Indonesian Journal of Data and Science

Volume 6 Issue 3
ISSN 2715-9936
https://doi.org/10.56705/ijodas.v6i3.261

Research Article

Classifying Honors Class Eligibility Using SVM

Khulfani Hendrawan [1,*]; Ika Arfiani [2]
[1] Ahmad Dahlan University, Yogyakarta, Indonesia, khulfani2100018308@webmail.uad.ac.id
[2] Ahmad Dahlan University, Yogyakarta, Indonesia, ika.arfiani@tif.uad.ac.id

Correspondence should be addressed to Khulfani Hendrawan; khulfani2100018308@webmail.uad.ac.id

Received 02 September 2025; Accepted 20 November 2025; Published 31 December 2025

Abstract:
Introduction: The honors class program aims to group outstanding students, but an objective data-based classification system is not yet available. This can result in high-potential students going undetected due to the selection process relying on self-registration and causing a lack of student interest. Method: This research uses the Support Vector Machine (SVM) algorithm to classify the eligibility of students to participate in the honors program based on academic and non-academic data. The dataset consists of 453 entries with an imbalanced class distribution, which was then balanced using the SMOTE technique. The model was trained using GridSearchCV to find the optimal parameters and compared with four types of SVM kernels: linear, polynomial, sigmoid, and radial basis function (RBF). Result: The RBF kernel achieved the best performance with an accuracy of 84%, precision of 0.77, recall of 0.84, and an F1-score of 0.79. However, it was found that the precision in the minority class (high-achieving students) is still lower than in the majority class. Conclusion: The SVM model, particularly with the RBF kernel, has proven effective in automating the classification of students for the honors program. However, further improvements are needed to enhance performance on the minority class to make the selection system fairer and more accurate.

Keyword: Classification, Support Vector Machine, Academic Achievement, Supervised Learning, Elite Class.

1. Introduction:

Improving the quality of human resources has become one of the main factors in facing the rapid development of technology [1]. in the digital era, the ability to adapt and master new technologies becomes the key to success for both individuals and organizations [2]. The advancement of increasingly sophisticated technology, which is widely spreading across various aspects of life, also has a significant impact on the world of education, where technology plays an important role in the learning process by providing various media designed to be easy and engaging [3]. In the academic context, technology has become one of the strategic innovations for monitoring and analyzing the progress of each student effectively and efficiently [4]. The use of technology-based academic information systems has proven to assist higher education institutions in designing more targeted learning strategies and interventions [5].

The graduation rate of students is one of the important indicators in assessing the success of a higher education institution in conducting the Teaching and Learning Activities [6]. Higher education institutions with high graduation rates tend to be more trusted by the public and are considered capable of producing competent graduates. This also adds value to the institution's promotional aspects, as prospective new students usually consider graduation rates when choosing their target university [7]. Therefore, the management and improvement of the quality of educational programs become a primary focus, especially in flagship programs such as the achievement classes. This program is specifically designed for selected students with high academic potential and is expected to produce outstanding graduates in their respective fields [8]. The success of the excellence class program will have a positive impact not only on the students but also on the image of the university both nationally and internationally. Higher education institutions that can produce outstanding graduates from their excellent programs will increasingly be recognized and trusted as high-quality educational institutions [9].

One of the study programs at a private university that has been accredited with an A and has been established since 1994 aims to produce Bachelor of Computer Science graduates who are capable of solving problems in the field of information and communication technology. This study program also organizes an achievement class program that has been running since 2023. The program is intended for the 2022 cohort students who have achieved well in both academic and non-academic fields, with the aim of grouping them into one class to share information, form discussion forums, and foster a positive culture of competition. The requirements to participate in this achievement class program include a minimum GPA of 3.25, a minimum TOEFL score of 450, achievements in academic or non-academic fields, and an interest in continuing to excel and compete.

One of the issues faced in the implementation of the excellence class program is the low interest of students to register independently. Based on a survey conducted among students of the excellence class and the Secretary of the Informatics Study Program, only 41 out of a total of 453 students from the 2022 cohort registered for this program. This figure indicates that student interest in the honors class is relatively low, which is suspected to be due to the perception that the program is too time-consuming, difficult, and competitive. However, the fundamental issue lies not only in the low interest but also in the absence of a system capable of automatically identifying and classifying students who actually have academic and non-academic potential to enter the excellence class program. Dependence on voluntary registration can cause high-achieving students to go undetected and miss out on the opportunities they deserve. This research aims to develop a data mining-based classification system capable of categorizing all students from the 2022 cohort into achievement and non-achievement categories without relying on the registration process.

Classification is the process of grouping an object into a specific category or class based on the characteristics or features possessed by that object [10]. Classification becomes very important in the data mining process, as a large amount of raw data will be processed and grouped based on certain characteristics to generate meaningful information[11], [12]. In Classification, there are several methods commonly used, such as Naïve Bayes, k-Nearest Neighbor, Random Forest, Decision Tree, Logistic Regression, and Support Vector Machine. Therefore, the classification process is very important in selection and assessment, especially in classifying students in an achievement class program, one of the commonly used methods in this case is the Support Vector Machine method. SVM is one of the leading classification methods in data mining due to its high accuracy, flexibility with non-linear data, and strong generalization capability. The basic operation of SVM begins with linear classification, which is then developed to handle non-linear data problems by adding a kernel function to the model. With this approach, SVM is capable of performing classification well on both linear and non-linear data. SVM is also used to find the best hyperplane by maximizing the distance between classes and maximizing the margin between the two classes [13]-[18].

This research uses the Support Vector Machine (SVM) method, one of the supervised learning algorithms that excels in classification tasks, both for linear and non-linear data. Compared to other methods such as Decision Tree, Naïve Bayes, or k-Nearest Neighbor. The application of the SVM algorithm to build a classification system for student eligibility in the excellence class program, considering both academic and non-academic attributes in an integrated manner. In addition, this research contributes to the development of educational data mining by demonstrating how classification techniques can be used to support data-driven selection processes and minimize potential bias in placing students into top programs at universities. SVM has the advantage of finding the optimal hyperplane with a maximum margin, and it can work effectively even with a high number of features or imbalanced data distribution. These advantages make SVM very suitable for classifying students based on complex academic and non-academic data. This research uses data from active students of the 2022 batch of the Informatics Study Program, which includes information such as GPA, study duration, TOEFL scores, as well as records of academic and non-academic achievements available in the campus information system. The aim of this research is to develop an SVM-based classification system that can objectively categorize students into merit-based programs, thereby supporting a more efficient and fair selection process, including for students who do not participate in self-registration.

2. Method:

This research follows systematic stages in the data mining process. Each stage plays an important role in ensuring the quality of the data and the accuracy of the classification model being built. **Figure 1** illustrates the complete flow of the research process, which consists of five main stages, from data collection to model evaluation [19].

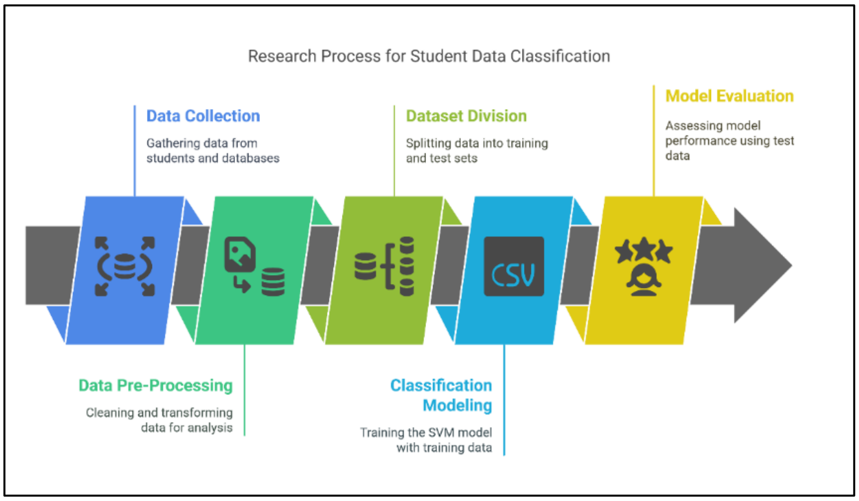

Figure 1. Stages of Data Mining Research

Data Collection

Data collection is the initial step in this research, using data from 453 students of the 2022 batch of the Informatics Study Program. The data was obtained through interviews and direct requests to the Student and Alumni Bureau (Bimawa) and the Administration (TU), which includes student ID numbers, names, GPA, as well as academic achievements (e.g., competitions, scientific works) and non-academic achievements (e.g., organizations, social activities) in Excel format. This dataset was then used as training and testing data in the classification process using the Support Vector Machine (SVM) method.

Data Pre-Processing

This stage includes several important processes in data preparation, namely data cleaning, normalization, and transformation. First, normalization is performed on numerical attributes such as the Cumulative Grade Point Average (GPA) and TOEFL scores to ensure that all data is on a uniform scale [20], [21]. If there are categorical attributes, an encoding process is performed to convert the data into a numerical format that can be processed by the model. Finally, to handle missing data, two approaches are applied. For numerical attributes, missing values are imputed using the mean (mean imputation), while for categorical attributes or labels with missing data, the related entries are deleted to maintain the integrity and quality of the dataset.

Dataset Division

The process of dividing the dataset, the initial stage is to separate the data into two main parts: training data and testing data. The division of the dataset into 70% for training and 30% for testing was chosen to ensure that the proportion of test data is sufficiently representative to objectively evaluate the model's performance. This proportion is used in the literature for medium-sized datasets and is considered a midpoint between the risks of overfitting and underfitting. Before the training process is carried out, an initial analysis of the class distribution in the dataset is conducted. The results show an imbalance, where the Non-Performing class is far more dominant compared to the Performing class. Checking the class distribution is an important step at this stage to ensure that the developed model can generalize well. The results of the data distribution visualization are presented in **Figure 2**.

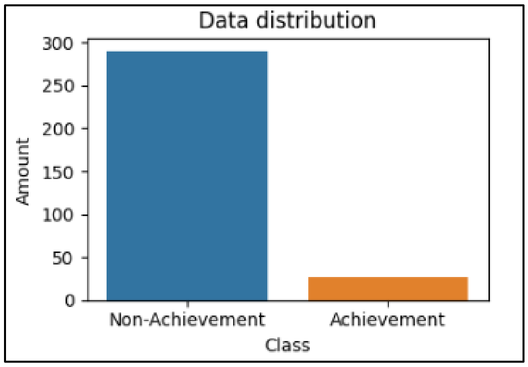

Figure 2. Data Distribution

Building a Classification Model with Support Vector Machine

The model used to classify study periods is the Support Vector Machine (SVM) with various types of kernels. SVM is a supervised learning algorithm used to address classification problems by finding the best hyperplane that can separate data from two or more classes with a maximum margin. To handle data that cannot be linearly separated, SVM uses kernel functions to map the data into a higher-dimensional feature space, allowing for more flexible decision boundaries to be formed [22], [23]. This research will conduct tests on four types of kernels, namely linear, radial basis function (RBF), polynomial, and sigmoid. Each kernel has different characteristics in mapping data. The linear kernel is used for data that tends to be linearly separable. The RBF and polynomial kernels are suitable for non-linear data, with RBF being more flexible in forming complex non-linear patterns. The sigmoid kernel mimics the activation function in neural networks and is used in the context of logistic relationships between features. RBF Kernel is often used in classification cases involving non-linear data with complex patterns. RBF implicitly maps data to a very high-dimensional feature space and calculates the influence of each data point based on its distance. This approach allows the model to flexibly adjust the shape of the decision boundary, especially when the data does not have a clear linear pattern. This characteristic makes the RBF kernel highly relevant for use in student classification cases, where academic and non-academic data often have non-linear and varying relationships. The hyperparameter tuning process was conducted using GridSearchCV with 5-fold cross-validation to evaluate various parameter combinations. The range of parameters tested includes C = [0.1, 1, 10, 100], gamma = [0.01, 0.1, 1, 10], and kernel types = ['linear', 'poly', 'rbf', 'sigmoid']. The optimal combination was found at C=10, gamma=10, and RBF kernel, which provided the highest accuracy on the test data. Therefore, selecting the optimal combination of C, gamma, and kernel is crucial to ensure maximum classification performance [24], [25].

Model Evaluation

Model evaluation aims to measure the performance and effectiveness of the model in making predictions. In this study, two main techniques are used, namely cross-validation and confusion matrix. Cross-validation is a validation method that divides data into two main parts using the k-fold cross-validation technique, which is one of the commonly used approaches in classification [26]. In this technique, the dataset is divided into several folds. This approach helps obtain a more comprehensive picture of the model's performance and reduces the risk of overfitting caused by unrepresentative data splitting.

After the cross-validation process, the evaluation continues using a confusion matrix to assess the prediction quality in more detail. The confusion matrix presents information in the form of a table that illustrates the number of correct and incorrect predictions for each class, namely True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN). Based on this matrix, evaluation metrics such as accuracy, precision, recall, and F1-score can be calculated, providing a comprehensive overview of the classification model's performance [27] - [32].

There are 4 types of matrices used for calculating the final score and classified as follows :

a. Accuracy is an evaluation metric that measures the proportion of total correct predictions from all cases.

$$Accuracy = \frac{TP+TN}{TP+TN+FN+FP} \tag{1}$$

b. Precision is an evaluation metric in classification that measures the proportion of true positive predictions out of all positive predictions.

$$Precission = \frac{TP}{TP+FP} \tag{2}$$

c. Recall is a metric used to measure a system's ability to identify all relevant positive instances.

$$Recall = \frac{TP}{TP+FN} \tag{3}$$

d. F1-Score is an evaluation metric that presents a balance between precision and recall in the form of the harmonic.

$$F1 - Score = 2\ x\ \frac{precission\ x\ recall}{precission+recal} \tag{4}$$

By combining cross-validation and the confusion matrix, model evaluation becomes more comprehensive and provides a stable estimation of model performance, while the confusion matrix shows the details of prediction errors. Both techniques are crucial in ensuring the reliability and accuracy of the classification model. The confusion matrix model used in this study is displayed in **Table 1** [33], [34].

Table 1. Confusion Matrix Model

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

Confusion Matrix displays the number of true and false as follows :

a. True Positive (TP), which is a sample of positive data predicted as positive.

b. True Negative (TN), which means a sample of negative data predicted as negative.

c. False Positive (FN), which is a sample of negative data predicted as positive.

d. False Negative (FN), which is a positive data sample predicted as negative.

3. Result and Discussion:

Result:

The classification model for the eligibility of outstanding class students is divided into five main stages: Data Collection, Data Preprocessing, Dataset Division, Classification Model, and Model Evaluation

Data Collection

Data was obtained through interviews and direct requests to the Student and Alumni Affairs Bureau (Bimawa) and the Administration Office of Ahmad Dahlan University. The data includes student ID numbers, names, GPAs, TOEFL scores, and academic achievement records manually labeled into two classes, namely achievement and non-achievement, and stored in Excel format. This dataset is used as training and testing data in the classification process using the Support Vector Machine method.

Data Pre-Processing

Processing Stage Data refers to the Knowledge Discovery in Database (KDD) process, which consists of data cleaning, data selection, and data transformation. The data of the 2022 student cohort was cleaned of duplicates and null values, and then relevant attributes such as GPA and achievement data were selected. Next, the data is transformed into a numerical format for classification purposes. Achievement labels are encoded into binary values, with 1 for students who have achievements and 0 for those who do not have achievements. The data of the 2022 Computer Science students that have been processed is presented in **Table 2**.

Table 2. Dataset Quote

Nim	Name	Ip_Semester	Achievement	Toefl_Score	Description
2200018484	Faisal Hairullah	3.02	0	580	*Tidak Prestasi*
2200018485	Fahreza Rifky Ferdiansyah	2.86	0	420	*Tidak Prestasi*
2200018486	Dani Sulaiman	3.39	0	376	*Tidak Prestasi*
2200018487	Pulung Kartiko Aji	2.99	0	374	*Tidak Prestasi*
2200018488	Yusuf Adi Jaya	1.05	0	300	*Tidak Prestasi*
2200018165	Yoga rusydi arifin	4.0	1	493	*Prestasi*
2200018197	Muhamad fadhli akbar	3.55	1	456	*Prestasi*
2200018183	Andi bintang toar dondok	3.75	1	460	*Prestasi*
2200018189	Akyas muhammad zaidan	3.75	1	620	*Prestasi*
2200018410	Aisyah syafi'i nurjannah	3.8	0	560	*Prestasi*

Dataset Division

A dataset of 453 students was divided into 70% training data and 30% testing data. Before modeling, the data was normalized using StandardScaler so that each feature has a distribution with a mean of zero and a standard deviation of one. The initial analysis results show an imbalance in class distribution, where non-achieving students are far more numerous than achieving students. To address this, the Synthetic Minority Over-sampling Technique (SMOTE) method was applied to the training data. SMOTE generates synthetic data for the minority class, thereby balancing the class distribution. This step aims to prevent model bias towards the majority class and improve accuracy in recognizing both classes. The results of the data distribution visualization after SMOTE are presented in **Figure 3**.

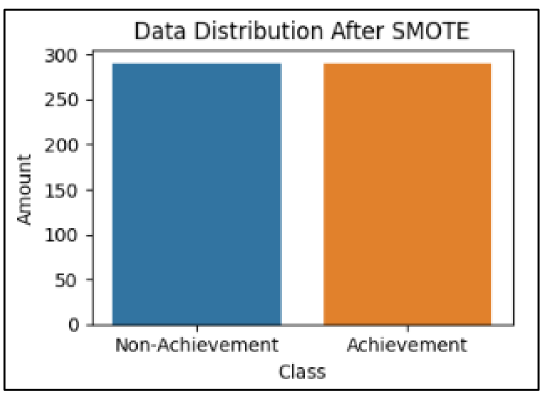

Figure 3. Data Distribution After SMOTE.

Classification Results with the Support Vector Machine Algorithm

At this stage, student classification is performed using the Support Vector Machine (SVM) algorithm implemented through a scikit-learn pipeline. The pipeline consists of three main stages: data normalization with StandardScaler, class balancing using SMOTE, and classification with SVC. To achieve optimal performance, parameter tuning was performed using GridSearchCV with 5-fold cross-validation, which evaluated various combinations of the parameters C, gamma, and kernel. The best results were obtained with the combination of C = 10, gamma = 10, and kernel = 'rbf', with an accuracy of 0.912. The selection of the RBF kernel indicates that the relationships between features are non-linear, thus requiring non-linear transformations for more accurate classification. The integration of preprocessing techniques, data balancing, and parameter tuning has proven effective in improving model performance.

The analysis of the SVM model's performance continued by comparing the accuracy of each type of kernel, specifically testing four main kernels, linear, polynomial, radial basis function (RBF), and sigmoid. Each kernel is applied to the data resulting from preprocessing and balancing using the SMOTE method. The visualization of accuracy comparison shown in **Figure 3** indicates that the rbf kernel provides the highest accuracy of 0.9121, followed by the polynomial kernel with an accuracy of 0.8310, and both the linear and sigmoid kernels achieving an accuracy of 0.8707. This difference indicates that the rbf kernel is better at capturing non-linear patterns in the data, resulting in better classification performance compared to other kernels. These results reinforce previous findings from the grid search process that the selection of the best parameters involves using the RBF kernel with C = 10 and gamma = 10. Thus, the rbf kernel can be considered the optimal configuration for the classification model in this study. The results of the accuracy comparison visualization for each kernel are presented in **Figure 4**.

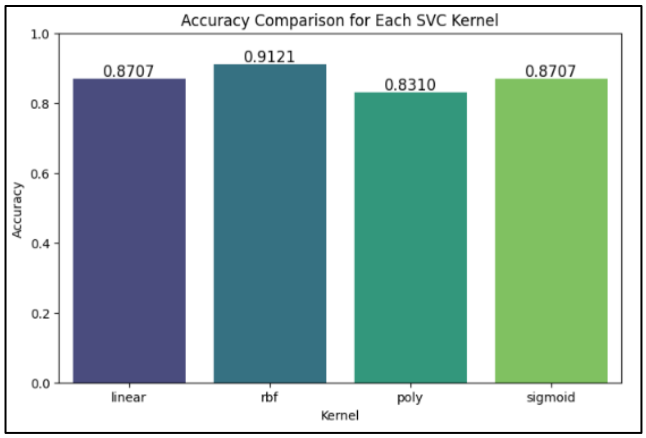

Figure 4. Accuracy comparison visualization

Model Evaluation

To evaluate the performance of each kernel that has been run in the SVM (Support Vector Machine) algorithm, calculations were performed on three main metrics: Precision, Recall, and F1-Score. **Figure 5** presents a comparison of the values of these three metrics for each type of kernel, namely linear, rbf, poly, and sigmoid. This evaluation is important to determine the extent to which the model can perform classification accurately and evenly, especially in the context of datasets that may be imbalanced or have certain characteristics.

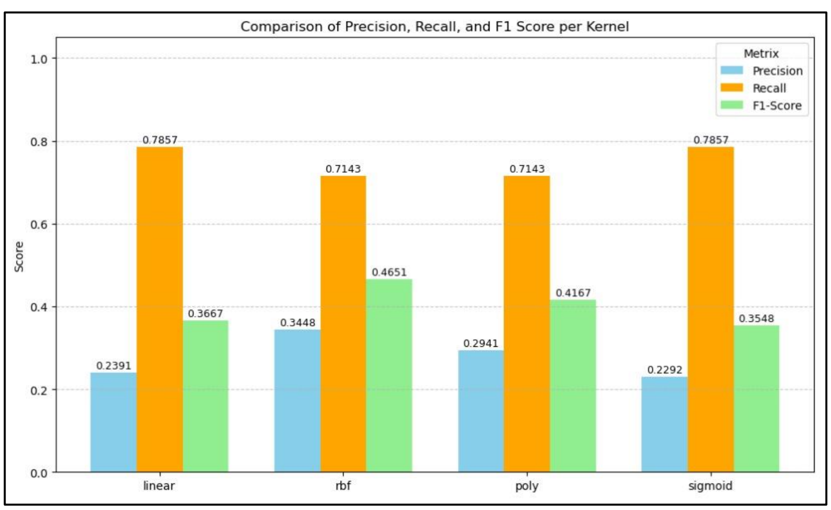

Figure 5. Comparison of Precision, Recall, and F1-Score per kernel

Figure 5 shows a comparison of Precision, Recall, and F1-Score values for four types of SVM kernels. It is evident that the RBF kernel demonstrates the best overall performance, with the second-highest Precision value (0.3448) and the highest F1-Score (0.4651) compared to other kernels. This indicates that the RBF kernel is able to provide a better balance between precision and recall, thereby minimizing classification errors more consistently. The linear kernel has the highest Recall value (0.7857), but the lowest Precision value (0.2391), causing the F1-Score to drop to 0.3667. This indicates that the linear kernel can recognize more positive classes, but with many false positive predictions. The polynomial kernel has stable performance with a Precision value of 0.2941 and a Recall of 0.7143, resulting in an F1-Score of 0.4167, which is still lower than RBF, indicating balanced performance but not yet as optimal as RBF in minimizing errors. The sigmoid kernel has the highest Recall value equal to linear (0.7857), but the Precision value is very low (0.2292), with an F1-Score of only 0.3548, making it less stable in maintaining the balance between precision and recall. Overall, this graph shows that the choice of kernel significantly affects the performance of the SVM model, and the RBF kernel is the best choice in this case based on the balance of the three displayed metrics (Precision, Recall, and F1-Score).

The classification evaluation using the Support Vector Machine algorithm is presented in **Table 3**. Based on the evaluation results with model_best = grid_search.best_estimator_, which represents the overall best model performance from the combination of the most optimal parameters, the SVM model with the RBF kernel achieved an accuracy of 84% on the test data. In the Non-Performing class (label 0), the model shows high performance with a precision of 0.96, recall of 0.85, and an f1-score of 0.90, indicating its excellent ability to identify the majority class. However, the performance on the Achieving class (label 1) still needs improvement. Although the recall score is quite high, the performance imbalance between classes is still noticeable. Overall, the weighted average f1-score reached 0.86, reflecting the model's quite good performance.

Table 3. Clasification Report.

Evaluation of the Best Model				
	Precision	Recall	F1-score	Support
0	0.96	0.85	0.90	122
1	0.36	0.71	0.48	14
Accuracy			0.84	136
Macro Avg	0.66	0.78	0.69	136
Weighted Avg	0.90	0.84	0.86	136

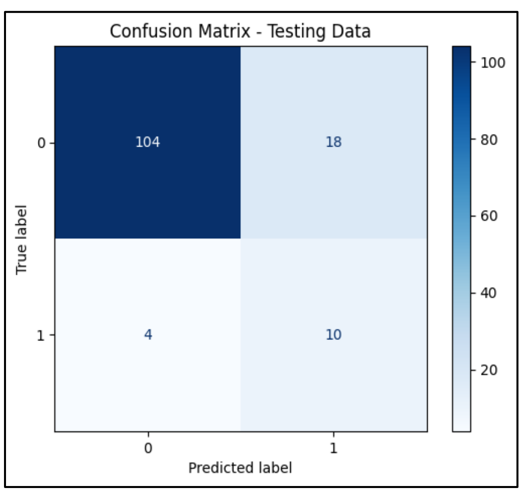

Figure 6. Graph of the confusion matrix results for the RBF model

To provide a more detailed overview of the performance of each kernel in classifying the data, **Figures 6** and **7** display the confusion matrix graphs for each SVM kernel model used, namely linear, RBF, polynomial, and sigmoid. The confusion matrix helps us not only see how accurate the model is overall but also evaluate how the model handles each class specifically, especially in the context of binary classification: Achievement and Non-Achievement. In **Figure 5**, it can be seen that the RBF model shows high performance.

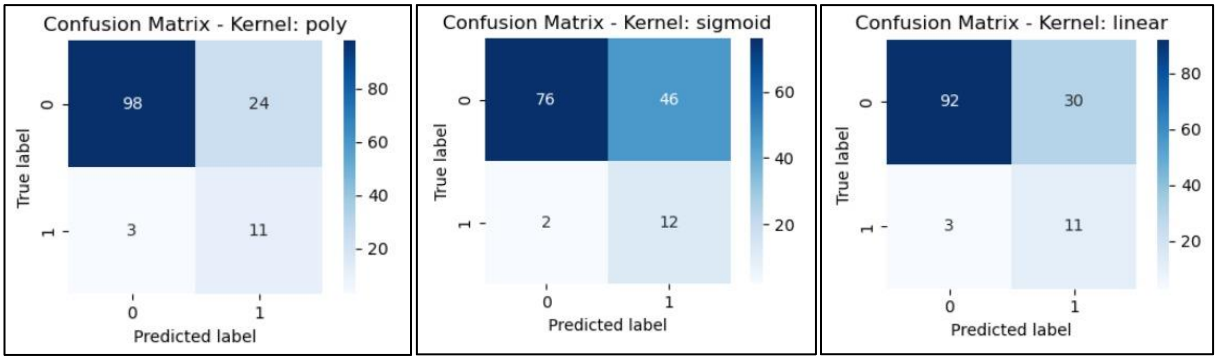

Figure 7. Comparison of Confusion Matrix Graphs for Linear, Poly, and Sigmoid Models

The Confusion Matrix results show that the model was able to correctly classify 104 out of 122 students in the Non-Achiever class (True Negative), but incorrectly classified 18 of them as Achievers (False Positive). Meanwhile, for the Achiever class, 10 students were correctly classified (True Positive), and 4 students were incorrectly classified as Non-Achievers (False Negative). In general, these results indicate that the model is more accurate in recognizing the majority class (Non-Performing) compared to the minority class (Performing). However, the model's ability to detect the minority class remains quite good, indicating that the SMOTE approach during the training phase successfully helped improve sensitivity to that class. The complete results of the confusion matrix are displayed in **Table 4.**

Table 4. Confusion Matrix Result

		Actual Value		
Prediction Value		**Actual Class Non-Achievement**	**Actual Class Achievement**	**Amount**
	Prediction of Non-Achievement Class	104	18	122
	Prediction of Achievement Class	4	10	14
Amount		108	28	136

Discussion

This research shows that the Support Vector Machine (SVM) algorithm with the Radial Basis Function (RBF) kernel is capable of classifying students into the categories of high achievers and non-high achievers with an accuracy of 84%. However, challenges remain in the minority class, namely high-achieving students, who tend to have lower precision values. This indicates a risk of classification errors, where non-performing students may be incorrectly classified as performing students. In the context of placing students in an honors program, this error can impact the fairness and effectiveness of the selection process. To address the data imbalance, the SMOTE technique was applied, which successfully improved the recall for the minority class. Further research is recommended to combine SMOTE with other approaches such as threshold adjustment, ensemble learning, or cost-sensitive classification, as well as exploring other algorithms like Random Forest and XGBoost. Thus, it is expected that the developed classification model can produce fairer and more accurate predictions for all classes, as well as support objective decision-making in student placement into programs.

4. Conclusion:

This study shows that the Support Vector Machine (SVM) algorithm with the Radial Basis Function (RBF) kernel is capable of classifying students into the achievement class program with an accuracy of 84%, precision of 0.77, recall of 0.84, and an F1-score of 0.79 using academic and non-academic data. These results reinforce the potential of SVM as a tool to assist in an objective and efficient selection process, especially for students who do not participate in self-registration. This research has limitations, particularly regarding the small size of the positive class (high-achieving), which results in the model's performance in detecting the minority class being suboptimal, especially in terms of precision. Therefore, for future research, it is recommended to apply additional techniques such as threshold tuning, class weighting adjustment, or exploration of other algorithms like Random Forest and XGBoost to improve recall and prediction accuracy for the minority class. With this approach, the classification model is expected to support the placement process of students into the excellence program more fairly and comprehensively.

Acknowledgments:

The author expresses the deepest gratitude to the Bureau of Student Affairs and Alumni (Bimawa) and the Administrative Office of the Faculty of Industrial Technology, who have provided data and support throughout the research process. Special thanks also to Mrs. Ika Afriani, the supervising lecturer, who has provided invaluable guidance and input, as well as to family and friends who have always offered encouragement and moral support. I hope this research can provide benefits and serve as a reference for similar studies in the future.

References:

[1] S. S. Gadzali, “Impact of Technology in Improving the Quality of Education and Human Resource Development,” *Indo-MathEdu Intellectuals J.*, vol. 4, no. 2, hal. 1337–1348, Okt 2023, doi: 10.54373/imeij.v4i2.362.

[2] N. L. W. S. R. Ginantra *et al.*, *Data Mining dan Penerapan Algoritma*, vol. 1. Yayasan Kita Menulis, 2021.

[3] A. Bozkurt, “Educational Technology Research Patterns in the Realm of the Digital Knowledge Age,” *J. Interact. Media Educ.*, vol. 2020, no. 1, hal. 18, Sep 2020, doi: 10.5334/jime.570.

[4] N. Bianchi, Y. Lu, dan H. Song, “The Effect Of Computer-Assisted Learning On Students’ Long-Term Development,” Cambridge, MA, 28180, Des 2020.

[5] A. M. R. B. Makkaraka, Akbar Iskandar, dan Wang Yang, “Design of Web-Based Student Academic Information System,” *Ceddi J. Educ.*, vol. 3, no. 2, hal. 9–15, Des 2024, doi: 10.56134/cje.v3i2.102.

[6] A. Salam, J. Zeniarja, dan D. M. Anthareza, “Student Graduation Prediction Model using Deep Learning Convolutional Neural Network (CNN),” in *2022 International Seminar on Application for Technology of Information and Communication (iSemantic)*, IEEE, Sep 2022, hal. 362–366. doi: 10.1109/iSemantic55962.2022.9920449.

[7] N. Por, C. Say, dan S. Mov, “Factors Influencing Students’ Decision In Choosing Universities: Build Bright University Students,” *J. As-Salam*, vol. 8, no. 1, hal. 1–15, Jan 2024, doi: 10.37249/assalam.v8i1.646.

[8] Z. Lazić, A. Đorđević, dan A. Gazizulina, “Improvement of Quality of Higher Education Institutions as a Basis for Improvement of Quality of Life,” *Sustainability*, vol. 13, no. 8, hal. 4149, Apr 2021, doi: 10.3390/su13084149.

[9] W. Zhao, “Honor classes in China’s higher education: perspectives from four universities,” *Natl. Sci. Rev.*, vol. 9, no. 3, Mar 2022, doi: 10.1093/nsr/nwac002.

[10] Sumiyatun, Y. Cahyadi, dan E. Faizal, “Implementation of Support Vector Machine Algorithm for Classification of Study Period and Graduation Predicate of Students,” *Indones. J. Data Sci.*, vol. 6, no. 1, hal. 56–64, Mar 2025, doi: 10.56705/ijodas.v6i1.214.

[11] R. Hazra, M. Banerjee, dan L. Badia, “Machine Learning for Breast Cancer Classification With ANN and

Decision Tree," in *2020 11th IEEE Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)*, IEEE, Nov 2020, hal. 0522–0527. doi: 10.1109/IEMCON51383.2020.9284936.

[12] Y. Wang, H. Wang, dan Z. Peng, "Rice diseases detection and classification using attention based neural network and bayesian optimization," *Expert Syst. Appl.*, vol. 178, hal. 114770, Sep 2021, doi: 10.1016/j.eswa.2021.114770.

[13] G. Manikandan, B. Pragadeesh, V. Manojkumar, A. L. Karthikeyan, R. Manikandan, dan A. H. Gandomi, "Classification models combined with Boruta feature selection for heart disease prediction," *Informatics Med. Unlocked*, vol. 44, hal. 101442, 2024, doi: 10.1016/j.imu.2023.101442.

[14] M. S, A. Sharma, S. P. Singh, V. Solanki, S. Sethuramalingam, dan S. P. Singh, "RETRACTED ARTICLE: SVM-based compliance discrepancies detection using remote sensing for organic farms," *Arab. J. Geosci.*, vol. 14, no. 14, hal. 1334, Jul 2021, doi: 10.1007/s12517-021-07700-4.

[15] M. B. Haile, A. O. Salau, B. Enyew, dan A. J. Belay, "Detection and classification of gastrointestinal disease using convolutional neural network and SVM," *Cogent Eng.*, vol. 9, no. 1, Des 2022, doi: 10.1080/23311916.2022.2084878.

[16] W. Wu *et al.*, "An Intelligent Diagnosis Method of Brain MRI Tumor Segmentation Using Deep Convolutional Neural Network and SVM Algorithm," *Comput. Math. Methods Med.*, vol. 2020, hal. 1–10, Jul 2020, doi: 10.1155/2020/6789306.

[17] P. Purnawansyah *et al.*, "Comparative Study of Herbal Leaves Classification using Hybrid of GLCM-SVM and GLCM-CNN," *Ilk. J. Ilm.*, vol. 15, no. 2, hal. 382–389, Agu 2023, doi: 10.33096/ilkom.v15i2.1759.382-389.

[18] C. Chen, H. Seo, C. H. Jun, dan Y. Zhao, "Pavement crack detection and classification based on fusion feature of LBP and PCA with SVM," *Int. J. Pavement Eng.*, vol. 23, no. 9, hal. 3274–3283, Jul 2022, doi: 10.1080/10298436.2021.1888092.

[19] E. Ahmed, "Student Performance Prediction Using Machine Learning Algorithms," *Appl. Comput. Intell. Soft Comput.*, vol. 2024, no. 1, Jan 2024, doi: 10.1155/2024/4067721.

[20] L. Peng, Z. Lu, T. Lei, dan P. Jiang, "Dual-Structure Elements Morphological Filtering and Local Z-Score Normalization for Infrared Small Target Detection against Heavy Clouds," *Remote Sens.*, vol. 16, no. 13, hal. 2343, Jun 2024, doi: 10.3390/rs16132343.

[21] M. Sholeh dan E. K. Nurnawati, "Comparison of Z-score, min-max, and no normalization methods using support vector machine algorithm to predict student's timely graduation," 2024, hal. 040003. doi: 10.1063/5.0202505.

[22] N. Mamat, S. F. Mohd Razali, dan F. B. Hamzah, "Enhancement of water quality index prediction using support vector machine with sensitivity analysis," *Front. Environ. Sci.*, vol. 10, Jan 2023, doi: 10.3389/fenvs.2022.1061835.

[23] A. A. Kayode, N. O. Akande, A. A. Adegun, dan M. O. Adebiyi, "An automated mammogram classification system using modified support vector machine" *Med. Devices Evid. Res.*, vol. Volume 12, hal. 275–284, Agu 2019, doi: 10.2147/MDER.S206973.

[24] G. H. Coffie dan S. K. F. Cudjoe, "Toward predictive modelling of construction cost overruns using support vector machine techniques," *Cogent Eng.*, vol. 10, no. 2, Des 2023, doi: 10.1080/23311916.2023.2269656.

[25] V. Patel dan H. Bhavsar, "Optimizing Kernel Transformations to Handle Binary Class Imbalanced Dataset Classification," *Appl. Artif. Intell.*, vol. 38, no. 1, Des 2024, doi: 10.1080/08839514.2024.2408933.

[26] O. Oyedele, "Determining the optimal number of folds to use in a K-fold cross-validation: A neural network classification experiment," *Res. Math.*, vol. 10, no. 1, Des 2023, doi: 10.1080/27684830.2023.2201015.

[27] F. Westin, “Time Period Categorization in Fiction: A Comparative Analysis of Machine Learning Techniques,” *Cat. Classif. Q.*, vol. 62, no. 2, hal. 124–153, Feb 2024, doi: 10.1080/01639374.2024.2315548.

[28] H. A. Abdelaziz, M. Sallah, A. Elgarayhi, dan F. E. Al-Tahhan, “Accurate automatic classification system for 3D CT images of some vertebrate remains from Egypt,” *J. Taibah Univ. Sci.*, vol. 16, no. 1, hal. 632–645, Des 2022, doi: 10.1080/16583655.2022.2096391.

[29] M. Stusek *et al.*, “Accuracy Assessment and Cross-Validation of LPWAN Propagation Models in Urban Scenarios,” *IEEE Access*, vol. 8, hal. 154625–154636, 2020, doi: 10.1109/ACCESS.2020.3016042.

[30] Z. A. Haq, Z. A. Jaffery, dan S. Mehfuz, “Precision harvesting: comparative analysis of machine learning and generative AI-based classifiers for guava fruit maturity assessment using thermal imaging,” *CyTA - J. Food*, vol. 22, no. 1, Des 2024, doi: 10.1080/19476337.2024.2401588.

[31] P. Kartikasari, I. T. Utami, S. Suparti, dan S. D. F. Rahman, “Breast Cancer Classification Using Support Vector Machine (Svm) And Light Gradient Boosting Machine (Lightgbm) Models,” *MEDIA Stat.*, vol. 16, no. 2, hal. 182–193, Mei 2024, doi: 10.14710/medstat.16.2.182-193.

[32] M. S. Darweesh *et al.*, “Early breast cancer diagnostics based on hierarchical machine learning classification for mammography images,” *Cogent Eng.*, vol. 8, no. 1, Jan 2021, doi: 10.1080/23311916.2021.1968324.

[33] L. Dai, L. Chen, dan M. Luo, “Optimization model and application of linear and nonlinear MBSVM based on pinball loss function,” *Math. Comput. Model. Dyn. Syst.*, vol. 30, no. 1, hal. 898–923, Des 2024, doi: 10.1080/13873954.2024.2424854.

[34] B.-E. Boudriki Semlali, C. Molina, H. Park, dan A. Camps, “Association of land surface temperature anomalies from GOES/ABI, MSG/SEVIRI, and Himawari-8/AHI with land earthquakes between 2010 and 2021,” *Geomatics, Nat. Hazards Risk*, vol. 15, no. 1, Des 2024, doi: 10.1080/19475705.2024.2324982.