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

The 2nd International Conference on Multidisciplinary Engineering: 00018 (2024) https://doi.org/10.29103/icomden.v2.xxxx
Volume 2, 2024
eISSN: xxxx-xxxx
Research Original Article

The Implementation of Support Vector Machine to Analyze Compliance of Land and Building Taxpayers

Nurul Nafisa[✉1] **Rozzi Kesuma Dinata**[2] **Rizki Suwanda**[3]
[1]Department of Informatic, Universitas Malikussaleh, Bukit Indah, Lhokseumawe, 24353, Indonesia, nurul.200170047@mhs.unimal.ac.id
[2]Department of Informatic, Universitas Malikussaleh, Bukit Indah, Lhokseumawe, 24353, Indonesia, rozzi@unimal.ac.id
[3]Department of Informatic, Universitas Malikussaleh, Bukit Indah, Lhokseumawe, 24353, Indonesia, rizkisuwanda@unimal.ac.id
✉Corresponding Author: **nurul.200170047@mhs.unimal.ac.id | Phone: +6282272627986**

Abstract
Land and Building Tax (LBT) is an important source of revenue for local governments, supporting development and community welfare. However, low taxpayer compliance rates often pose a challenge in achieving the targets for Local Own-Source Revenue (LOSR). This study aims to develop a data-driven classification system to map areas with varying levels of LBT taxpayer compliance in Lhokseumawe City and to implement the Support Vector Machine (SVM) method to improve the accuracy of predicting taxpayer compliance. The research data was obtained from the Regional Financial Management Agency (RFMA) of Lhokseumawe City, encompassing LBT data from 2021 to 2023, with variables such as principal amount, penalties, total payments, due dates, and payment dates. This classification system divides taxpayers into two categories: Compliant and Non-Compliant. The results of testing the SVM model indicate that Banda Sakti sub-district has a compliance rate of 98%, Muara Satu has a compliance rate of 99%, Muara Dua has a compliance rate of 99%, and Blang Mangat has a compliance rate of 100%. The accuracy metrics from the implementation of the Support Vector Machine method for assessing land and building tax compliance show a Precision of 86%, a Recall of 100%, and an Accuracy of 86%. By applying the SVM method, it is hoped that there will be an increase in efficiency in the tax collection and management processes, thereby optimally increasing Local Own-Source Revenue (LOSR) and supporting better regional development.
Keywords: Taxpayer Compliance, Land and Building Tax, Support Vector Machine

Introduction
Taxes are mandatory contributions from citizens to the government, which uses these funds for national interests and public welfare. Taxes take various forms, such as the Motor Vehicle Tax (MVT) and the Land and Building Tax (LBT). The Land and Building Tax (LBT) is imposed on the public and is a significant source of government revenue, contributing to the national income that supports development and enhances societal well-being [1]. Tax payments by citizens have a substantial impact on increasing Local Own-Source Revenue (LOSR). Managed by the Regional Financial Management Agency (RFMA), RFMA reflects income from regional resources. Higher tax revenue contributes to a more optimal level of regional income. However, RFMA targets are frequently unmet due to low public awareness of tax obligations, which can hinder regional governments in revenue management and impede local development.

In Lhokseumawe City in 2020, RFMA realization from LBT amounted to only IDR 3.99 billion, against a target of IDR 4.52 billion. This underachievement highlights the need to improve public compliance in tax payments, as non-compliance negatively impacts regional revenue management and reduces the effectiveness of regional development.

A previous study titled Application of Data Mining using Classification Techniques to Assess Potential Land and Building Taxpayer Compliance by Arif Rahmat Shaumi et al. used a dataset of 200 samples with three classifications: Very Punctual, Punctual, and Not Punctual. The study achieved an accuracy rate of 99.33% across 1,647 taxpayers, with non-compliance potentials of 0.437 in Medan Amplas District and 0.229 in Medan Area District [2].

A previous study titled "Support Vector Machine Algorithm for Classifying Political Attitudes Toward Indonesian Political Parties" used two data categories—positive and negative—with a training dataset of 900 entries. This study applied the Support Vector Machine (SVM) method in conjunction with K-Fold Cross Validation to evaluate the model's accuracy. The results showed that the SVM method achieved an accuracy rate of 86%, while K-Fold Cross Validation produced an accuracy rate of 71% with an error rate of 29% [3].

A related study titled "Classification of Hate Speech on Twitter Using Support Vector Machine" utilized a dataset of 700 entries for training and 300 entries for testing. This study experimented with several kernel functions, including RBF, Linear, and Sigmoid kernels. Each kernel function produced different accuracy levels: the RBF kernel achieved an accuracy of 93%, while the Linear and Sigmoid kernels both achieved 92%. These results indicate that using the SVM method for classifying hate speech on Twitter is highly effective, with a high accuracy rate of 93%, a precision of 84%, and a recall of 86% [4].

Based on these issues, there is a need to classify districts by their compliance levels. This study applies the Support Vector Machine (SVM) method to assess land and building taxpayer compliance and to evaluate the accuracy of the SVM method in this context.

Literature Review

1. Land and Building Tax (LBT)

The Land and Building Tax (LBT) is a tax imposed on individuals or entities that benefit from land and buildings. Although the LBT was initially managed by the central government, recent amendments to the law have allowed local governments to administer this tax as well [5]. The Land and Building Tax is collected by local governments and serves as a significant source of revenue derived from land and building assets. LBT plays a vital role in supporting regional development [6]. According to Sainang and Aji (2021), the Land and Building Tax (LBT) is levied on individuals who benefit from land and buildings, with the revenue generated from this tax being utilized for national interests. This includes funding for infrastructure projects such as improving highways, constructing government buildings, and building or repairing bridges, among other developments [7].

2. Taxpayer Compliance

Taxpayer compliance refers to the adherence to tax regulations in accordance with the law, aimed at enhancing the country's economic growth [8]. One significant factor influencing taxpayer compliance is the public's understanding of tax obligations. The greater the public's understanding of tax responsibilities, the higher the level of taxpayer compliance [9].

3. Support Vector Machine (SVM)

Support Vector Machine (SVM) was first introduced by Vapnik in 1992 as a prominent concept in the field of pattern recognition, which involves mapping data to predefined classes or categories [10]. SVM is a supervised learning method that analyzes data and recognizes patterns, and it is used for classification and regression tasks. The SVM algorithm works by identifying the optimal hyperplane that maximizes the distance between classes [11][12].

$$f(x) = w_1 * x_1 + \cdots + wn * xn + b \qquad (1)$$

Description:
$f(x)$ = Predicted value
w = Weight value
x = Input value
b = Bias value

Here is the formula for Support Vector Machine using the Gradient Descent algorithm:

$$w_j = w_j + \eta \times \left(y_i - f(x_i)\right) \times x_{ij} \qquad (2)$$

$$b = b + \eta \times \left(y_i - f(x_i)\right) \qquad (3)$$

Description:
η = Learning rate (e.g., $\eta = 0{,}01$)
y_i = Actual target class value
$f(x_i)$ = Predicted result
x_{ij} = j-th feature of the i-th data point

4. Confusion Matrix

The confusion matrix is a model evaluation method that estimates correct and incorrect data predictions. It provides deeper insights into errors, weaknesses, and model performance. The confusion matrix is typically a table that displays the actual results compared to the predicted outcomes generated by the algorithm under study [13][14].

Materials & Methods

1. Data Collection

Data was obtained from the Lhokseumawe City RFMA office, totaling 20,054 records. After the pre-processing stage, 19,340 records were available for processing. This calculation will yield outputs indicating compliance or non-compliance, as well as the accuracy level of the SVM method used.

Table 1. Data for Classification Process

No.	YEAR	PRINCIPAL	PENALTY	TOTAL	PAYMENT DATE	DUE DATE	LEVEL OF COMPLIANCE	SUB-DISTRICT
1	2021	133.219	0	133.219	04/01/2021	September	COMPLIANT	BANDA SAKTI
2	2021	98.458	0	98.458	5/24/2021	September	COMPLIANT	BANDA SAKTI
3	2021	98.458	0	98.458	9/27/2021	September	COMPLIANT	BANDA SAKTI
4	2021	98.458	0	98.458	9/27/2021	September	COMPLIANT	BANDA SAKTI
5	2021	233.541	4.671	238.212	10/28/2021	September	NON-COMPLIANT	BANDA SAKTI
…	…	…	…	…	…	…	…	…
19340	2023	667.300	0	667.300	7/31/2023	September	COMPLIANT	BLANG MANGAT

Table 1 presents the complete data that will be used in the classification process using the SVM method. A total of 19,340 records will be used for training, and 3,868 records will be used for testing, which will be taken from the normalized training data.

2. Methods

Below is the system scheme from the research conducted.

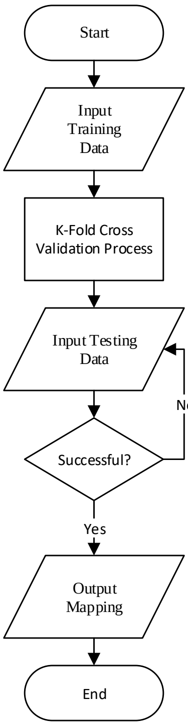

Figure 1. System Scheme

Figure 1 illustrates the system scheme for assessing land and building taxpayer compliance. The first step is the input of training data. The training data is input in the form of an Excel file and then trained. Next, the accuracy of the input training data will be assessed using K-Fold Cross Validation. After determining the accuracy of the training data, the process will continue with the testing data. The testing data is inputted, and testing is performed. If the testing process is successful, the results will be displayed in the form of mapping. However, if it is unsuccessful, the testing process should be repeated. The application of the Support Vector Machine (SVM) method results in two classes: Compliant and Non-Compliant.

3. Evaluation

To calculate the accuracy between the actual results and the predicted outcomes, the evaluation stage can use a

Confusion Matrix. Table 2 presents the Confusion Matrix for calculating the accuracy of the research results.

Table 2. Confusion Matrix

	Positive Prediction	Negative Prediction
Actual Positive	True Positive (TP)	False Negative (FN)
Actual Negative	False Positive (FP)	True Negative (TN)

Results and Discussion

Table 3 presents the training data that will be used for the SVM calculation process.

Table 3. Training Data

No.	YEAR	PRINCIPAL	PENALTY	TOTAL	PAYMENT DATE	DUE DATE	LEVEL OF COMPLIANCE	SUB-DISTRICT
1	2021	133.219	0	133.219	04/01/2021	September	COMPLIANT	BANDA SAKTI
2	2021	98.458	0	98.458	5/24/2021	September	COMPLIANT	BANDA SAKTI
3	2021	98.458	0	98.458	9/27/2021	September	COMPLIANT	BANDA SAKTI
4	2021	98.458	0	98.458	9/27/2021	September	COMPLIANT	BANDA SAKTI
5	2021	233.541	4.671	238.212	10/28/2021	September	NON-COMPLIANT	BANDA SAKTI
…	…	…	…	…	…	…	…	…
19340	2023	667.300	0	667.300	7/31/2023	September	COMPLIANT	BLANG MANGAT

In the training data available, only three parameters related to tax compliance are necessary: the principal, fine, and total columns. Then, in the application of Support Vector Machine, we need to convert the target parameter, which is the compliance level, into calculable values. In this study, a conversion value of 1 is assigned to the compliant category and -1 to the non-compliant category. Therefore, the data used for training can be found in Table 4.

Table 4. Conversion Data Labels

PRINCIPAL	PENALTY	TOTAL	COMPLIANCE'
133.219	0	133.219	1
98.458	0	98.458	1
98.458	0	98.458	1
98.458	0	98.458	1
233.541	4.671	238.212	-1
…	…	…	…
667.300	0	667.300	1

In the initial stage, the first step is to perform data normalization.

Table 5. Results of Training Data Normalization

PRINCIPAL	PENALTY	TOTAL
0,000115897	0	0,000114969
0,000085649	0	0,000084964
0,000085649	0	0,000084964
0,000085649	0	0,000084964
0,000203194	0,000336055	0,0002056
…	…	…
0,000580637	0	0,000575991

After obtaining the normalized value data, the next step is to build the training model for the Support Vector Machine. This study utilizes a linear kernel in the development of the model and the prediction of tax reporting compliance data. The formula used is given by equation (1). In the initial calculations, the initial weights can be initialized as w_1 = 0, w_2 = 0, w_3 = 0 dan b = 0. Thus, the calculations can be implemented on the first data point as follows.

$$
\begin{aligned}
f(x_0) &= w_1 * x_1 + w_2 * x_2 + w_3 * x_3 + b \\
&= (0 * 0{,}000115897) + (0 * 0) + (0 * 0{,}000114969) + 0 \\
&= 0 + 0 + 0 + 0 \\
&= 0
\end{aligned}
$$

In the calculation, the data point x_0 results in a value of 0. In the model development, there are conditions for weight formation based on the iteration results. To determine whether a data point belongs to class 1 or -1, SVM calculates the predicted result as follows:

a. If $f(x) \geq 0$ then the predicted class is 1 (COMPLIANT).
b. If $f(x) < 0$, then the predicted class is -1 (NON-COMPLIANT).

In the first iteration, the value of y for the first data point is COMPLIANT or 1. For $f(x_0)$ which equals 0, it indicates a prediction of class 1. Since the prediction result matches the first iteration, there is no need to update the weights and bias. Conversely, if the predicted value differs from the actual class, the weights and bias must be updated. SVM uses optimization algorithms such as Gradient Descent to update the weights w and bias b based on the prediction error. The formulas used are equations (2) and (3).

Table 6. Results of SVM Calculations using Gradient Descent

	w1	w2	w3	b
F(x)	0	0	0	0
0	0	0	0	0,01
0,01	0	0	0	0,0199
0,0199	0	0	0	0,0297
0,0297	0	0	0	0,0394
0,0394	0	0	0	0,02901
…	…	…	…	…
0,66043	0,000484	-0,00064	0,000473	0,66382

After building the model and updating the weights and biases through all iterations, the testing resulted in the final weight values w and bias b as follows: w_1= 0,000484, w_2= -0,00064, w_3= 0,000473, b = 0,66382. Next, we will calculate the predicted results for the testing data, which consists of 3,868 reporting data points that will be used as the testing data as follows.

Table 7. Test Data

PRINCIPAL	PENALTY	TOTAL	COMPLIANCE'
0,000033853	0	0,000033582	1
0,000043874	0	0,000043523	1
0,000053572	0	0,000053143	1
0,000014408	0	0,000014293	1
0,00005706	0	0,000056604	1
…	…	…	…
0,000580637	0	0,000575991	1

Table 8. Test Data Results

COMPLIANCE'	F(x)	STATUS
1	0,663824	1
1	0,663824	1
1	0,663824	1
1	0,663824	1
1	0,663824	1
…	…	…
1	0,663824	1

Table 9. Confusion Matrix Results

	Positive Prediction	Negative Prediction

Actual Positive	True Positive = 3329	False Negative = 0
Actual Negative	False Positive = 539	True Negative = 0

To view the number of predicted data points that match and do not match the actual data, please refer to Figure 2.

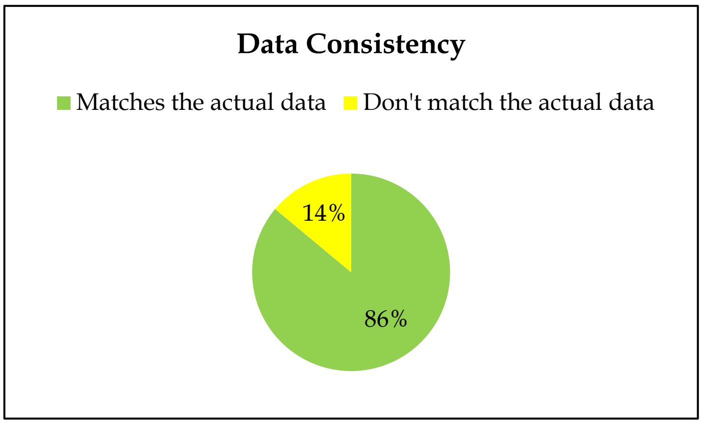

Figure 2. Data Consistency

It can be seen from Table 9 that there are 3,329 predicted data points using the SVM method that match the actual data. However, there are 539 predicted data points that do not match the actual data. Referring to Figure 2, 86% of the predicted data aligns with the actual data, while 14% does not. Therefore, the accuracy, precision, and recall results from implementing the Support Vector Machine method to assess compliance with land and building tax obligations can be observed in Figure 3 below.

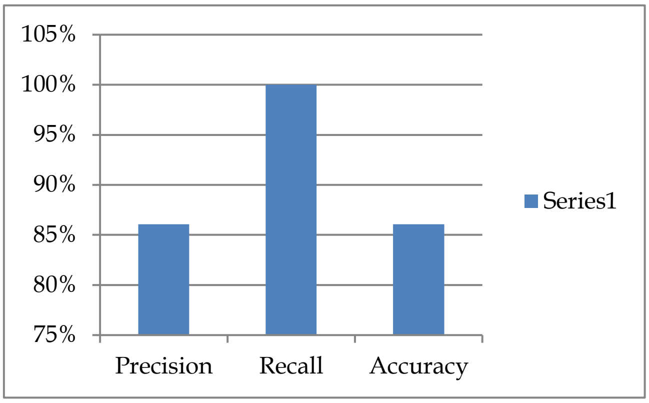

Figure 3. Evaluation SVM of Confusion Matrix

The accuracy results of the Support Vector Machine (SVM) method are as follows: an accuracy of 86%, a precision of 86%, and a recall of 100%. In Figure 4, the final results of the system implementation for assessing compliance with land and building tax obligations using the Support Vector Machine method can be seen.

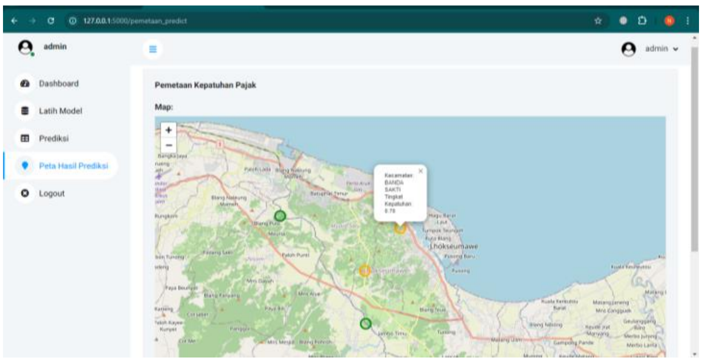

Figure 4. Mapping Results

In the mapping results, admin users can view the mapping outcomes for the reporting sub-districts that have been tested. The display includes a map and the compliance levels for land and building tax reporting. The compliance rate for Banda Sakti sub-district is 98%, Muara Satu is 99%, Muara Dua is 99%, and Blang Mangat is 100%.

Conclusions

With the classification system in place, areas with varying levels of compliance in land and building tax obligations can be identified efficiently, assisting the government in grouping regions with low tax compliance. The implementation of the Support Vector Machine (SVM) algorithm enables high-accuracy classification of land and building tax compliance. Using features such as principal amount, penalties, total amount, payment date, and due date, the SVM model can automatically classify tax data into two target classes: compliant and non-compliant.

The testing results of the SVM model indicate that Banda Sakti sub-district has a compliance rate of 98%, Muara Satu has 99%, Muara Dua has 99%, and Blang Mangat has 100%. Thus, it can be concluded that, in 2023, tax compliance in Lhokseumawe City has improved. The accuracy metrics for the Support Vector Machine method in assessing land and building tax compliance show a Precision of 86%, a Recall of 100%, and an Accuracy of 86%.

Acknowledgments

The author would like to express sincere gratitude to Mr. Rozzi Kesuma Dinata, S.T., M.Eng, the primary supervisor, for his invaluable guidance throughout the completion of this research. The author also extends thanks to Mr. Rizki Suwanda, S.T., M.Kom, the secondary supervisor, for his patient direction and support during the research process. Special appreciation goes to the Regional Financial Management Agency for their assistance and for facilitating the research. Additionally, the author is grateful to friends who have provided encouragement and support throughout the completion of this study.

References

[1] W. D. Febrian and R. Ristiliana, "Pengaruh Pengetahuan dan Kesadaran Wajib Pajak Terhadap Kepatuhan Wajib Pajak dalam Membayar Pajak Bumi dan Bangunan (PBB) pada Kantor Badan Pendapatan Daerah Kota Pekanbaru," *Eklektik J. Pendidik. Ekon. dan Kewirausahaan*, vol. 2, no. 1, hal. 181, 2019, doi: 10.24014/ekl.v2i1.7563.

[2] A. R. Shaumi, M. F. Ali, and M. T. A. M. Simbolon, "Penerapan Data Mining Menggunakan Metode Teknik Classification Untuk Melihat Penerapan Data Mining Menggunakan Metode Teknik Classification Untuk," *JUKI J. Komput. dan Inform.*, vol. 4, no. 2, hal. 171–182, 2022, [Daring]. Tersedia pada: www.pajak.go.id

[3] S. Y. Pangestu, Y. Astuti, and L. D. Farida, "ALGORITMA SUPPORT VECTOR MACHINE UNTUK KLASIFIKASI SIKAP POLITIK TERHADAP PARTAI Politik Indonesia," *J. Mantik Penusa*, vol. 3, no. 1, hal. 236–241, 2019, [Daring]. Tersedia pada: https://t.co/eF

[4] O. H. Rahman, Gunawan Abdillah, and Agus Komarudin, "Klasifikasi Ujaran Kebencian pada Media Sosial Twitter Menggunakan Support Vector Machine," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 1, hal. 17–23, 2021, doi: 10.29207/resti.v5i1.2700.

[5] H. Pratiwi, M. Muhaimin, and W. O. Rayyani, "Kontribusi Pajak Bumi Dan Bangunan (Pbb) Dalam Meningkatkan Penerimaan Pajak Daerah," *Amnesty J. Ris. Perpajak.*, vol. 3, no. 1, hal. 24–30, 2020, doi: 10.26618/jrp.v3i1.3402.

[6] M. Y. Noor, "Efektivitas Penerimaan Pajak Bumi Dan Bangunan (Pbb-P2) Di Desa Tibona Kecamatan Bulukumpa Kabupaten Bulukumba," *Amnesty J. Ris. Perpajak.*, vol. 3, no. 2, hal. 135–150, 2020, doi: 10.26618/jrp.v3i2.4409.

[7] S. Sainang and A. W. Aji, "Pengaruh Persepsi Kemanfaatan, Persepsi Kemudahan Dan Kepuasan Pengguna Terhadap Minat Membayar Pajak Bumi Dan Bangunan (Pbb) Menggunakan Shopeepay," *Amnesty J. Ris. Perpajak.*, vol. 4, no. 1, hal. 129–140, 2021, doi: 10.26618/jrp.v4i1.6317.

[8] N. P. Rahmayanti, R. M. Arini, S. D. Indiraswari, and R. R. Dara, "Pengaruh Pengetahuan Pajak, dan Sanksi Pajak Terhadap Kepatuhan Wajib Pajak," *J. Komun. Bisnis dan Manaj.*, vol. 10, no. 2, hal. 290–298, 2023, doi: 10.59422/margin.v1i02.125.

[9] R. Adawiyah, Y. Rahmawati, and I. Eprianto, "Literature Review: Pengaruh Sosialisasi Perpajakan, Sanksi Perpajakan, Pemahaman Peraturan Perpajakan Terhadap Kepatuhan Wajib Pajak," *J. Econ.*, vol. 2, no. 9, hal. 2310–2321, 2023, doi: 10.55681/economina.v2i9.812.

[10] A. Z. Praghakusma and N. Charibaldi, "Komparasi Fungsi Kernel Metode Support Vector Machine untuk Analisis Sentimen Instagram dan Twitter (Studi Kasus : Komisi Pemberantasan Korupsi)," *JSTIE (Jurnal Sarj. Tek. Inform.*, vol. 9, no. 2, hal. 33–42, 2021, doi: 10.12928/jstie.v9i2.20181.

[11] N. Aula, M. Ula, and L. Rosnita, "Analisis Sentimen Review Customer Terhadap Perusahaan Ekspedisi Jne, J&T Express Dan Pos Indonesia Menggunakan Metode Support Vector Machine (Svm)," *J. Informatics Comput. Sci.*, vol. 9, no. 1, hal. 81–86, 2023.

[12] S. Nurhaliza, Y. Yusra, and M. Fikry, "Klasifikasi Sentimen Masyarakat di Twitter Terhadap Kenaikan Harga BBM dengan Metode Support Vector Machine," *J. Sist. Komput. dan Inform.*, vol. 4, no. 4, hal. 586–593, 2023, doi: 10.30865/json.v4i4.6322.

[13] Fitrianingsih and B. Zuraeni, "Analisis Ramalan Cuaca di Sekupang, Kota Batam Menggunakan Algoritma Decision Tree dan Confusion Matrix," *J. Ekon. Pembang. dan Manaj.*, vol. 1, no. 3, hal. 15–26, 2024, [Daring]. Tersedia pada: https://ibnusinapublisher.org/index.php/EKOSPHERE

[14] R. K. Dinata, H. Akbar, and N. Hasdyna, "Algoritma K-Nearest Neighbor dengan Euclidean Distance dan Manhattan Distance untuk Klasifikasi Transportasi Bus," *Ilk. J. Ilm.*, vol. 12, no. 2, hal. 104–111, 2020, doi: 10.33096/ilkom.v12i2.539.104-111.