicted = Not Eligible
   c. ID 30 (Teuku Fadhil): Actual = Not Eligible, Predicted = Eligible

Conclusion:
From the 6 test data provided:
3 data points match the predicted results with the actual categories.
3 data points do not match the predicted results with the actual categories.

Category	Count	Percentage (%)
Matched	3	50.00
Not Matched	3	50.00

Conclusions

Based on the classification results using the Support Vector Machine (SVM), a significant majority of the training data, specifically 90.6% or around 135 data points, were categorized as 'Eligible'. This model demonstrates good performance in identifying eligible recipients of assistance. However, there are still several data points near the hyperplane, which leads to some overlap between the 'Eligible' and 'Ineligible' categories. Approximately 9.4% of the data (about 14 data points) were classified as 'Ineligible'. This indicates that there are certain data points that are more difficult to predict accurately, possibly due to characteristics that are not sufficiently represented in the feature space used by the current model. It may be necessary to add new features or employ more complex classification techniques to reduce the overlap in these harder-to-identify data points. Further analysis at the village level shows significant variation in the distribution of eligibility for assistance. Some villages, such as Keude Bagok Dua and Bantayan, exhibit uniform eligibility classifications, with the majority of recipients in these villages categorized as 'Eligible' (96.00% and 92.00%).

References

[1] Amin, H. (2022). Pengelolaan Alokasi Dana Desa (ADD) Terhadap Pembangunan Gampong. Journal of Social and Policy Issues, 2, 71–76. https://doi.org/10.58835/jspi.v2i2.42

[2] Arief, M. (2019). Klasifikasi Kematangan Buah Jeruk Berdasarkan Fitur Warna Menggunakan Metode SVM. Jurnal Ilmu Komputer Dan Desain Komunikasi Visual, 4(1), 9–16.

[3] Ha, J., Kambe, M., & Pe, J. (2011). Data Mining: Concepts and Techniques. In Data Mining: Concepts and Techniques. https://doi.org/10.1016/C2009-0- 61819-5

[4] Hariandja, T. R., & Budiman, N. T. (2020). Pelaksanaan Bantuan Langsung. Indonesian Journal of Law and Islamic Law (IJLIL), 2, 271–293.

[5] Nurdin, & Astika, D. (2015). Penerapan Data Mining Untuk Menganalisis Penjualan Barang Dengan Pada Supermarket Sejahtera Lhokseumawe. 6(1), 134–155. https://doi.org/10.29103/TECHSI.V7I1.

[6] Salam, R. R., Jamil, M. F., Ibrahim, Y., Rahmaddeni, R., Soni, S., & Herianto, H. (2023). Analisis Sentimen Terhadap Bantuan Langsung Tunai (BLT) Bahan Bakar Minyak (BBM) Menggunakan Support Vector Machine. MALCOM: Indonesian Journal of Machine Learning and Computer Science, 3(1), 27–35. https://doi.org/10.57152/malcom.v3i1.590

[7] Sallaby, A. F., & Kanedi, I. (2020). Perancangan Sistem Informasi Jadwal Dokter Menggunakan Framework Codeigniter. Jurnal Media Infotama, 16(1), 48–53. https://doi.org/10.37676/jmi.v16i1.1121

[8] Sari, L. P., Kadir, A., Marbun, U., Tinggi, S., Kesehatan, I., & Hasanuddin, N. (2024). Faktor yang Mempengaruhi Minat Keluarga dalam Memilih Pelayanan Home Care pada Bayi. 19(1), 316–324.

[9] Sutanto, H., & Hardiningsih, P. (2021). Akuntabilitas Pengelolaan BLT-Dana Desa Pada Masa Pandemi Covid-19. InFestasi, 17(1), InPres. https://doi.org/10.21107/infestasi.v17i1.9932

[10] Suwanda, R., Efendi, S., & Budhiarti, E. (2019). Kinerja Algoritma Support Vector Machine Berdasarkan Ekstraksi Sinyal Suara Dengan Mel-Frequency Cepstrum Coefficients Pada Pelafadzan Ayat Al-Quran. 4(2), 2502–7131.

[11] Syaputra, A., Muslim, B., Prawira, N. S., & Edowinsyah. (2023). Implementasi Metode Support Vector Machine Dengan Algoritma Genetika Pada Prediksi Konsumsi Energi Untuk Gedung Beton Bertulang. Faktor Exacta, 16(2), 142– 153. https://doi.org/10.30998/faktorexacta.v16i2.16657

[12] Utomo, D. P., & Mesran, M. (2020). Analisis Komparasi Metode Klasifikasi Data Mining dan Reduksi Atribut Pada Data Set Penyakit Jantung. Jurnal Media Informatika Budidarma, 4(2), 437. https://doi.org/10.30865/mib.v4i2.2080

[13] Widyawati, R., Suyono, E., Hasanah, U., & Setiawan, W. (2022). Pengaruh Penyaluran Bantuan Langsung Tunai (Blt), Persepsi Status Wilayah, Dan Standar Hidup Terhadap Kesejahteraan Masyarakat Desa Baseh, Kecamatan Kedungbanteng, Kabupaten Banyumas. *MidYear National Conference and Call for Paper*, *1*(1), 723–735.

[14] Yeni Kustiyahningsih, D. R. A. (2011). Pemrograman Basis Data Berbasis Web Menggunakan PHP & MySQL. Yogyakarta: Graha Ilmu, 8–33.