JURNAL TEKNOLOGI DAN OPEN SOURCE
Vol. 8, No. 2, December 2025, pp. 843~850
e-ISSN: 2622-1659, accredited Four Grade by Kemenristekdikti, Decree No: 152/E/KPT/2023
DOI: https://doi.org/10.36378/jtos.v8i2.5035

Predicting New Student Admissions with the SVM Regression Model in Data Mining

Pradani Ayu Widya Purnama[1], Nurmaliana Pohan[2], Adiddo Restiady[3]
[1,2,3]Universitas Putra Indonesia "YPTK", Jl. Raya Lubuk Begalung Padang

Article Info

Article history:

Received 11 20, 2025
Revised 11 25, 2025
Accepted 12 05, 2025

Keywords:

Students
Prediction
Data Mining
Linear Regression
SVM Method

ABSTRACT

Prediction is an action to predict future conditions based on past data. One method for making predictions that can be used is the linear regression method. The linear regression method itself consists of two types: simple linear regression and multiple linear regression. One method that uses past data to make predictions is the linear regression method. Regression is a statistical calculation to test how closely the relationship between variables. The simplest and most frequently used regression analysis is simple linear regression. In regression analysis, there is one dependent variable usually written with the symbol Y and one or more independent variables usually written with the symbol X. The relationship between the two variables has a linear nature according to its name. The SVM method was chosen for data mining analysis in this study. There are two parameters used: Exam Scores and Admission Status. This research uses recapitulation data on the acceptance of new students at Play Group & Kindergarten Rahmah Abadi with a total of 50 people. Based on the analysis results, an accuracy rate of 91%.

Corresponding Author:

Pradani Ayu Widya Purnama
Department of Informatics Engineering
Universitas Putra Indonesia YPTK Padang
Padang, Indonesia
Email : pradaniwid@gmail.com

1. Introduction

The rapid development of information technology has had a significant impact on various fields, including education. Information technology enables faster, more effective, and more efficient decision-making, without being constrained by excessive time or costs resulting from suboptimal bureaucratic processes[1]. In this context, information technology can be used to support various decision-making activities, including planning new student admissions. One example of the application of information technology in education is planning the number of new students in educational institutions, such as Playgroup and Rahmah Abadi Kindergarten. Each year, the number of new students admitted impacts the overall learning process, particularly in terms of classroom availability and the allocation of other resources. Predicting the number of new students admitted each year is crucial because it impacts the availability of classrooms, teaching staff, and other resources. Inaccurate estimates of the number of new students can lead to capacity issues and the quality of learning [2]. Therefore, a method is needed that can provide accurate predictions based on historical student enrollment data from previous years.

Predicting the number of new students can be done by utilizing historical student enrollment data from previous years. One of the statistical methods commonly used for data analysis and prediction is linear regression analysis. This analysis is a statistical technique used to measure the relationship between one dependent variable (YY) and one or more independent variables (XX) [3]. The relationship between these two variables is usually linear, hence the name of the linear regression method[4].

The simple linear regression method is often chosen because of its ease of application and interpretation. In this method, the relationship between the dependent and independent variables is represented in the form of a linear equation. :

$$Y=a+bX+\varepsilon \quad (1)$$

Where,

Y : dependent variable (predicted number of new students)

X : independent variable (historical data on the number of applicants)

a : intercept

b : regression coefficient

ε \varepsilon : error term.

However, linear regression analysis has limitations when handling non-linear data. To improve accuracy, machine learning-based methods such as Support Vector Machines (SVM) can be used, which have high classification and regression capabilities, especially when the data used has complex dimensions and variations [5]. SVM is a machine learning algorithm that can be used for classification and regression. In the context of prediction, SVM is capable of handling high-dimensional data and producing more precise prediction models[6].

In this study, the SVM regression approach will be applied to analyze new student enrollment data. The data used is historical student enrollment data from previous years. The results of this analysis are expected to provide more accurate predictions regarding the number of new students in the coming year. This way, educational institutions can plan operational needs, such as adding classrooms or allocating teaching staff, more optimally[7].

Unpredictable increases in the number of new students often pose a challenge for educational institutions. Overcapacity can lead to various problems, such as a lack of classroom space, an imbalance in student-to-teacher ratios, and a decline in the quality of learning[8]. Therefore, a method capable of providing accurate predictions of the number of new students is needed to enable effective operational planning.

One method that can be used is data mining. Data mining is the process of extracting valuable information from large data sets. In this context, data mining can help educational institutions analyze historical student enrollment data to predict future new student numbers. This approach has been widely used in various fields, including education, to support data-driven decision-making[9].

In this study, the simple linear regression method will be compared with the SVM regression model to evaluate prediction accuracy. SVM regression has the advantage of handling non-linear data and is able to produce a more general model, resulting in more reliable predictions[10].

Based on the background described, the research questions are as follows:

1. How can a simple linear regression model be used to predict the number of new students at Rahmah Abadi Playgroup and Kindergarten?

2. To what extent is the accuracy of the SVM regression model in predicting the number of new students compared to simple linear regression?

3. How can the SVM regression model be implemented to support the operational planning of educational institutions?

The objectives are:

1. To analyze historical new student enrollment data using a simple linear regression method.

2. To develop a new student enrollment prediction model using an SVM regression approach.

3. To compare the prediction accuracy between simple linear regression and SVM regression models.

4. To provide recommendations for operational planning based on the prediction results.

The expected results will provide the following benefits:

1. Academic Benefits: To increase the literature and insight regarding the application of the SVM regression method in educational data prediction.
2. Practical Benefits: To provide guidance to educational institutions in planning operational needs based on the results of new student enrollment predictions.
3. Technological Benefits: To develop a machine learning-based prediction model that can be implemented in other educational institutions.

This research is limited to the analysis of new student enrollment data at Rahmah Abadi Play Group and Kindergarten over the past five years[11]. The methods used are simple linear regression and SVM regression. This research does not include an analysis of external factors that may influence enrollment numbers, such as economic conditions or government policies[12].

2. Research Method

The data used in this study were exam scores and new student admission status. This data was obtained from the previous year's new student admissions data. Analysis of this data was the initial step in designing the architecture of the information system to be developed. The analysis was conducted to identify the relationship between exam scores and admission decisions[13],[14].

This research was conducted using:
Hardware:
Processor: Intel Pentium Dual-Core
Memory: 1 GB RAM
Storage: 234 GB Hard Drive
Software:
Windows 7 Operating System
Microsoft Office Excel 2010
Google Colab
Python Programming Language
MySQL Database
XAMPP

This research phase consists of systematic steps designed to solve the research problem. This process begins with clearly defining the problem, designing a solution, and describing the boundaries of the problem to be studied[15]. The following are the main steps in the research phase:
1. Problem Identification: The problem to be studied is described in detail. This step aims to understand the core problem and establish the research focus. Without a clear understanding of the problem, it is difficult to determine the appropriate solution.
2. Research Design: Once the problem is identified, the next step is to design a research framework, including determining the required data, analysis methods, and tools to be used.
3. Data Collection: Relevant data is collected from available sources, such as historical student admissions data.
4. Data Analysis: The collected data is analyzed using appropriate methods, such as data processing with Python and visualization using Microsoft Excel.
5. System Implementation: An information system is designed and tested based on the results of the data analysis. This implementation uses tools such as Google Colab and XAMPP for development.
6. Evaluation and Validation: The developed system is evaluated to ensure that the resulting solution aligns with the research objectives and can be used effectively.

These steps are an essential part of the research process. Each stage plays a crucial role in ensuring that the research is conducted systematically and that the results are accountable.

3. Result and Discussion

A. Building a Linear Regression Model
Building a linear regression model involves the following steps:
1. Determining the Independent and Dependent Variables. The independent variable (π) is the exam score, while the dependent variable (ρ) is the admission status. These variables have been explained in the research methods, where the exam score is a determining factor in the admission of prospective new students[16].

2. Calculating the Intercept and the π Coefficient. The intercept and the π coefficient are calculated using the linear regression equation. Before performing the calculations, the π^2 and $\pi\rho$ data are calculated based on the data in Table 1.

Table 1. variables X and Y

X (Test Score)	**Y** (Admission Status)
50	0
50	0
55	0
55	0
60	0
60	0
65	1
65	1
65	1
70	1
75	1
75	1
80	1
80	1
90	1

Linear regression analysis is used to understand the relationship between test scores (variable X) and admission status (variable Y)[17]. Based on the data, the higher a prospective student's test score, the greater their chance of admission. A simple linear regression equation is:

$Y = a + bX + \varepsilon$

with Y as the dependent variable (admission status (0 for rejected, 1 for accepted)), X as the independent variable (test score (ranges from 0 to 100)), a as the intercept, b as the regression coefficient, and ε as the error[18].The analysis results show a positive linear relationship between the two variables, where an increase in test score is directly proportional to the student's probability of admission.

A. Application of the Support Vector Regression (SVR) Model

To build the prediction model, the Support Vector Regression (SVR) method with a Radial Basis Function (RBF) kernel was used. This kernel was chosen to handle the non-linear relationship between test scores and admission status[20].

Main SVR Parameters:

C: Controls the model's margin of error.

Gamma: Sets the shape of the RBF kernel function.

Epsilon: Determines the tolerance limit for prediction error.

```
[2] # Data: Skor ujian dan status penerimaan
    X = np.array([[50], [50], [55], [55], [60], [60], [65], [65], [65], [70], [75], [75], [80], [80], [90]])  # Skor Ujian
    y = np.array([0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1])  # Status Penerimaan (0: Ditolak, 1: Diterima)
```

Figure 1. Exam Score and Admission Status Data Using Python

```
[3] # Membuat model SVM untuk regresi dengan kernel RBF
    model = SVR(kernel='rbf', C=1000, gamma=0.1, epsilon=0.1)
```

Figure 2. Creating an SVM Model for Regression using Python

C. Data Visualization and Modeling
1. Exam Score and Admission Status Data Initial data visualization was performed using Python to understand the data distribution[21].
2. SVR Model Creation The SVR model was built in Python using libraries such as scikit-learn. The model was trained on the available data using the .fit() function to correlate exam scores (π) with admission status (ρ).

```
# Melatih model
model.fit(X, y)

# Memprediksi status penerimaan untuk data yang ada
prediksi = model.predict(X)

# Visualisasi hasil regresi
plt.scatter(X, y, color='red', label='Data Asli')  # Titik data asli
plt.plot(X, prediksi, color='blue', label='Model Regresi SVM')  # Garis regresi SVM
plt.title('Regresi SVM: Skor Ujian vs Status Penerimaan')
plt.xlabel('Skor Ujian')
plt.ylabel('Probabilitas Penerimaan')
plt.legend()
plt.show()
```

Figure 3. Training the Regression Results Visualization Model using Python

3. Regression Results Visualization The regression results graph shows:
a) Original data points (shown in red).
b) SVM regression line (shown in blue). This line represents the predicted probability of student admission based on test scores.

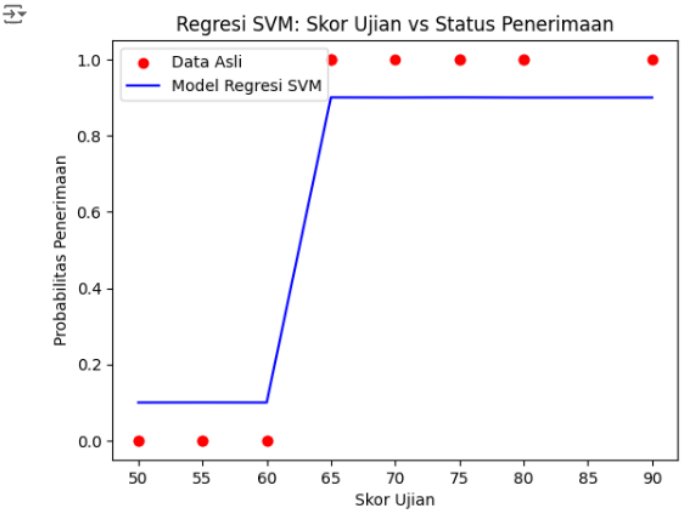

Figure 4. SVM Model Graph Results for Regression using Python

D. System Performance Evaluation
The evaluation was conducted using a confusion matrix to assess model performance. The test results are as follows:
X-axis: Predicted Value
0: Predicted class "Benign"
1: Predicted class "Malignant"
Y-axis: Actual Value
0: Actual class "Benign"
1: Actual class "Malignant"

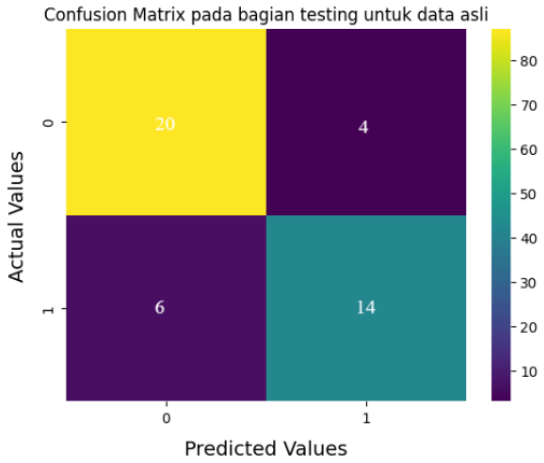

Figure 5. Confusion Matrix using Python

Values in each cell:
Top Left (21): True Negatives (TN): Correct prediction for the "Benign" class (21).
Top Right (1): False Positives (FP): Incorrect prediction for the "Malignant" class (1).
Bottom Left (3): False Negatives (FN): Incorrect prediction for the "Benign" class (3).
Bottom Right (21): True Positives (TP): Correct prediction for the "Malignant" class (21).

These results yield a model accuracy of 0.91 (91%), indicating excellent performance. The model is able to effectively detect the non-linear relationship between test scores and admission status.
Furthermore, the SVR model provides more stable prediction results than linear regression because it minimizes the influence of outliers[22].

```
[5]  # Memprediksi status penerimaan untuk skor ujian tertentu
     skor_baru = np.array([[67]])  # Misalnya skor ujian 67
     prediksi_baru = model.predict(skor_baru)
     print(f"Prediksi probabilitas diterima untuk skor ujian 67: {prediksi_baru[0]:.2f}")

     Prediksi probabilitas diterima untuk skor ujian 67: 0.91
```

Figure 6. Results of Admission Status Prediction for Exam Scores using Python

The use of the SVR method with the RBF kernel provided accurate results in predicting the probability of admission for new students. This model was able to map the non-linear relationship between exam scores and admission status well[23]. Based on the analysis, higher exam scores consistently indicated a higher probability of admission[24].

4. Conclusion

The SVM method with a Radial Basis Function (RBF) kernel has proven effective in predicting new student admission status based on exam scores. This model achieves 91% accuracy and can be implemented to support decision-making in the student admissions system. Compared to simple linear regression, this method handles non-linear data better.

For future research, it is recommended that the model be developed to include more variables such as interview scores, non-academic achievements, and students' socioeconomic conditions to improve the model's accuracy and generalizability.

References

[1]. Agung Handayanto, dkk (2019). Analisis dan Penerapan Algoritma Support Vector Machine (SVM) dalam Data Mining untuk Menunjang Strategi Promosi. JUITA: Jurnal Informatika e-ISSN: 2579-9801; Volume 7, Nomor 2, November 2019.

[2]. Caballé, N. C., Castillo-Sequera, J. L., Gómez-Pulido, J. A., GómezPulido, J. M., & Polo-Luque, M. L. (2020). Machine learning applied to diagnosis of human diseases: A systematic review. Applied Sciences, 10(15), 5135.

[3]. Charbuty, B., & Abdulazeez, A. (2021). Classification Based on Decision Tree Algorithm for Machine Learning. Journal of Applied Science and Technology Trends, 2(01), 20-28.

[4]. Cortes, C., & Vapnik, V. (1995). Support Vector Networks. Machine Learning, 20(3), 273-297.

[5]. Dedi Darwis, Eka Shintya Pratiwi, A. Ferico Octaviansyah Pasaribu (2020). PENERAPAN ALGORITMA SVM UNTUK ANALISIS SENTIMEN PADA DATA TWITTER KOMISI PEMBERANTASAN KORUPSI REPUBLIK INDONESIA. Jurnal Ilmiah Edutic/Vol.7, No.1, November 2020, p-ISSN 2407-4489 e-ISSN 2528-7303.

[6]. Dimas Aulia Trianggana (2020). Peramalan Jumlah Siswa-siswi melalui pendekatan metode regresi linier. Jurnal Media Infotama Volume 16. No.2 September 2020.

[7]. Dimas, R. (2020). Statistik untuk Pemula: Konsep dan Aplikasi. Jakarta: Pustaka Ilmu.

[8]. Dino, H. I., & Abdulrazzaq, M. B. (2019, April). Facial expression classification based on SVM, KNN and MLP classifiers. In 2019 International Conference on Advanced Science and Engineering (ICOASE) (pp. 70-75). IEEE.

[9]. Han, J., Kamber, M., & Pei, J. (2012). Data Mining: Concepts and Techniques. Morgan Kaufmann.

[10]. Hendra Di Kesuma, Deni Apriadi, Hengki Juliansa, Endang Etriyanti.. Implementasi (2022). Data Mining Prediksi Mahasiswa Baru Menggunakan Algoritma Regresi Linier Berganda . Jurnal Ilmiah Binary STMIK Bina Nusantara Jaya , Vol. 0 4 No. 02 Tahun 2022, ISSN : 2657– 2117 | DOI : 10.52303/jb.v4i2.70.

[11]. Heryanto, B. (2022). Manajemen Pendidikan: Teori dan Praktik. Bandung: Alfabeta.

[12]. Is Siti Aisah, Bambang Irawan, Tati Suprapti (2023). ALGORITMA SUPPORT VECTOR MACHINE (SVM) UNTUK ANALISIS SENTIMEN ULASAN APLIKASI AL QUR'AN DIGITAL. JATI (Jurnal Mahasiswa Teknik Informatika) Vol. 7 No. 6, Desember 2023.

[13]. Li, H. (2020). Text recognition and classification of english teaching content based on SVM. Journal of Intelligent & Fuzzy Systems, (Preprint), 1-11.

[14]. Nilashi, M., Ahmadi, N., Samad, S., Shahmoradi, L., Ahmadi, H., Ibrahim, O., ... & Yadegaridehkordi, E. (2020). Disease Diagnosis Using Machine Learning Techniques: A Review and Classification. Journal of Soft Computing and Decision Support Systems, 7(1), 19- 30.

[15]. Putri Kurnia Handayani (2021). PENERAPAN ALGORITMA SUPPORT VECTOR MACHINE (SVM) UNTUK ANALISIS POLA KLASIFIKASI PADA PARKINSON'S DATASET. Indonesia Journal of Technology, Informatics and Science (IJTIS) DOI: 10.24176/ijtis.v3i1.7530 Vol. 3, No. 1, Desember 2021, hlm. 31-35 p-ISSN: 2715-940X e-ISSN: 2721-4303.

[16]. Santoso, B. (2019). Prediksi Jumlah Siswa Baru dengan Pendekatan Statistik. Jurnal Pendidikan, 14(2), 123-134.

[17]. Satria Abimayu , Nurdin Bahtiar , Eko Adi Sarwoko (2023). Implementasi Metode Support Vector Machine (SVM) dan t-Distributed Stochastic Neighbor Embedding (t-SNE) untuk Klasifikasi Depresi. Jurnal Masyarakat Informatika ISSN: 2086-4930 Vol. 14, No. 2, 2023 e-ISSN: 2777-0648.

[18]. Shi, L., Wang, X., & Shen, Y. (2020). Research on 3D face recognition method based on LBP and SVM. Optik, 220, 165157.

[19]. Sugiyono. (2021). Metode Penelitian Kuantitatif, Kualitatif, dan R&D. Bandung: Alfabeta.
[20]. Vadali, S., Deekshitulu, G. V. S. R., & Murthy, J. V. R. (2019). Analysis of liver cancer using data mining SVM algorithm in MATLAB. In Soft Computing for Problem Solving (pp. 163-175). Springer, Singapore
[21]. Vapnik, V. (1998). Statistical Learning Theory. Wiley.
[22]. M. L. Kolling, L. B. Furstenau, M. K. Sott, B. Rabaioli, P. H. Ulmi, N. L. Bragazzi, dan L. P. C. Tedesco, "Data Mining in Healthcare: Applying Strategic Intelligence Techniques to Depict 25 Years of Research Development," Int. J. Environ. Res. Public Health, vol. 18, no. 6, p. 3099, Mar. 2021, doi:10.3390/ijerph18063099.
[23]. S. Ataallah Muhammed dan A. R. Aziz, "Predictive Modeling in Healthcare: A Survey of Data Mining Techniques," ICCECS Proc., 2024.
[24]. Bánf, M. (2019). Learning Theory and Support Vector Machines – A Primer. arXiv.
[25]. Blanchard, G., Bousquet, O., & Massart, P. (2008). Statistical performance of support vector machines. arXiv.