

Analysis of the Support Vector Machine Algorithm for Predicting New Student Admissions at MA Taruna AL Jabbar

Faustina Wardhana¹⁾, Felix²⁾, Elvis Ompusunggu³⁾

^{1,2,3} Universitas Prima Indonesia

¹⁾ faussw4@gmail.com, ²⁾ felixhuang16@gmail.com, ³⁾ sastraelvis@gmail.com

ABSTRACT

Madrasah Aliyah (MA) Taruna Teknik Al Jabar is an educational institution sponsored by the Ministry of Education and Culture that holds a new student admission program periodically every year. However, the number of students accepted tends to increase or decrease every year. This situation needs to be analyzed carefully because it affects the school's policy in achieving the school's educational goals both qualitatively and quantitatively. This problem requires systematic planning and design to develop effective solutions to improve the quality of new student admissions. The algorithm applied is Support Vector Machine which is a method commonly used for classification and regression. In classification modeling, SVM has a more mature and mathematically clear concept than other classification methods. The results achieved 100% prediction accuracy from 20 data samples, and the results of the prediction data test showed that students with registration number 25023 were declared passed, while students with registration number 25024 were declared failed.

Keywords: Classification, SVM, Data Mining

INTRODUCTION

New student admissions (PPDB) is a strategic process within educational institutions aimed at selecting prospective students according to predetermined criteria and requirements. Madrasah Aliyah (MA) Taruna Teknik Al Jabar is an educational institution sponsored by the Ministry of Education and Culture and holds a new student admission program periodically every year. However, the number of accepted students tends to increase or decrease each year. This situation requires careful analysis because it affects school policies in achieving educational goals, both qualitatively and quantitatively. This problem requires systematic planning and design to develop effective solutions to improve the quality of new student admissions. Accurate analysis is expected to provide a basis for reflection in the decision-making process. One method that can be used to assist decision-making is estimating the number of prospective new students for the next period.

Data mining is a tool for extracting data from large databases with a certain level of complexity and is widely used in many application fields such as banking and telecommunications (Septian, et al., 2021). Some opinions regarding data mining are that it is a process for discovering interesting patterns or information in data using specific techniques or methods (Lisnawati, et, al., 2022). Another view states that data mining is a process that utilizes statistical methods, mathematics, artificial intelligence, and machine learning to extract and identify valuable information and relevant knowledge from various large databases (Ariansyah, et, al., 2021). Data mining is developed using predictive techniques that aim to predict the likely number of new students enrolling in the following year. Forecasting is the process of systematically estimating what is most likely to happen in the future, based on past and present information, to minimize errors. Predictions should not provide clear answers, but should try to find answers that are as close as possible to what will happen (Panggabean, et, al., 2021). According to another perspective, forecasting is the art and science of predicting future events. Forecasting or forecasting in the business world can be translated as the term projection. The advantage of prediction is that it helps in decision making (Tumanggor, et, al., 2021). Another view of forecasting is the

* Corresponding author



[Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.](https://creativecommons.org/licenses/by-nc-sa/4.0/)

estimation of the values of a variable based on the known values of that variable or related variables. According to a different view, prediction is an assumption or estimate regarding the occurrence of an event or occurrence in the future. Predictions can be qualitative (not in numbers) or quantitative (in numbers). Qualitative predictions are difficult to achieve satisfactory results because the factors involved are highly relative. Quantitative predictions are divided into two categories: point predictions and interval predictions. A single prediction covers a single value, while an interval prediction covers multiple values, forming an interval bounded by a minimum (lower-bound prediction) and a maximum (upper-bound prediction) (Astuti, et al., 2022). Various prediction methods are used to identify new students, one of which is the Support Vector Machine (SVM) algorithm.

SVM is a supervised learning method commonly used for classification and regression. In classification modeling, SVM has a more mature and mathematically clear concept than other classification methods. The goal of this algorithm is to find the largest hyperplane. A hyperplane is a function that can separate two classes. In this process, SVM maximizes the margin or distance between the training model and the decision boundary (Abdusyukur, et, al., 2023). In this study, what differentiates it from previous research is the use of different data and different research locations with the title Comparison of SVM and Decision Tree Algorithms for Tourist Attraction Recommendation Systems which produces an accuracy value of 98.97% (Oktafiani, et, al., 2023). Another study entitled Analysis of Student Satisfaction with E-Learning Using Support Vector Machine Algorithms with an accuracy level of 95.65% (Mahendro, et, al., 2022). In another study entitled Application of Support Vector Machine Methods for Tweet Sentiment Analysis about the Change of Halal Logo in Indonesia, the results obtained reached 71% (Afandi, et, al., 2023). Based on the explanation above, the author wishes to create a system for accuracy and speed in the new student registration procedure. From here the author chose the title "Analysis of Support Vector Machine Algorithms in Predicting New Student Admissions at MA Taruna Teknik Al Jabbar".

LITERATURE REVIEW

Data mining is the analysis of large amounts of observational data to discover previously unknown relationships and two new methods for summarizing data for ease of understanding and use by data collectors. Data mining is defined as the process of discovering meaningful new relationships, patterns, and trends by sifting through vast amounts of data stored in storage using pattern recognition techniques such as statistics and mathematics (Listriani, et, al., 2022).

Support Vector Machine is a technique that can separate two data sets from two different classes by maximizing the boundary of the separating function (hyperplane). One of the advantages of this method is its ability to classify and handle both linear and nonlinear regression. SVM has a better classification accuracy rate compared to other classification methods.

1. Linear Support Vector Machine

Suppose there are m training data sets $x_1, y_1, x_2, y_2, \dots, x_m, y_m$, where $x_i \in \mathbb{R}^n$ is the data sample and $y_i \in \{-1, 1\}$ is the target class of the data sample. Suppose also that the data for both classes are linearly separable, then we want to find the separation function.

2. Non-Linear Support Vector Machine

In classification problems, most data samples are not linearly separable. Therefore, using a linear SVM will produce suboptimal results and result in poor classification results. One of the advantages of SVM lies in its ability to be extended to solve non-linear problems. A linear SVM can be transformed into a non-linear SVM using the kernel method. This method works by mapping the input data to a higher-dimensional feature space using a function. For example, $u = (u_1, u_2)$ is the input data on $\mathbb{R}^2$ and $\phi(u) = (1, \sqrt{2}u_1, \sqrt{2}u_2, u_1^2, u_2^2, \sqrt{u_1}u_2)$. (Dasmase, et, al., 2022)

METHOD

This research framework is implemented through observation, followed by data collection. The data is then entered into Microsoft Excel, where it is processed through calculations and follows the steps of the SVM algorithm. The results of these calculations can then be implemented in the RStudio



Figure 1. Method

2.1. Literature Review

The literature review begins with an in-depth review of the algorithm used in this research, namely the Support Vector Machine. This review covers the basic concepts, working principles, and main characteristics of the SVM algorithm. Furthermore, the discussion covers the stages in the algorithm development process, from parameter determination and data training to model application to solve the research problem. This literature review aims to provide a strong theoretical foundation as a basis for designing and implementing the SVM algorithm in this research.

2.2. Data Collection

This research uses data from new students at the Al Jabbar Taruna Teknik MA (Islamic Senior High School) for the past year, starting in 2024.

2.3. Needs Analysis

1. At this stage, a system requirements analysis is conducted, encompassing both functional and non-functional requirements. The functional requirements analysis relates to the system's ability to provide the main features required, including:
 - a. The system is able to display the main page as a user navigation center.
 - b. The system is able to display the student data management page.
 - c. The system is able to provide a page for dataset creation.
 - d. The system is capable of displaying the training data process page.
 - e. The system is capable of displaying the new student prediction page.
2. Meanwhile, the non-functional requirements analysis focuses on the quality aspects of the system. These non-functional requirements include:
 - a. The system has an attractive, intuitive, and user-friendly interface.
 - b. The system is capable of displaying prediction results with a response time of no more than 10 seconds.

2.4. Implementation

This stage presents the overall system design results, including the integration of the Support Vector Machine algorithm into the developed system. This integration aims to ensure the system can process

* Corresponding author



available data and perform prediction functions effectively and accurately. This stage includes implementing the SVM model in the system, testing the prediction process, and adjusting the necessary parameters to ensure that the prediction results meet user needs. With the integration of the SVM algorithm, the system is expected to support the decision-making process by providing reliable prediction information.

2.5. Results Analysis and Testing

The next step after the implementation stage is results analysis and system testing. In this study, the tests conducted included functional testing and accuracy testing. Functional testing was conducted by observing the system's test results during operation to ensure all features were functioning properly. This testing aimed to assess the alignment between the designed functional requirements and the resulting system's performance, thus ensuring that the system was operating according to the specified specifications.

RESULT

3.1. Problem Analysis

Identifying the root cause of a problem is known as problem analysis. The information used in this analysis comes from 200 new students enrolled at the Al-Jabbar Technical High School (MA Taruna Teknik Al-Jabbar). The researchers used 20 new student data samples to calculate the SVM algorithm, with each sample containing four variables: Registration Number, Indonesian Language Grade (AK1), Mathematics Grade (AK2), Science and Social Studies Grades (AK3), and Memorization Grade (AK4). The SVM algorithm was then calculated after selecting the new student data. The following is a sample of data obtained from the Al-Jabbar Technical High School (MA Taruna Teknik Al-Jabbar).

Table 1. Sample New Student Data

Registration Number	AK1	AK2	AK3	AK4	Outcome
TR25001	83	90	70	73	Not Passed
TR25002	90	82	87	84	Passed
TR25003	70	80	88	85	Not Passed
TR25004	82	81	87	84	Passed
TR25005	85	75	87	70	Not Passed
TR25006	80	81	70	84	Not Passed
TR25007	82	81	88	73	Passed
TR25008	90	82	87	84	Passed
TR25009	82	82	87	84	Passed
TR25010	83	72	86	73	Passed
TR25011	70	83	87	75	Not Passed
TR25012	82	81	87	84	Passed
TR25013	82	82	73	70	Not Passed
TR25014	75	82	87	84	Not Passed
TR25015	83	82	70	71	Not Passed
TR25016	90	83	87	84	Passed
TR25017	85	80	90	83	Passed
TR25018	82	82	86	83	Passed
TR25019	72	85	90	72	Not Passed
TR25020	84	90	87	84	Passed

This stage needs to be improved to ensure precise and accurate data analysis results. Data preprocessing is the initial method for transforming raw data into more effective and useful formats and information. Effective data preprocessing can impact data outcomes, resulting in increased accuracy and clearer new data processing. Data cleaning, data selection, and data balancing are all part of the data preprocessing stage. The results of the data preprocessing are shown below.

* Corresponding author



Table 2. New Student Data Preprocessing

Registration Number	AK1	AK2	AK3	AK4	Outcome
25001	83	90	70	73	1
25002	90	82	87	84	0
25003	70	80	88	85	1
25004	82	81	87	84	0
25005	85	75	87	70	1
25006	80	81	70	84	1
25007	82	81	88	73	0
25008	90	82	87	84	0
25009	82	82	87	84	0
25010	83	72	86	73	0
25011	70	83	87	75	1
25012	82	81	87	84	0
25013	82	82	73	70	1
25014	75	82	87	84	1
25015	83	82	70	71	1
25016	90	83	87	84	0
25017	85	80	90	83	0
25018	82	82	86	83	0
25019	72	85	90	72	1
25020	84	90	87	84	0

Information is a strategy that can be applied in survey implementation by adjusting the level of information readiness and verification. In this section, data preparation includes 80% of the training data, while 20% is allocated for testing data. This study will conduct several experiments to assess the percentage and quantity of data sharing to achieve the highest accuracy. Further information relates to the data expected for the testing data. The following is the data that will be used for prediction:

Table 3. Prediction Data

Num	Registration Number	AK1	AK2	AK	AK4	Result
1	25021	90	95	80	78	?
2	25022	80	73	78	82	?

3.2. Implementation of the SVM Algorithm

Implementation of the SVM algorithm is the process of building the best classification model by separating data based on specific features. From the 20 data samples used in this study, a system test will be conducted using the Python programming language to determine the best classification results. The following are the system test results:

	Registration Number	AK1	AK2	AK3	AK4	Outcome
0	25001	83	90	70	73	1
1	25002	90	82	87	84	0
2	25003	70	80	88	85	1
3	25004	82	81	87	84	0
4	25005	85	75	87	70	1
5	25006	80	81	70	84	1
6	25007	82	81	88	73	0
7	25008	90	82	87	84	0
8	25009	82	82	87	84	0
9	25010	83	72	86	73	0
10	25011	70	83	87	75	1
11	25012	82	81	87	84	0
12	25013	82	82	73	70	1
13	25014	75	82	87	84	1
14	25015	83	82	70	71	1
15	25016	90	83	87	84	0
16	25017	85	80	90	83	0

Figure 2. Sample Data

```
df.Hasil.value_counts()
```

```
Hasil
0    11
1     9
Name: count, dtype: int64
```

Figure 3. Data Results

The figure above shows that the resulting data column has a fairly balanced class distribution, with a smaller majority in category 0 (11 data points) compared to category 1 (9 data points). This dataset is suitable for classification without significant class imbalance issues.

```
df.isnull().sum()
No Pendaftaran    0
AK1               0
AK2               0
AK3               0
AK4               0
Hasil            0
dtype: int64

df.duplicated().sum()
np.int64(0)
```

Figure 4. Data Cleaning

The data cleaning results show excellent quality, as all columns are complete, with no blank values and no duplicate data. Therefore, the data is ready for statistical analysis and machine learning algorithm implementation without additional cleaning processes.

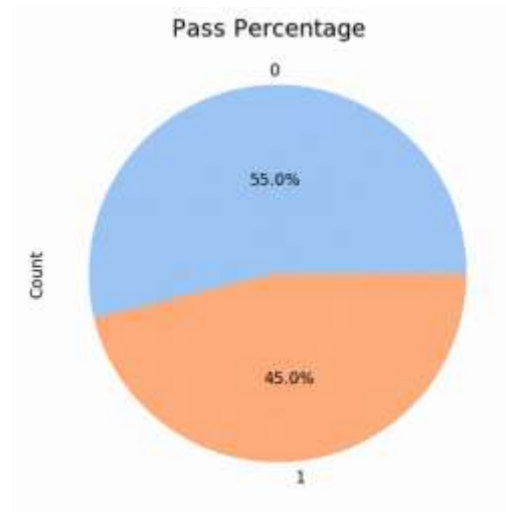


Figure 5. Data Visualization

The diagram above shows that the majority of students (55%) are in the pass category, while 45% of students fail out of 20 data samples. This relatively balanced distribution makes the dataset suitable for classification analysis and evaluation of pass patterns.

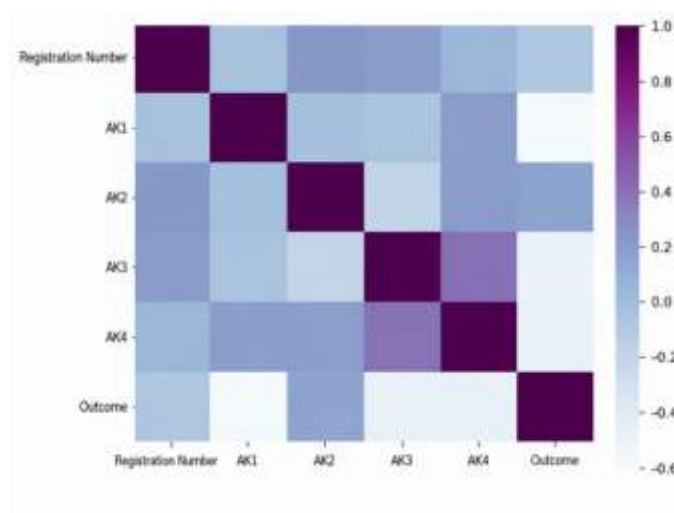


Figure 6. Data Correlation

The correlation heatmap above shows that the relationship between student scores tends to be positive, especially between AK3 and AK4, while the Outcome variable is not strongly dependent on a single attribute. This suggests that the graduation decision is likely influenced by a combination of several assessment aspects, so a multivariate classification analysis such as SVM is more appropriate.

```
x = df.drop(columns=['Has1'])
y = df['Has1']

print("x : ", x.shape)
print("y : ", y.shape)

x = (20, 5)
y = (20,)

scaler = StandardScaler()
scaler.fit(x)
x = scaler.transform(x)

x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=0.2, random_state=42)
```

Figure 7. Data Preprocessing

This figure shows the data preparation stages before training a classification model for the SVM algorithm. The dataset used consists of 20 data points. The data has been properly processed, split, normalized, and divided into train and test.

	precision	recall	f1-score	support
0	1.00	1.00	1.00	3
1	1.00	1.00	1.00	1
accuracy			1.00	4
macro avg	1.00	1.00	1.00	4
weighted avg	1.00	1.00	1.00	4

Akurasi SVM : 100.00%

Figure 8. Classification Model Evaluation

The figure above shows that the model achieved 100% accuracy on the test data. The precision, recall, and F1-score values were all 1.00 for both classes (0 and 1), indicating that the model made no errors on the data and all predictions were correct.

```
new_data = {
    'Registration Number': [25023, 25024],
    'AK1': [90, 78],
    'AK2': [95, 80],
    'AK3': [80, 75],
    'AK4': [78, 85]
}

new_data = pd.DataFrame(new_data)
new_data

Registration Number  AK1  AK2  AK3  AK4
0                25023   90   95   80   78
1                25024   78   80   75   85

scaled_new_data = scaler.transform(new_data)
y_pred_new = clf.predict(scaled_new_data)
print("Prediction Results:", y_pred_new)

Prediction Results: [0 1]
```

The test results in the image above show that the predictions obtained from the image above resulted in the student with registration number 25023 receiving a score of "0," meaning the student passed. The student with registration number 25024 received a score of "1," meaning the student failed.

DISCUSSIONS

This study applies the Support Vector Machine (SVM) algorithm to classify the graduation of new students at the Al-Jabbar Technical High School (MA Taruna Teknik Al-Jabbar) based on four assessment variables: Indonesian Language Score (AK1), Mathematics Score (AK2), Science and Social Studies Score (AK3), and Memorization Score (AK4). The dataset used consists of 20 student data samples with a relatively balanced class distribution: 11 data points in category 0 and 9 data points in category 1. This situation benefits the classification process because the model does not encounter class imbalance, which often leads to prediction bias. Preprocessing results indicate that the dataset is of

excellent quality, as there are no missing values or duplicate data. Furthermore, the normalization process and data division into 80% training data and 20% test data have been carried out correctly. This stage is crucial in the application of the SVM algorithm because it is sensitive to differences in data scale. With good data quality, the model can optimally learn classification patterns.

Based on the correlation analysis, the relationship between the assessment variables tends to be positive, especially between AK3 and AK4. However, the outcome variables did not show a strong dependence on a single attribute. This finding indicates that student graduation decisions are not influenced solely by a single academic score, but rather by a combination of several assessment aspects. Therefore, the use of a multivariate classification method such as SVM is deemed appropriate because it can simultaneously consider all attributes in the decision-making process. Model evaluation results showed that the SVM algorithm achieved 100% accuracy, with precision, recall, and F1-score values each reaching 1.00 for both classes. These results indicate that all test data was correctly classified without prediction errors. This high level of accuracy indicates that the data patterns in the dataset can be very well separated by the hyperplane formed by the SVM algorithm. These results align with several previous studies showing that the SVM algorithm performs very well in educational data classification tasks. Overall, this study demonstrates that the SVM algorithm is capable of providing excellent classification performance on new student selection data and has great potential for application in decision support systems in educational settings. However, testing on larger and more diverse datasets is still needed to ensure the model's reliability in real-world conditions.

CONCLUSION

The conclusion that can be drawn from this study is that this study only uses the SVM algorithm to predict new students at MA Taruna Teknik Al Jabbar, using 4 exam score variables. The use of the SVM algorithm in the python programming language achieved 100% prediction accuracy from 20 data samples, and the results of the prediction data test showed that students with registration number 25023 were declared to have passed, while students with registration number 25024 were declared to have failed.

REFERENCES

- Isnanto, Septian et al. 2021. "Penerapan Data Mining Pada Penerimaan Mahasiswa Baru Dengan Algoritma K-Means Clustering." 4: 158–67.
- Listriani, Dewi, Anif Hanifa Setyaningrum, and Fenty Eka M A. 2022. "Penerapan Metode Asosiasi Menggunakan Algoritma Apriori Pada Aplikasi Analisa Pola Belanja Konsumen (Studi Kasus Toko Buku Gramedia Bintaro)." 9(2): 120–27.
- Hendra Di Kesuma, Deni Apriadi, Hengki Juliansa, and Endang Etriyanti. 2022. "Implementasi Data Mining Prediksi Mahasiswa Baru Menggunakan Algoritma Regresi Linear Berganda." *Jurnal Ilmiah Binary STMIK Bina Nusantara Jaya Lubuklinggau* 4(2): 62–66.
- Lisnawati et al. 2022. "Algoritma Linear Regression Dalam Memprediksi Pertumbuhan Jumlah Penduduk." *Jurnal Ilmiah Wahana Pendidikan* <https://jurnal.unibrah.ac.id/index.php/JIWP> 8(3): 178–83. <http://jurnalmahasiswa.unesa.ac.id/index.php/jurnal-penelitian-pgsd/article/view/23921>.
- Ariansyah, T, P Purwadi, and S Yakub. 2022. "Implementasi Data Mining Untuk Mengestimasi Kebutuhan Persediaan Roti Panggang Di Junction Cafe Dengan Menggunakan Metode Regresi Linier Berganda." *Jurnal Cyber Tech* 1(1): 1–8.
- Hendrian, Senna. 2018. "Algoritma Klasifikasi Data Mining Untuk Memprediksi Siswa Dalam Memperoleh Bantuan Dana Pendidikan." *Faktor Exacta* 11(3): 266–74.
- Simbolon, Cici Erlangga. 2021. "Penerapan Algoritma Regresi Linier Sederhana Dalam Memprediksi Keuntungan Dan Kerugian Kelapa Sawit Pt . Sri Ulina." *Journal of Information System Research (JOSH)* 2(2): 169–72.

- Putri, Ayu Azlina. 2021. "Penerapan Data Mining Untuk Memprediksi Penjualan Buah Dan Sayur Menggunakan Metode K-Nearest Neighbor (Studi Kasus : PT . Central Brastagi Utama)." 1(6): 354–61.
- Panggabean, Devi Sari Oktavia, Efori Buulolo, and Natalia Silalahi. 2020. "Penerapan Data Mining Untuk Memprediksi Pemesanan Bibit Pohon Dengan Regresi Linear Berganda." *JURIKOM (Jurnal Riset Komputer)* 7(1): 56.
- Julianto, Indri Tri, Dede Kurniadi, Muhammad Rikza Nashrulloh, and Asri Mulyani. 2022. "Comparison Of Data Mining Algorithm For Forecasting Bitcoin Crypto Currency Trends Komparasi Algoritma Data Mining Untuk Forecasting Tren." 3(2): 245–48.
- Setyoningrum, Nanny Raras et al. 2022. "Implementasi Algoritma Regresi Linear Dalam Sistem Prediksi Pendaftar Mahasiswa Baru Sekolah Tinggi Teknologi Indonesia Tanjungpinang." *Prosiding Seminar Nasional Ilmu Sosial dan Teknologi (SNISTEK)* (4): 13–18. <https://ejournal.upbatam.ac.id/index.php/prosiding/article/view/5200>.
- Tumanggor, E M. 2021. "Analisa Dan Implementasi Data Mining Untuk Memprediksi Jumlah Material Bangunan Menggunakan Algoritma Autoregressive Intergrated Moving Average (ARIMA)." *TIN: Terapan Informatika Nusantara* 2(6): 373–77. <http://ejurnal.seminar-id.com/index.php/tin/article/view/887%0Ahttp://ejurnal.seminar-id.com/index.php/tin/article/download/887/643>.
- Astuti, Diyah, Dyah Yunita Hartanti, Susi Tri Nurhayanti, and Herlin Fransiska. 2022. "Clustering and Forecasting of Covid-19 Data in Indonesia." *Jurnal Matematika, Statistika dan Komputasi* 18(3): 324–35.
- Abdusyukur, Fatwa. 2023. "Penerapan Algoritma Support Vector Machine (Svm) Untuk Klasifikasi Pencemaran Nama Baik Di Media Sosial Twitter." *Komputa : Jurnal Ilmiah Komputer dan Informatika* 12(1): 73–82.
- Dasmasela, R, B P Tomasouw, and Z A Leleury. 2022. "Penerapan Metode Support Vector Machine (SVM) Untuk Mendeteksi Penyalahgunaan Narkoba." *Jurnal Parameter* 01(02): 111–22.
- Oktafiani, Rian, and Rianto Rianto. 2023. "Perbandingan Algoritma Support Vector Machine (SVM) Dan Decision Tree Untuk Sistem Rekomendasi Tempat Wisata." *Jurnal Nasional Teknologi dan Sistem Informasi* 9(2): 113–21.
- Mahendro, Iwan, and Dhanan Abimanto. 2022. "Analisa Kepuasan Mahasiswa Terhadap E-Learning Menggunakan Algoritma Support Vector Machine." *Jurnal Sains Dan Teknologi Maritim* 23(1): 97.
- Afandi, Widi, Tri Ginanjar Laksana, and Nia Annisa Ferani Tanjung. 2023. "Penerapan Metode Support Vector Machine Analisis Sentimen Tweet Pergantian Logo Halal Di Indonesia." *Jurnal Elektronika dan Komputer* 16(1): 45–52. <https://journal.stekom.ac.id/index.php/elkom/article/view/964%0Ahttps://journal.stekom.ac.id/index.php/elkom/article/download/964/837>.