

Analysis of Social Assistance Donor Classification at the Muhammadiyah Medan Orphanage Using SVM

Ahmad Helmy¹⁾, Zulham Sitorus²⁾, Abdul Chaidir Hrp³⁾, Dwika Ardy⁴⁾, Siti Isna Syahri T⁵⁾, Sukrianto⁶⁾
¹⁾²⁾³⁾⁴⁾⁵⁾⁶⁾ Master of Information Technology, Pancabudi Development University, Medan, Indonesia
¹⁾ahmadhelmy.dev@gmail.com, ²⁾zulhamsitorus@dosen.pancabudi.ac.id³⁾abdulchaidir52@gmail.com,
⁴⁾dwikardy@gmail.com, ⁵⁾rhieachmad338@gmail.com, ⁶⁾sukrianto.hambaallah@gmail.com,

Submitted: Dec 6, 2024 | Accepted : Dec 15, 2024 | Published : Jan 9, 2025

Abstract: The Putra Muhammadiyah Orphanage in Medan City is a social institution that relies on donor assistance to support various social programs. The problem that occurs at the Putra Muhammadiyah Orphanage in Medan is the difficulty in identifying potential and non-potential donors who have the potential to provide sustainable social assistance contributions. This study aims to conduct a comprehensive analysis and classification of donors using the Support Vector Machine method, an effective method in machine learning to handle classification problems with SVM with high accuracy. The research data consists of donor data with several main characteristics such as the amount of donation, the frequency of donations given, and the type of assistance. The data is processed through a preprocessing stage including data normalization and data division into training and testing data. Then, the SVM model is trained to classify donors into two categories, namely Potential Donors and Non-potential Donors. Based on the data obtained from the donation bookkeeping records of the Putra Muhammadiyah Orphanage in Medan City, it can be concluded that around 55 potential donors out of 90 donors and 35 non-potential donors out of 90 donor data. From the results of the analysis and testing of the model conducted, it can be concluded that the SVM method can classify "Potential Donors" and "Non-Potential Donors" with a fairly high level of accuracy. The level of accuracy obtained reached up to 89% with a precision value of 93%, a recall value of 89% and an f1-score of 90%. With these results, this study can provide significant benefits in the management of social assistance, especially helping orphanages to determine who are potential and non-potential donors. Therefore, this study is expected to have an impact on improving the sustainability of social programs at the Putra Muhammadiyah Orphanage in Medan City.

Keywords: Donors; Muhammadiyah Putra Orphanage; Social assistance; Classification; Support Vector Machine (SVM) ;

INTRODUCTION

Orphanages are a very popular institution for shaping the development of children who do not have families. Mental emotional problems are increasing in the age group over 15 years, which has an impact on behavioral problems as adults (Haryanti et al., 2019). An orphanage is a social welfare institution that is responsible for providing social services to abandoned children. The orphanage functions as a substitute for the role of parents or families in meeting the physical, mental, and social needs of foster children, including orphans, half-orphans, half-orphans, and other abandoned children. With this service, children get a decent and adequate opportunity to develop optimally according to their potential. The ultimate goal is to create a quality next generation of the nation who are able to actively contribute to national development. In the data collection of orphans, and assistance at the Yatim Madani Orphanage is still in the form of notes in books. and regarding the delivery of information on funds, the orphanage only notifies when a donor visits the orphanage. So that information about the use of donor funds is still not conveyed properly. Meanwhile, in introducing the Yatim Madani Orphanage, the orphanage manager usually introduces it when there is a routine study, puts up a signpost on the side of the road that approaches the location of the Yatim Madani Orphanage, and in an emergency the orphanage manager makes a proposal to donors to convey what is currently needed. Among these, Muhammadiyah stands out as a significant organization that has established orphanages, such as the Muhammadiyah Orphanage for Orphaned Families (PAKYM), which

* name of corresponding author



This is a Creative Commons License. This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

cares for children from various backgrounds, including orphans, neglected children, and those from financially disadvantaged families.

Data processing can be visualized in the form of tables and diagrams/graphs. Presenting data in the form of tables and graphs can make it easier for readers to read and understand data. As data grows day by day, it will result in a wasteful accumulation of data if the data owner cannot process the data properly. The role of data is very large for the progress of a business, data can be processed in order to find out trends that will occur in the future. Currently, there are many ways to analyze data in the application of data science or

Data analysis. Data visualization is a real depiction of moving data in the form of tables, bar charts, pie charts, line graphs, maps and diagrams (Mujilahwati et al., 2021).

Fundraising or donation is a way to collect voluntary financial contributions without expecting any reward or profit. This process often involves various institutions, both religious and social, which aim to develop donation concepts that are relevant to their specific fields. It is common for large and small organizations, such as youth groups, family welfare organizations, non-governmental organizations (NGOs), student organizations, Quran study centers, community learning centers, shelters, and orphanages, to experience challenges in obtaining funding. This support vector plays a role in the formation of the hyperplane, one way that can be used to find the value of this support vector is by using the kernel function (Pambudi & Suprpto, 2021). One of the strengths of SVM lies in its high data generalization ability, producing effective classification models even when trained on relatively small datasets compared to the problem space. SVM is characterized by a clear concept and formulation, requiring minimal parameter tuning. However, SVM also faces challenges when applied to datasets with large sample sizes and dimensions. Although SVM can be extended to multiclass classification, each multiclass strategy has its own limitations.

An orphanage is a social welfare institution that is responsible for providing substitute services in fulfilling the physical, mental, and social needs of its foster children, so that they get broad, appropriate, and adequate opportunities for personality development according to expectations. The purpose of an orphanage is to provide social services to foster children by helping and guiding them towards reasonable personal development and having work skills, so that they become members of society who can live decently and responsibly, both towards themselves, their families, and the community (Rahmawati et al., 2023). The problem that occurred at the Muhammadiyah Putra Orphanage in Medan was the difficulty in identifying potential and non-potential donors, who have the potential to provide sustainable social assistance contributions. Accreditation of Social Welfare Institutions, including Child Welfare Institutions (LKSA) is an important need so that LKSA can play an appropriate role in fulfilling the needs of neglected children. Accreditation also encourages an increase in the quality of services provided by the institution, so that it will directly increase protection for children in the LKSA.

Accreditation is the determination of the level of eligibility and standardization of Social Welfare Institutions based on the assessment of programs, human resources, management and organization, facilities and infrastructure, and results of social welfare services. While the Accreditation Body for Social Welfare Institutions, hereinafter referred to as the Accreditation Body, is an institution that carries out the assessment to determine the level of eligibility and standardization of Institutions in the field of Social Welfare (Sukmana et al., 2021). The Support Vector Machine (SVM) algorithm is used for modeling. This is because SVM is an accurate algorithm because it is able to create models with various feature dimensions (linear and nonlinear) (Rianti et al., 2021). Furthermore, which classification can help find previously unknown class labels, then the labels obtained from this classification process can be used to make predictions (Trya Ayu Pratiwi, 2021)

Data processing and presentation in the current technological era can be easily processed using a programming language called Python. Python is a high-level, object-oriented programming language created by Guido van Rossum. Python is a maximally interpreted cross-functional language that has many functions. This object-oriented programming language is usually used to streamline large complex data sets. Another feature that makes Python famous among data scientists and analysts is its flexibility. Due to its flexibility, this language allows for creating data models, systematizing data sets, creating ML (Machine Learning) supported algorithms, web services, and implementing data mining to complete various tasks in a short time (Junaidi et al., 2023).

The benefits of this study provide significant value in the management of social assistance, especially helping orphanages to be able to determine who are potential and non-potential donors. Therefore, this study is expected to have an impact on improving the sustainability of social programs at the Putra Muhammadiyah Orphanage in Medan City.

LITERATURE REVIEW

Orphanage Donors

Data collection of orphans and assistance at the Yatim Madani Orphanage is still done manually using notes in notebooks. This causes challenges in efficient data management, including the risk of losing or damaging physical documents. In addition, the delivery of information about funds and their use to donors is only done directly when donors visit the orphanage, so that transparency in the use of funds is less than optimal. The lack of a digital

* name of corresponding author



recording system and structured reporting mechanism results in the potential for a decrease in the level of donor trust, which can indirectly affect the sustainability of support for the orphanage. Therefore, an information technology solution is needed that can increase recording efficiency, facilitate access to information, and increase transparency in reporting donation funds (Dadi Riskiono et al., 2020).

Datasheet

Datasheet is a collection of raw data that is the initial material for the analysis process. This data can be in the form of a dataset in a table format that contains relevant features. Data becomes an asset that can be used to find patterns that can be used in policy making. Information from the processed model can be used in projecting strategies or policies carried out for the business development process. The process of searching for large data must be carried out carefully. The larger the data, the greater the process of sorting data according to needs (Sholeh et al., 2023).

Pre-Processing

Generally, the steps of data preprocessing that are completed are text cleaning, deleting stopwords, deleting non-letter characters, and converting up-percase to lowercase. Furthermore, tokenization is carried out to separate words from text so that the data contains several words (Berutu et al., 2023). Data cleaning is one of the important steps in pre-processing to ensure that the data used is optimal, accurate, and free from noise or inconsistencies that can reduce the performance of the SVM model. In the context of orphanage donor data collection, data cleaning becomes even more crucial because donor data management usually involves various sources of information, such as Donor Name, Gender, Donation Type, Donation Amount, Frequency Per Year and Number of Children in the Orphanage. The feature selection process is then carried out after the preprocessing stage. Delete or fill in missing values so that the analysis remains valid. After the data has passed the preprocessing and feature selection stages, the data will be ready to be stored as training data.

Calculation Kernel

Kernel is used by SVM to map data to a higher dimensional space. This is useful for making data that is initially not linearly separable into separable. The most commonly used SVM kernels are Linear, Radial Basis Function (RBF), and Polynomial. In previous studies, each kernel has been used. However, there has been no comparison of the three kernel function methods on the same dataset using accuracy, sensitivity, and specificity parameters. For this reason, this study is proposed to obtain comparative information on the three kernel functions using accuracy, sensitivity, and specificity parameters. The expected results can later be used as a reference for implementing the best kernel function (Rochim et al., 2021). The dataset used is Donor Data to analyze what is potential or not potential later.

Training and Testing

Provides an extensive survey of existing methods for selecting SVM training data from large donation datasets. We divide existing techniques into several categories to help understand the fundamental ideas behind these algorithms (Nalepa & Kawulok, 2019). Data training aims to train the machine through learning stages, so that it is able to understand data of unknown type and achieve the most optimal accuracy. In this study, the implementation of training data was carried out using the SVM kernel by utilizing the scikit-learn library as a machine learning tool in the Python programming language.

SVM Classification

Data division into training sets. The sentiment analysis process begins with building training data, then this training data is processed with a certain classification algorithm to produce a classification model. This model will later be used by the classifier as a basis for carrying out the classification process. In this study, donations will be classified into two sentiments, namely potential donors and non-potential donors (Fitriyah et al., 2020). This model will later be used by the classifier as a basis in the sentiment classification process, for example to analyze donor perceptions of the transparency of orphanages or ongoing donations, then testing and the final stage of the results is evaluation.

Evaluation Results

The purpose of testing or evaluating the results is to find out how well the machine performs at the learning stage using the SVM kernel. Performance evaluation uses the confusion matrix technique and then produces an accuracy value which is the most important information from the success of a more classy algorithm in processing classified data (Furqan et al., 2020). By using the resulting kernel type, a comparison is carried out with the aim of finding out the best SVM recognition on donor data.

* name of corresponding author



METHOD

The diagram shows the flow in this study to achieve the results of data processed in training and then evaluated to determine the performance of the SVM in learning to provide optimal results, shown in Figure 1.

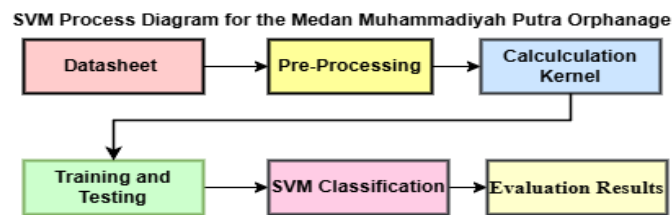


Figure 1. Support Vector Machine (SVM) Process Diagram

RESULTS

Before conducting the analysis process, several stages are required starting from collecting orphanage datasets for the last 3 years, compiling training tables, conducting pre-processing stages, to testing and visualization. In the discussion explained that there are 4 stages of the process carried out in this study, where the first stage of the process is to tidy up the data shown in Table 1.

Table 1. Orphanage Sample Dataset

ID	Donor Name	Gender	Donation Types	Donation Amount	Frequency Per (Year)	Orphanage Street
1	A. Mubarak	Man	Money	2000000	1	50
2	Laila Hidayati	Woman	Money	200000	2	50
3	Taslim Tan	Man	Basic necessities	450000	2	50
4	M. Idris	Man	Basic necessities	150000	1	50
5	The Story of Hainul Alfi Syahril	Woman	Money	200000	2	50
6	Dr. Rina Yunita	Woman	Food	1000000	1	50
7	Qadri Khalid	Man	Money	300000	1	50
8	Husna Ulfa	Woman	Food	1000000	1	50
9	Fauzan Azima	Man	Money	200000	2	50
10	Yurima	Woman	Basic necessities	300000	1	50

After compiling the Orphanage dataset, the next stage is entering the training data compilation stage which includes Donor Category, Gender, Type of Assistance, Amount of Donation, Frequency Per Year, and Number of Children in the Orphanage. These stages are shown in Table 2 below.

Table 2. Training Data

Donor Category	L	P	Money	Basic necessities	Food	Clothes	Quran	Donation Amount	Frequency Per Year	Orphanage Street
Potential	1	0	1	0	0	0	0	2000000	1	50
Potential	0	1	1	0	0	0	0	200000	2	50
Potential	1	0	0	1	0	0	0	450000	2	50
No	1	0	0	1	0	0	0	150000	1	50
Potential	0	1	1	0	0	0	0	200000	2	50
Potential	0	1	0	0	1	0	0	1000000	1	50
No	1	0	1	0	0	0	0	300000	1	50
Potential	0	1	0	0	1	0	0	1000000	1	50
Potential	1	0	1	0	0	0	0	200000	2	50
No	0	1	0	1	0	0	0	300000	1	50

* name of corresponding author



Then after compiling the training data, the next stage is that the data must be processed to convert the donor category variables into a numeric format, this stage is called the Preprocessing stage which is shown in Table 3.

Table 3. Data Preprocessing

#	Goal	Like	Profile	Money	Basic necessities	Food	Clothes	Quran	Amount Donation	Frequency Per Year	Orphanage Street
0	0	1	0	1	0	0	0	0	2000000	1	50
1	0	0	1	1	0	0	0	0	200000	2	50
2	0	1	0	0	1	0	0	0	450000	2	50
3	1	1	0	0	1	0	0	0	150000	1	50
4	0	0	1	1	0	0	0	0	200000	2	50
...
85	1	1	0	0	1	0	0	0	150000	1	50
86	1	0	1	0	1	0	0	0	150000	1	50
87	1	1	0	1	0	0	0	0	100000	1	50
88	0	1	0	0	0	0	1	0	1500000	1	50

At this stage, the data normalization process will be carried out which aims to align the feature scale, so that all features have an average of 0 and a standard deviation of 1, then dividing the data into training and testing is an important step in machine learning to evaluate model performance. The data is divided by a certain proportion, for example 80% for training and 20% for testing, using the `train_test_split` function from `scikit-learn`. The `random_state` parameter ensures that the division is consistent every time it is run. The training data is used to train the model, while the testing data is used to measure the accuracy of the model on data that has never been seen before. This step helps evaluate the model's generalization ability to new data. The function of `scikit-learn` used to build the SVM model is shown in Figure 2 below.

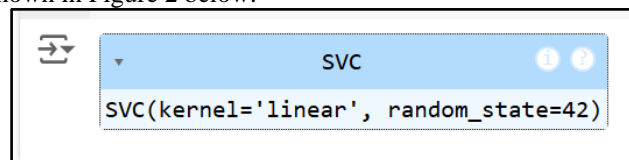
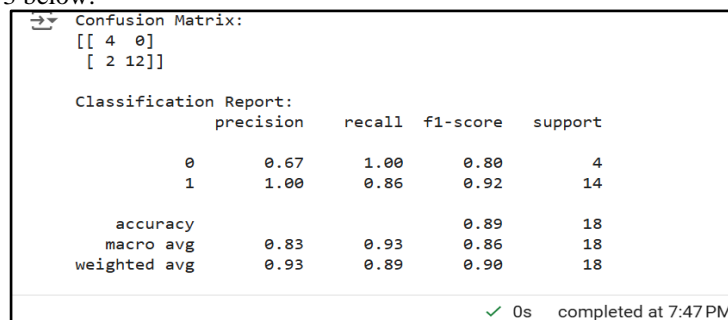


Figure 2. Initializing and Training the SVM Model

Then the next stage make predictions on test data using the Support Vector Machine (SVM) model. With the command `pred = svm_model.predict(X_test)`, the SVM model predicts labels for the features present in `X_test`, and the results are stored in `pred`. The process of predicting test data using a trained SVM model aims to evaluate the model's performance on data that has never been seen before. By utilizing the test feature data (`X_test`), the model produces predictions that are stored in the `pred` variable. These results are then compared to the actual labels (`y_test`) to calculate accuracy, which reflects the percentage of correct predictions. Further evaluation is carried out through the classification report, which includes precision, recall, and F1-score matrices, to understand how well the model handles each category. The classification report results from the SVM model can be seen in Figure 3 below:



* name of corresponding author

Figure 3. Report Classification Results

The Matrix Results in Figure 3 above show that the model has 4 correct predictions for class 0 (Non-Potential Donors) and 12 correct predictions for class 1 (Potential Donors), with few errors (2 for class 1). This indicates that the model is quite effective in identifying both classes. Report classification resultsshowed good overall performance, with an accuracy of 89%. In particular, the model was very effective in identifying class 1, with perfect precision and a high F1-score. Although recall for class 0 was perfect, the lower precision suggests that some class 0 predictions are likely inaccurate. Therefore, although the model performed well, there is still room for improvement, especially in handling the underrepresented class (class 0). Further improvement in class 0 precision could be a focus to improve the overall performance of the model. For Visualization of SVM classification is done by displaying a graph that represents the model's decision region, complete with the original data mapped based on its class label. In the graph, the hyperplane that separates the classes is visualized as a color change in the background, while the data points are displayed to show the distribution of the original samples. The axes of the graph are labeled to describe the features used, such as the amount of donation and the category of the donor, making it easier to interpret the results. This visualization helps in understanding how well the SVM model separates the classes based on the available features, as well as providing insight into areas that are difficult to classify. The visualization results of the social assistance analysis process for the classification of donors for the Muhammadiyah Medan orphanage, shown in Figure 4 below.

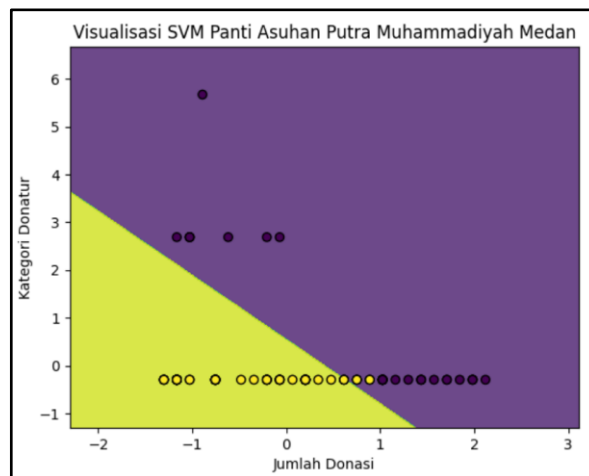


Figure 4. Results of SVM Data Visualization of Muhammadiyah Putra Orphanage

The figure shows the visualization results of the classification using the Support Vector Machine (SVM) model to separate donor categories based on the amount of donation data given. In the figure, the purple and yellow areas depict areas separated by the SVM model. The purple and yellow areas represent two different donor categories. The dividing line in the middle shows the decision boundary determined by the model to distinguish the two categories. The dots in the figure show donor data, where the color of the dot depicts the actual donor category. The results of the SVM testing process show that the higher the amount of donation, the better the classification produced. The yellow dots on the graph represent data from donors who are considered not potential, while the purple dots show potential donors, which provide good classification results for the orphanage. This graph shows the decision areas and dividing lines created by the SVM model based on two features: Donation Amount and Donor Category. This dividing line distinguishes between potential donors, indicated by the dots to the right of the dividing line, and non-potential donors, indicated by the dots to the left of the dividing line. Meanwhile, to find out the number of potential and non-potential donors, the author visualizes it in the form of a bar chart as shown in Figure 5 below:

* name of corresponding author

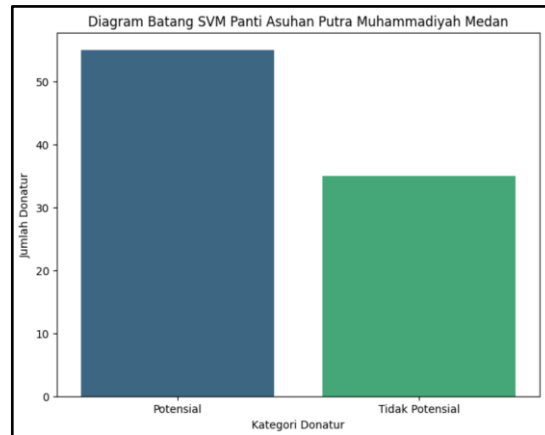


Figure 5. Bar Chart of SVM of Muhammadiyah Putra Orphanage

Figure 5 shows that there are about 55 donors out of 90 donor data who are categorized as potential donors. While there are 35 donors who are categorized as non-potential.

DISCUSSION

In this study, the linear kernel Support Vector Machine (SVM) method is used to analyze the classification of donations from the Muhammadiyah Putra Orphanage in Medan City into two categories, namely "Potential Donors" and "Non-Potential Donors". This study was conducted based on data taken from the Muhammadiyah Putra Orphanage in Medan City consisting of the amount of donations, frequency of giving, and type of donation. By using this linear kernel SVM method, the accuracy results reached up to 89%. However, there are several aspects that need to be improved, such as overcoming data imbalance and considering the efficiency of computing time on larger datasets. This study provides an important contribution in supporting more structured and data-based donation management, which can ultimately help the Muhammadiyah Putra Orphanage in Medan City to improve the effectiveness and sustainability of its social programs.

CONCLUSION

Based on the data obtained from the donation bookkeeping records of the Putra Muhammadiyah Orphanage in Medan City, it can be concluded that there are around 55 potential donors from 90 donors and 35 non-potential donors from 90 donor data. From the results of the analysis and testing of the model carried out, it can be concluded that the SVM method can classify "Potential Donors" and "Non-Potential Donors" with a fairly high level of accuracy. The level of accuracy obtained reaches up to 89% with a precision value reaching 93%, a recall value reaching 89% and an f1-score reaching up to 90%.

REFERENCES

- Berutu, S.S., Budiati, H., Jatmika, J., & Gulo, F. (2023). Data preprocessing approach for machine learning-based sentiment classification. *Infotel Journal*, 15(4), 317-325. Retrieved from <https://doi.org/10.20895/infotel.v15i4.1030>
- Dadi Riskiono, S., Hamidy, F., & Ulfia, T. (2020). Sistem informasi manajemen dana berbasis web pada panti asuhan yatim madani. *In Journal of Social Sciences and Technology for Community Service (JSSTCS)*, 1(1), 21-26. Retrieved from <https://doi.org/10.33365/jta.v1i1.670>
- Fitriyah, N., Warsito, B., Asih, D., & Maruddani, I. (2020). Analisis sentiment gojek pada media social twitter dengan klasifikasi support vector machine (SVM). *Gaussian Journal*, 9(3), 376-390. Retrieved from <https://ejournal3.undip.ac.id/index.php/gaussian/>
- Furqan, M., Kurniawan, R., & HP, KI (2020). Evaluasi performa support vector machine classifier terhadap penyakit mental. *Journal Of Business Information Systems*, 10(2), 203-210. Retrieved from <https://doi.org/10.21456/vol10iss2pp203-210>
- Haryanti, D., Pamela, E., & Susanti, Y. (2019). Mental emotional development of adolescents in orphanages. *Journal of Mental Health Nursing*, 4(2), 97-104. Retrieved from <https://doi.org/10.26714/jkj.4.2.2016.97-104>
- Junaidi, S., Dewi, M., & Kurniawan, H. (2023). Pelatihan pengelohan dan visualisasi data penduduk menggunakan python. *ADMA: Journal of Community Service and Empowerment*, 4(1), 151-162. Retrieved from <https://doi.org/10.30812/adma.v4i1.2963>
- Mujilahwati, S. (2021). Visualisasi Data Hasil Klasifikasi Naive Bayes Dengan Matplotlib Pada Python. *Prosiding Seminar Nasional Sains dan Teknologi Ke-11 Tahun 2021*, 1(1), 205-211. Retrieved from <http://dx.doi.org/10.36499/psnst.v1i1.5164>

- Nalepa, J., & Kawulok, M. (2019). Selecting training sets for support vector machines: A review. *Artificial Intelligence Review*, 52(2), 857–900. Retrieved from <https://doi.org/10.1007/s10462-017-9611-1>
- Pambudi, A., & Suprpto, S. (2021). Effect of sentence length in sentiment analysis using support vector machine and convolutional neural network method. *Indonesian Journal of Computing and Cybernetics Systems*, 15(1), 21-30. Retrieved from <https://doi.org/10.22146/ijccs.61627>
- Sukmana, O., Agustino, H., & Hidayat, W. (2021). Pendampingan pengelolaan panti asuhan putri aisiyah kota malang alam upaya persiapan akreditasi lembaga kesejahteraan sosial anak (LKSA). *Community Development Journal*, 5(1), 143-154. Retrieved from <https://doi.org/10.35326/pkm.v5i1>
- Rahmawati, FM., & Safitri, TA. (2023). Analisis SWOT Panti Asuhan Aisyiyah Putri Yogyakarta. *Community Development Journal*, 4(2), 1590-1596. Retrieved from <https://journal.universitaspahlawan.ac.id/index.php/cdj/article/view/13534/18119>
- Rianti, DL., Umaidah, Y., & Voutama, A. (2021). Tren marketplace berdasarkan klasifikasi ulasan pelanggan menggunakan perbandingan kernel support vector machine. *STRING (Technology Research and Innovation Writing Unit)*, 6(1), 98-105. Retrieved from <http://dx.doi.org/10.30998/string.v6i1.9993>
- Rochim, AF., Widyaningrum, K., & Eridani, D. (2021). Performance comparison of support vector machine kernel functions in classifying covid-19 sentiment. *2021 4th International Seminar on Research of Information Technology and Intelligent Systems (ISRITI)*, Yogyakarta, Indonesia, 2021, pp, 224–228. Retrieved from <https://doi.org/10.1109/ISRITI54043.2021.9702845>
- Sholeh, M., Nurnawati, K. E., & Lestari, U. (2023). Application of data mining with linear regression method to predict exam result value data using Rapidminer. *In Jurnal Informatika Sunan Kalijaga*, 8(1), 10-21. Retrieved from <https://doi.org/10.14421/jiska.2023.8.1.10-21>
- Pratiwi, TA., Irsyad, M., & Kurniawan, R. (2021). Classification of forest and land fires using the Naive Bayes algorithm in (Case Study: Riau Province). *Jurnal Sistem dan Teknologi Informasi*, 09(2), 101-107. Retrieved from <http://dx.doi.org/10.26418/justin.v9i2.42823>