

Forward Selection as a Feature Selection Method in the SVM Kernel for Student Graduation Data

Hafis Nurdin^{*}, Irmawati Carolina²⁾, Resti Lia Andharsaputri³⁾, Anus Wuryanto⁴⁾, Ridwansyah⁵⁾

¹⁾Universitas Nusa Mandiri, Indonesia

²⁾³⁾⁴⁾⁵⁾ Universitas Bina Sarana Informatika, Indonesia

¹⁾ hafis.nnr@nusamandiri.ac.id, ²⁾ irmawati.imc@bsi.ac.id, ³⁾ resti.ria@bsi.ac.id, ⁴⁾ anus.awu@bsi.ac.id, ⁵⁾ ridwansyah.rid@bsi.ac.id

Submitted : Oct 8, 2024 | **Accepted** : Oct 21, 2024 | **Published** : Oct 30, 2024

Abstract: In the era of information technology development, accurate graduation predictions are important to improve the quality of higher education in Indonesia. This research evaluates the effectiveness of the Support Vector Machine (SVM) with various kernels, including Radial Basis Function (RBF), linear, and polynomial, as well as the application of Forward Selection (FS) as an optimization method. The dataset used consists of student graduation data which includes nine independent attributes and one label. This research aims to increase the accuracy of student graduation predictions using the SVM method which is optimized through Forward Selection (FS). The SVM method is applied using 10-fold cross validation to predict on-time graduation. The results show that the combination of SVM and FS improves prediction accuracy significantly. The SVM model with an RBF kernel optimized with FS achieved the highest accuracy of 87.06% and recall of 53.68%, showing increased sensitivity in identifying student graduation cases compared to SVM without FS. Although there is a trade-off between precision and recall, the model optimized with FS shows better performance overall. This research contributes to the development of a more efficient graduation prediction method, which can help universities in planning strategies to improve academic quality. Further studies are recommended to overcome weaknesses in the recall value by using other optimization methods or combinations of other optimization algorithms.

Keywords: Feature Selection; Forward Selection; Student Graduation; SVM.

INTRODUCTION

In the midst of the increasingly rapid development of information technology, the Indonesian Ministry of Research, Technology and Higher Education (Kemenristekdikti) is facing new challenges in ensuring the quality of higher education. One crucial aspect that is of particular concern and plays an important role in determining the quality of graduates and the effectiveness of educational institutions.

The aim of this research is to evaluate the effectiveness of FS in improving feature selection performance in the SVM Kernel in the context of student graduation data. This research is expected to produce new insights regarding effective feature selection strategies for predicting student graduation and contribute to the development of more accurate predictive methods in higher education. Through an experimental approach and analysis of student graduation data, it can improve the accuracy of the SVM Kernel and make a significant contribution to efforts to improve the quality of higher education in Indonesia, in line with the goals and vision of the Ministry of Research, Technology and Higher Education in facing the challenges of the digital era.

This research introduces new innovation by retesting the SVM method that has been used by previous researchers on the same dataset, but this time by applying different kernels in the SVM algorithm. Not only that, this research also involves a further optimization process for the SVM kernel through the application of the FS method. This step intends to improve the overall performance of the model. The use of different kernels is expected to provide more accurate variations in results, while optimization with FS is expected to increase the efficiency of feature selection, so that student graduation prediction models can be more precise and relevant to existing data. This was done to see to what extent the performance of the SVM algorithm can be improved through a combination of selecting the right kernel and more effective feature optimization.

*name of corresponding author



LITERATURE REVIEW

The SVM kernel is famous for its ability to process non-linear data (Purnama, Nawawi, Rosyida, Ridwansyah, & Risandar, 2019), being one of the methods that is widely used and one of the popular methods in machine learning (Ridwansyah, Ariyati, & Faizah, 2019). However, the main challenge in this research is that selecting effective and efficient features to improve model performance is still a challenge in developing accurate predictive models. So a method for selecting features is needed, namely FS (Wijaya, 2024).

FS is a feature selection technique that gradually adds features based on certain criteria to improve model performance by means of computational efficiency and model accuracy (Kurniadi, Nuraeni, & Lestari, 2022). Previous research has shown that selecting features one by one based on certain precise criteria can significantly increase prediction accuracy in various applications (Purwaningsih, 2022). However, there is a gap in the application of this method, especially in predicting student graduation using the SVM Kernel.

In the literature review, various studies have explored the application of machine learning techniques for predictions in the field of education. Previous research in the field of student graduation data analysis has shown significant variations in results in the accuracy of classification and optimization algorithms. Classification algorithm using machine learning algorithms such as Decision Tree (DT), Neural Network (NN), and Support Vector Machine (SVM). Previous studies recorded an accuracy of 84.96% for DT, 84.68% for NN, and 85.18% for SVM indicating relatively similar performance between the three methods (Riyanto, Hamid, & Ridwansyah, 2019). This shows the existence of gaps and potential for further exploration in improving the effectiveness of classification algorithms through optimization methods. Where the integration of optimization algorithms such as Particle Swarm Optimization (PSO) (Suhardjono, Wijaya, & Hamid, 2019) and Genetic Algorithm (GA) (Ridwansyah, Wijaya, & Purnama, 2020) from the results has shown a significant increase in accuracy. In particular, the combination of DT and PSO (Hendra, Azis, & Suhardjono, 2020), NN and PSO (Nurdin, Sartini, Sumarna, Maulana, & Riyanto, 2023), then NN and GA (Pangesti, Ariyati, Priyono, Sugiono, & Suryadithia, 2024) showed the most significant improvement reaching a value of up to 87.56 %, 86.94%, and 87.33% respectively. However, current research focuses on implementing the Forward Selection method in the SVM kernel which is an innovative approach to feature selection to increase the effectiveness of the model in predicting student graduation data. This approach is expected to overcome the limitations of previous methods and provide important insights for current research with the aim of developing more efficient and accurate methods considering technological advances and increasingly complex application requirements.

METHOD

Data mining has an important role in society throughout the world, especially in the information systems sector for education (Ridwansyah, Riyanto, Hamid, Rahayu, & Purnama, 2022). This is due to the ability to process large amounts of data into information that is relevant and according to needs. There are various applications aimed at forecasting which function to predict future events based on new data originating from historical data (Ridwansyah, Faizah, & Achyani, 2021). This research focuses on building a prediction model for student graduation by comparing the effectiveness of the SVM method and the FS algorithm by comparing the level of accuracy and kernel parameter values of the two methods. Apart from that, this research also analyzes the impact on model performance and feature subgroups obtained from forward selection results. The data is then integrated into the SVM training system. Several types of SVM kernel functions used include radial (RBF), dot (linear) and polynomial with C parameters ranging from 0.0 to 100. To determine the best method, this research uses accuracy calculations using the 10 k-fold cross validation method. split the data into training and testing parts in each experiment. The entire research process consists of several stages, to provide a deeper understanding of how this methodology works and its application in practice a detailed illustration of this classification process is presented in Figure 1.

*name of corresponding author



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

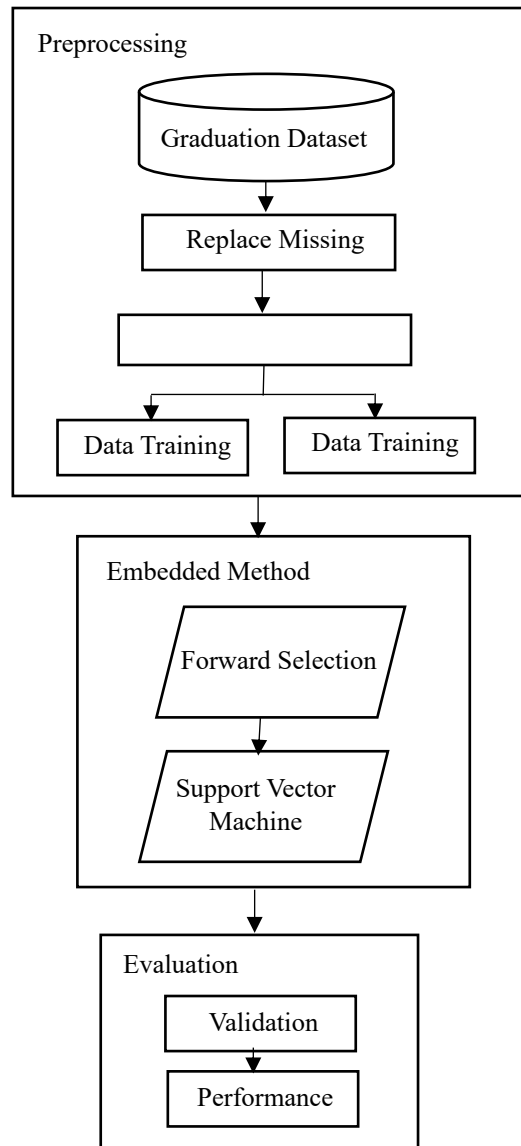


Fig. 1 Research Design

The SVM method begins with a normalization process, using a scale of minimum and maximum values. The Min-max technique used aims to convert the original data into a linear format in the range of values more than 1 and less than 1. The preprocessing process includes dividing the dataset into training and testing data, followed by 10-fold cross validation carried out using the rapid miner application. Next, the data is processed using the embedded feature selection method. Integral feature search in classification algorithms, as applied to SVM in this research (M Hafidz Ariansyah, Esmi Nur Fitri, & Sri Winarno, 2023).

SVM is integrated in the prediction algorithm. The application of SVM involves searching for distances in a dataset, which often shows non-linear data that cannot be separated linearly (Iqbal et al., 2020). Kernel functions used include radial (RBF), polynomial, and linear.

This problem can be overcome through Feature Space, which converts linear data into a high-dimensional space. This is known as the kernel trick, changed to K Kernel (x_i, x_j) (Zhao & Wu, 2020). The purpose of Feature Selection is to remove irrelevant attributes in the dataset. There are three main methods of feature selection:

1. Feature Based Selection: Assess each attribute with statistical methods. Rank the attributes, and identify whether they are relevant or not using the matrix that will be generated from the dataset first.
2. Embedded Feature Selection: Perform feature selection during training to get optimal results. This type of feature selection uses the principle of regularization with a limiting penalty factor.
3. Wrapper Feature Selection: Looking for the best attribute combination by comparing one attribute with another, using forward selection from an empty model until the attribute combination is met.

*name of corresponding author



Next, validation was carried out using a confusion matrix. Where this matrix has criteria such as True Positive (TP), which shows the ability of the prediction engine to identify positive values correctly, reflecting accurate results according to real conditions. This also applies to True Negative (TN), where the machine correctly predicts negative values. So, if the machine classifies a positive value as negative, then the result is wrong and can be considered a false negative, because the true value is positive. This means that the negative value produced by the prediction machine is wrong or can be called a false negative value, because the actual value should be positive. Validation was carried out using a confusion matrix. In this matrix, criteria such as True Positive (TP) and True Negative (TN) are used to measure prediction performance, including the variables False Positive (FP) and False Negative (FN) (Chicco & Jurman, 2020).

The evaluation results are a study of the results obtained from the optimized model and the trained model is evaluated using two criteria. First, validation is carried out using cross-validation to ensure the stability and trustworthiness of the model by involving the use of data sets that have never been seen by the model (usually a validation or test set) to assess how well the model can predict new data. These two Performances are model performance evaluated based on certain metrics such as accuracy, precision, recall and AUC. After both methods have been tested, both the SVM and Forward Selection methods, the results will be compared by comparing the level of accuracy between the support vector machine methods that have been optimized and those that have not been optimized to determine the success of the model optimization method.

RESULT

The dataset used is student graduation data obtained from private universities with nine independent attributes consisting of gender which has male or female gender data which means which gender has the enthusiasm to achieve graduation on time. High school major that influences the suitability of the major or study program before entering university. The student's Cumulative Achievement Index (GPA) from semester 1 to semester 6. Meanwhile, the attribute labels consist of 1 = Passed on time and 2 = Passed not on time with a total of 796 records in the dataset. So that the data is valid and the quality obtained is good, initial data processing is carried out or also called preprocessing, namely identifying data and cleaning problematic data by replacing lost data. Meanwhile, to change the record on the gender attribute (Male becomes 1), High School Major is grouped into 10 groups (1-10), High School Origin is grouped into 5 groups (1-5) and GPA every semester until the sixth semester which can be seen in table 1.

Table 1. Sample Graduation Data After Preprocessing

GEN	HSD	Origin_HS	CAI1	CAI2	CAI3	CAI4	CAI5	CAI6	OnTime
1	10	3	2,86	2,85	3,07	3,11	3,13	3,09	Yes
1	9	3	2,59	2,71	2,7	2,83	3	3,12	Yes
1	7	5	2,55	2,56	2,66	2,63	2,73	2,96	No
1	10	3	3,41	3,05	3,07	2,79	2,96	3,09	No
2	4	3	2,32	2,2	2,1	2,3	2,56	2,87	No
1	7	3	3,23	3,24	3,05	3,05	3,15	3,13	No
2	1	3	3,23	3,07	3,08	3,21	3,35	3,35	Yes
1	7	3	3,09	3,27	3,26	3,4	3,47	3,48	Yes
1	9	3	2,32	2,44	2,64	2,83	2,88	2,91	Yes
1	7	3	2,41	1,88	1,85	2,31	2,67	2,9	No
2	5	3	2,55	2,2	1,7	1,7	2,24	2,72	No
..
4	3	2,77	2,73	2,62	2,59	3,09	3,12	4	Yes

After data preprocessing has been carried out in Table 1 with the stage of replacing missing data, this data is then tested using the Support Vector Machine (SVM) method. In the student graduation process, the data is processed by applying 10 times cross validation by dividing it directly using the rapid data miner application into 90% training data and 10%.

Table 2. Confusion Matrix Kernel SVM

Algorithm	TP	TN	FP	FN
SVM (Dot)	612	22	68	94
SVM (RBF)	617	17	28	134
SVM (Ploynomial)	633	1	25	137

Based on Table 2. First SVM (Dot): The results show that this model has 612 True Positives (TP) and 22 True Negatives (TN), which means this model predicts 612 cases of student graduation and 22 cases of failure correctly. This model also has 68 FP and 94 FN, indicating errors in positive and negative predictions. Second

*name of corresponding author



SVM (RBF): This model has a slightly higher TP at 617 and a lower TN at 17, with a much lower FP at 28 and a higher FN at 134. Third SVM (Polynomial): This model has a slightly higher TP at 633 and a lower TN at 1, with a much lower FP at 25 and a higher FN at 137. This model is best at identifying positive cases but almost always wrong in identifying negative cases, almost all non-passing cases are predicted as passing.

After the graduation data has been tested using the SVM method with various kernels, testing will be carried out using forward selection. The results of this testing process will be presented in confusion matrix format, details of these results can be found in Table 3.

Table 3. Confusion Matrix Kernel SVM+FS

Algorithm	TP	TN	FP	FN
SVM (Dot)	618	16	96	66
SVM (RBF)	606	28	87	75
SVM (Polynomial)	634	0	27	135

The following is an explanation for each row in Table 2. First SVM (Dot): The results show that this model has 618 True Positives (TP) and 16 True Negatives (TN), which means this model predicts 618 cases of student graduation and 16 cases of failure correctly. This model also has 96 False Positives (FP) and 66 False Negatives (FN), indicating errors in positive and negative predictions. Second SVM (RBF): This model has a slightly higher TP at 606 and a lower TN at 28, with a much lower FP at 87 and a higher FN at 75. This indicates that this model is better at identifying cases positive (pass) but more often misidentify negative cases (not pass). Third SVM (Polynomial): Results show 634 TP and only 0 TN, with 27 FP and the highest FN at 135. This model is best at identifying positive cases but almost always wrong in identifying negative cases, almost all non-pass cases are predicted as passing.

This research conducted a comprehensive evaluation of the effectiveness of the Support Vector Machine (SVM) method and its combination with Feature Selection (SVM+FS) in processing graduation data. This process involves a series of careful tests and validation of the results obtained from both approaches. The results that attract attention especially are the significant improvements in key aspects such as accuracy, precision, recall, and Area Under Curve (AUC) when graduation data is analyzed using the SVM method integrated with Feature Selection. This evaluation is performed using specific metrics designed to measure model performance and effectiveness. Key indicators such as accuracy, precision, recall, and AUC are adopted as measuring tools to compare the two methods. The main objective of this research is to identify the level of accuracy of the method used and to determine whether the integration of Feature Selection with the SVM method results in increased effectiveness in dealing with issues related to student graduation. These improvements are explicitly noted and presented in Table 4, which highlights the performance of each method.

Table 4. Perbandingan performa metode

Algoritma	Accuracy	Precision	Recall	AUC
SVM (Dot)	85,43%	76,62%	42,17%	0,882%
SVM (RBF)	81,03%	64,55%	17,21%	0,831%
SVM (Polynomial)	82,66%	96,15%	15,48%	0,732%
SVM (Dot)+FS	85,93%	82,25%	40,70%	0,863%
SVM (RBF)+FS	87,06%	77,02%	53,68%	0,827%
SVM (Polynomial)+FS	83,04%	100,00%	16,76%	0,717%

From Table 4, we can see that the use of FS generally improves model performance, especially in terms of accuracy and precision.

SVM (RBF)+FS shows significant improvement in recall compared to SVM (RBF) without FS. There are trade-offs between various metrics, for example, SVM (Polynomial)+FS has very high precision (100%) but lower recall compared to several other models. Overall, this table provides a comparative overview of how the use of various SVM kernels and FS applications affects various aspects of model performance in the context of student graduation data. The following is a research evaluation based on the information available in the graph in Figure 2.

*name of corresponding author



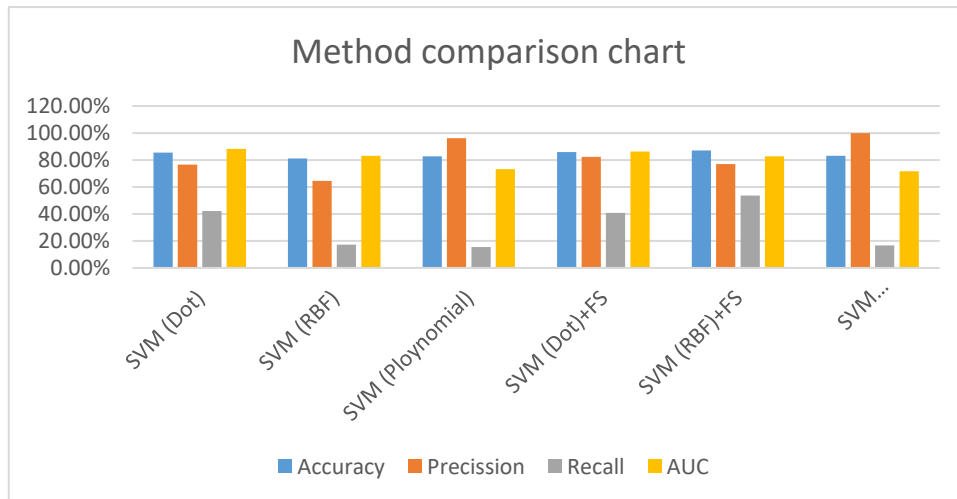


Fig 2. Comparison of the Results of the SVM Method with the SVM+FS Method

From Figure 2, a comparison graph of various SVM and FS kernels can be explained. Where the highest accuracy and recall values are found in the SVM (Radial)+FS method with an accuracy value of 87.06% and a recall of 53.68%. And the highest precision value is found in SVM (Polynomial)+FS with a value of 100%. And the highest AUC value is found in SVM (Dot) without FS with a value of 0.882%.

DISCUSSIONS

Based on the results of the trials carried out, there are several important findings related to the application of the Support Vector Machine (SVM) method with different kernels, as well as optimization using Forward Selection (FS) in predicting student graduation. The results show that the use of FS provides significant improvements in several model performance metrics, especially in terms of accuracy and precision, although there are some trade-offs between different metrics, such as recall and AUC. The best accuracy value obtained from this method was 87.06%, with a significant increase in recall of 53.68%. This indicates that the SVM+FS method is better at recognizing positive cases (student graduation) compared to SVM without FS, which shows that the addition of FS is able to increase the model's sensitivity to accurate graduation predictions. Meanwhile, the polynomial kernel in SVM optimized with FS shows the highest precision of 100%, but the recall is lower. This shows that although the model is very good at correctly predicting pass cases, there are still a large number of negative (not pass) cases that are incorrectly predicted as positive (pass).

From the results obtained, the author believes that SVM+FS provides a better approach for predicting student graduation, especially in terms of increasing accuracy and precision. The research results can help universities make better decisions regarding graduation and strategies to improve academic quality. Overall, this research provides new insights into the effectiveness of SVM methods optimized with Forward Selection, which can be applied more widely in the context of higher education. The increased accuracy and precision resulting from this combination shows great potential to improve predictions of student graduation, as well as helping universities identify key factors that influence student academic success.

CONCLUSION

This research has shown that the application of Support Vector Machine (SVM) with various kernels optimized using Forward Selection (FS), is able to improve the model's performance in predicting student graduation. The test results show a significant increase in accuracy and precision in several combinations of kernel and FS, especially in SVM (RBF) + FS which achieved the highest accuracy of 87.06% and recall of 53.68%. This improvement shows that the use of FS is successful in filtering out relevant features, thereby increasing the effectiveness of the prediction model.

The main limitation of this research is that further optimization is needed to increase the recall value in several kernels, so that predictions can be more balanced in identifying passing and non-passing students. In future studies, the application of other optimization methods that may be more efficient in overcoming the low recall problem in some kernels. In addition, further testing of SVM kernel variations and the application of hybrid methods with other optimization algorithms, such as Genetic Algorithm or Particle Swarm Optimization can be carried out to explore improving model performance.

The benefits of this research are very relevant to help universities identify factors that influence student graduation, so that they can be implemented as a predictive tool for better academic planning. With this

*name of corresponding author



optimization, it is hoped that educational institutions can take strategic steps to improve the quality of education and academic success of their students.

REFERENCES

- Chicco, D., & Jurman, G. (2020). The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics*, 21(1), 1–13. <https://doi.org/10.1186/s12864-019-6413-7>
- Hendra, Azis, M. A., & Suhardjono. (2020). ANALISIS PREDIKSI KELULUSAN MAHASISWA MENGGUNAKAN DECISION TREE BERBASIS PARTICLE SWARM OPTIMIZATION. *Jurnal Sisfokom (Sistem Informasi Dan Komputer)*, 9(1), 102–107. <https://doi.org/https://doi.org/10.32736/sisfokom.v9i1.756>
- Iqbal, M., Herliawan, I., Ridwansyah, Gata, W., Hamid, A., Purnama, J. J., & Yudhistira. (2020). Implementation of Particle Swarm Optimization Based Machine Learning Algorithm for Student Performance Prediction. *JITK (Jurnal Ilmu Pengetahuan Dan Teknologi Komputer)*, 6(2), 195–204. <https://doi.org/10.33480/jitk.v6i2.1695>
- Kurniadi, D., Nuraeni, F., & Lestari, S. M. (2022). Implementasi Algoritma Naïve Bayes Menggunakan Feature Forward Selection dan SMOTE Untuk Memprediksi Ketepatan Masa Studi Mahasiswa Sarjana. *Jurnal Sistem Cerdas*, 5(2), 63–82. <https://doi.org/https://doi.org/10.37396/jsc.v5i2.215>
- M Hafidz Ariansyah, Esmi Nur Fitri, & Sri Winarno. (2023). Improving Performance of Students' Grade Classification Model Uses Naïve Bayes Gaussian Tuning Model and Feature Selection. *Jurnal Teknik Informatika (Jutif)*, 4(3), 493–501. <https://doi.org/10.52436/1.jutif.2023.4.3.737>
- Nurdin, H., Sartini, Sumarna, Maulana, Y. I., & Riyanto, V. (2023). Prediction of Student Graduation with the Neural Network Method Based on Particle Swarm Optimization. *Sinkron : Jurnal Dan Penelitian Teknik Informatika*, 8(4), 2353–2362. <https://doi.org/10.33395/sinkron.v8i4.12973>
- Pangesti, W. E., Ariyati, I., Priyono, Sugiono, & Suryadithia, R. (2024). Utilizing Genetic Algorithms To Enhance Student Graduation Prediction With Neural Networks. *Sinkron : Jurnal Dan Penelitian Teknik Informatika*, 9(1), 276–284. <https://doi.org/https://doi.org/10.33395/sinkron.v9i1.13161> e-ISSN
- Purnama, J. J., Nawawi, H. M., Rosyida, S., Ridwansyah, & Risandar. (2019). Klasifikasi Mahasiswa Her Berbasis Algoritma Svm Dan Decision Tree. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 7(6), 1253–1260. <https://doi.org/10.25126/jtiik.202073080>
- Purwaningsih, E. (2022). Improving the Performance of Support Vector Machine With Forward Selection for Prediction of Chronic Kidney Disease. *JITK (Jurnal Ilmu Pengetahuan Dan Teknologi Komputer)*, 8(1), 18–24. <https://doi.org/10.33480/jitk.v8i1.3327>
- Ridwansyah, Ariyati, I., & Faizah, S. (2019). PARTICLE SWARM OPTIMIZATION BERBASIS CO-EVOLUSIONER DALAM EVALUASI KINERJA ASISTEN DOSEN. *Jurnal SAINTEKOM*, 9(2), 165–177. <https://doi.org/https://doi.org/10.33020/saintekom.v9i2.96>
- Ridwansyah, R., Faizah, S., & Achyani, Y. E. (2021). Mengidentifikasi Jenis Virus Menggunakan Sistem Pakar Berbasis Metode Forward Chaining. *Paradigma - Jurnal Komputer Dan Informatika*, 23(1), 49–54. <https://doi.org/10.31294/p.v23i1.10048>
- Ridwansyah, R., Riyanto, V., Hamid, A., Rahayu, S., & Purnama, J. J. (2022). Grouping Data in Predicting Infant Mortality Using K-Means and Decision Tree. *Paradigma*, 24(2), 168–174. <https://doi.org/10.31294/paradigma.v24i2.1399>
- Ridwansyah, R., Wijaya, G., & Purnama, J. J. (2020). Hybrid Optimization Method Based on Genetic Algorithm for Graduates Students. *Jurnal Pilar Nusa Mandiri*, 16(1), 53–58. <https://doi.org/10.33480/pilar.v16i1.1180>
- Riyanto, V., Hamid, A., & Ridwansyah. (2019). Prediction of Student Graduation Time Using the Best Algorithm. *Indonesian Journal of Artificial Intelligence and Data Mining*, 2(2), 1–9. <https://doi.org/http://dx.doi.org/10.24014/ijaidm.v2i1.6424>
- Suhardjono, Wijaya, G., & Hamid, A. (2019). PREDIKSI WAKTU KELULUSAN MAHASISWA MENGGUNAKAN SVM BERBASIS PSO. *Bianglala Informatika*, 7(2), 97–101. <https://doi.org/https://doi.org/10.31294/bi.v7i2.6654.g3731>
- Wijaya, G. (2024). *Improvement of Kernel SVM to Enhance Accuracy in Chronic Kidney Disease*. 9(1), 136–144. <https://doi.org/https://doi.org/10.33395/sinkron.v9i1.13112> e-ISSN
- Zhao, G., & Wu, Y. (2020). An efficient kernel-based feature extraction using a pull–push method. *Applied Soft Computing*, 96. <https://doi.org/https://doi.org/10.1016/j.asoc.2020.106584>

*name of corresponding author



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