

Improved SVM Classification Using Particle Swarm Optimization for Student Completion Prediction System

I Made Dwi Putra Asana^{1*}, I Dewa Gede Ari Oka², I Made Oka Widyantara³, I Made Subrata Sandhiyasa⁴

^{1,2,4}*Departement of Informatics, Institut Bisnis dan Teknologi Indonesia, Indonesia*

³*Departement of Electrical Engineering, Udayana University, Indonesia*

*corr_author: dwiputraasana@instiki.ac.id

Abstract – Timely completion of a study program is crucial for evaluating the quality of universities. To achieve timely completion, student's progress needs to be monitored early in order to ensure that they can complete the given task on time. This process is particularly important because universities often enroll thousands of students, thereby making individual supervision impractical. An effective solution to this problem is leveraging machine learning to develop a system that predicts whether student will complete the study without delay. Therefore, this study used Support Vector Machine (SVM) method for classification, with RBF kernel. Optimization of SVM classification was achieved by ensuring the values for Soft Margin C parameter and kernel parameter were correct. In addition, Particle Swarm Optimization (PSO) method was used to determine the optimal SVM parameter values. Consequently, the resulting model was evaluated using Cross Fold Validation. The optimized SVM parameter identified through PSO were gamma of 0.0085 and C of 0.4196. The average training accuracy recorded is 82.58%, with 81.22% validation, these results can be categorized into Good Classification. Finally, the application of PSO in optimization resulted in SVM models that avoided overfitting, as shown by the closeness of training and validation values.

Keywords: PSO, SVM, optimization, student completion prediction, data mining

I. INTRODUCTION

Timely completion of a study program is crucial for evaluating the quality of universities. Therefore, to achieve timely completion, student's progress needs to be monitored early in order to ensure that they can complete the given task on time. The process is very important because universities often enroll thousands of students, thereby making individual supervision impractical. One of the applications of technology that can be used in supporting preliminary information about the problem of timely completion of study for student is data mining. Conceptually, data mining is the process of discovering patterns or knowledge relationships in a data

set [1]. In the context of this study, data mining can be used to explore the result pattern of student who graduate on time and exceed the study period. An example of the data mining method that can be used to identify the pattern of student's study periods is classification method [2]. Specifically, classification is the grouping of data into a class based on the characteristics [3], and one of the methods used for this classification is Support Vector Machine (SVM) [4].

SVM is a supervised learning-based machine learning classification method developed by Vapnik [5]. Furthermore, SVM has the advantage of solving problems with small, nonlinear, and high-dimensional data samples [6]. In its application, SVM has a high level of prediction accuracy, and the level of classification accuracy is highly dependent on parameter settings and feature selection [7]. Specifically, the parameters used by SVM are the soft margin C and the kernel. It is crucial to be aware that SVM has several kernels useful for classifying non-linear data, namely, linear, polynomial, sigmoid, and RBF kernel [8], [9]. Based on several comparisons in a related study, the best kernel for data classification is RBF kernel [10]-[12]. The parameters required by RBF kernel are soft margin C and gamma (γ). Moreover, the challenge of obtaining optimal parameter values to classify data in SVM is that the process requires several trials and errors [13], [14]. This implies that the method is not very efficient and consumes a lot of time. Therefore, a method is required to perform parameter tuning on SVM in order to obtain the optimal parameter value in a short time [15]. An efficient solution for obtaining optimal parameter value is to apply optimization method [16].

Several optimization methods can be used to address different scenarios, such as PSO. Typically, PSO is an evolutionary computing method that can obtain the global optimal solution in the search space by considering the interaction between individuals in each particle. Each particle provides its best position

information to others, adjusting its position and rate of change based on the provided information [17]. When compared with similar algorithms such as Genetic Algorithm (GA), PSO can conduct better parameter tuning. The parameter tuning is represented in related studies such as controller design based on Flexible AC Transmission System (FACTS) [18], and analysis for Task Scheduling in Cloud Computing Environment [19]. Additionally, Yoga Religia used the Naive Bayes algorithm to predict airline customer satisfaction, comparing the performance of PSO and GA optimization methods. The results show that the best model was achieved with Naive Bayes algorithm optimized by PSO, producing accuracy, precision, recall, and AUC of 86.13%, 87.90%, 87.29%, and 0.923, respectively [20]. In another study, Irawan compared the performance of K-NN and Random Forest algorithms, both optimized with PSO, for predicting timely study completion. PSO significantly improved accuracy, especially for Random Forest, boosting accuracy from 95.79% to 97.89% and AUC from 0.991 to 0.993. Similarly, K-NN's accuracy increased from 93.49% to 96.74%, and AUC from 0.975 to 0.986 after PSO optimization. [21].

The current study proposes the application of PSO for SVM performance optimization in classifying study completion of student. The exploration aims to predict student's study completion by developing a framework that applies PSO. Additionally, this framework determines the optimal parameter values for classification using SVM method. It is crucial to be aware that the reason for applying these two methods is to increase the accuracy, precision, and AUC values.

II. METHOD

Cross-Industry Standard Process for Data Mining (CRISP-DM) method was used in this study, which included six phases, Fig. 1 showed the stages of CRISP-DM, with the first step being Business Understanding. Specifically, the process was performed by reviewing the existing systems at the study location.

The second stage was the Data Understanding phase, where available data for exploration was thoroughly checked. In the third stage, known as Data preparation, data was processed for use, while in the fourth stage, prediction model was developed using Python programming language, SVM, and PSO. Furthermore, the model was evaluated in the fifth stage, called Evaluation stage. The final phase is Deployment included deploying the developed model followed by conclusions and recommendations. It is crucial to be aware that recommendations were provided based on the exploration result.

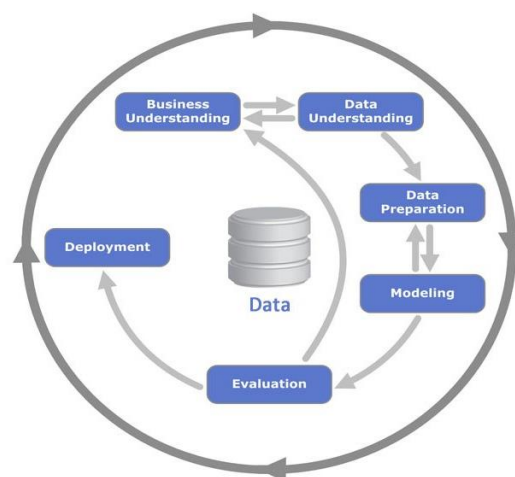


Fig. 1 CRISP-DM

A. Data Preparation

The dataset used in this study consisted of student data from the academic database of Indonesian Institute of Business and Technology Campus. Data was collected by executing SQL queries on various tables in the database, later imported in CSV format for easier processing with Python Programming Language. The dataset includes 4,660 student data from batch 2015 to 2019 with 102 attributes. Although this dataset is comprehensive, a selection process is still needed to identify attributes that are significant in influencing student graduation.

B. Data Pre-processing

Data pre-processing in this situation was needed because the raw data obtained from the database could not be processed by the learning machine due to the presence of corrupted data, hence, the pre-processing required several stages. The first stage was selecting data features or attributes related to the study completion. In selecting the features or data attributes used in modelling, 28 data attributes were obtained, as shown in Table I.

TABLE I
DATA ATTRIBUTE

No	Data Attribute	Data Type
1	Study Programme	Numeric
2	Cumulative GPA 1 - 6	Numeric
3	Grade Point Average 1-6	Numeric
4	Complete Course Credits 1-6	Numeric
5	Incomplete Course Credits 1-6	Numeric
6	Gender	Numeric
7	Region of Origin	Categoric
8	Work Status	Numeric

The second stage involved cleaning data with noise and missing values, adjusting them as required to enhance data usability for the machine learning model. This process was executed using Python and relevant libraries, with examples shown in Table II.

C. Modeling

During the modeling stage, the test data was divided using K-Fold cross-validation to build an effective model. This method splits the data into k equal-sized subsets, using k-1 subsets for training and the remaining one for testing. The process is repeated, with each subset serving as the test set once, ensuring comprehensive evaluation and parameter optimization [22]. Moreover, for classification, the accuracy estimate was the number of correct classification from k iterations divided by the total number of tuples in the initial data [23]. The next stage is modeling using the SVM classification method. In the modeling process, to determine the tuning hyperparameters of SVM, the PSO optimization method is used which the flowchart of the PSO method is shown in Fig. 2.

During the exploration, modeling was conducted using SVM algorithm for data classification, distinguishing between student who completed the study and those who did not. Additionally, PSO optimization method was applied to determine the parameters used in SVM algorithm. Fig. 3 showed an overview of the system created in this result using SVM method with PSO optimization. The prediction system starts with dataset preparation, which is then divided into training and testing sets. The training data is employed for the classification modeling process. The PSO algorithm is used to optimize the c and y parameters for the RBF kernel in the SVM model. The PSO algorithm iteratively updates the particles to find the best solution (Pbest) and the global best solution (Gbest). After determining the optimal parameters, the SVM model is trained, and its performance is evaluated using the confusion matrix. This process is repeated until a good classification accuracy value is obtained. After achieving good accuracy, the optimized model is then finalized for the student graduation prediction system.

TABLE II
DATA PREPROCESSING

No	Name	Study Program	Cummulative GPA 1	Cummulative GPA 2	...	Work Status
1	Student 1	IT	3.3	3.66	...	Employment
2	Student 2	IT	3.15	2.92	...	<u>Unemployment</u>
...
4660	Student 4660	IT	3.22	2.75	...	<u>Unemployment</u>

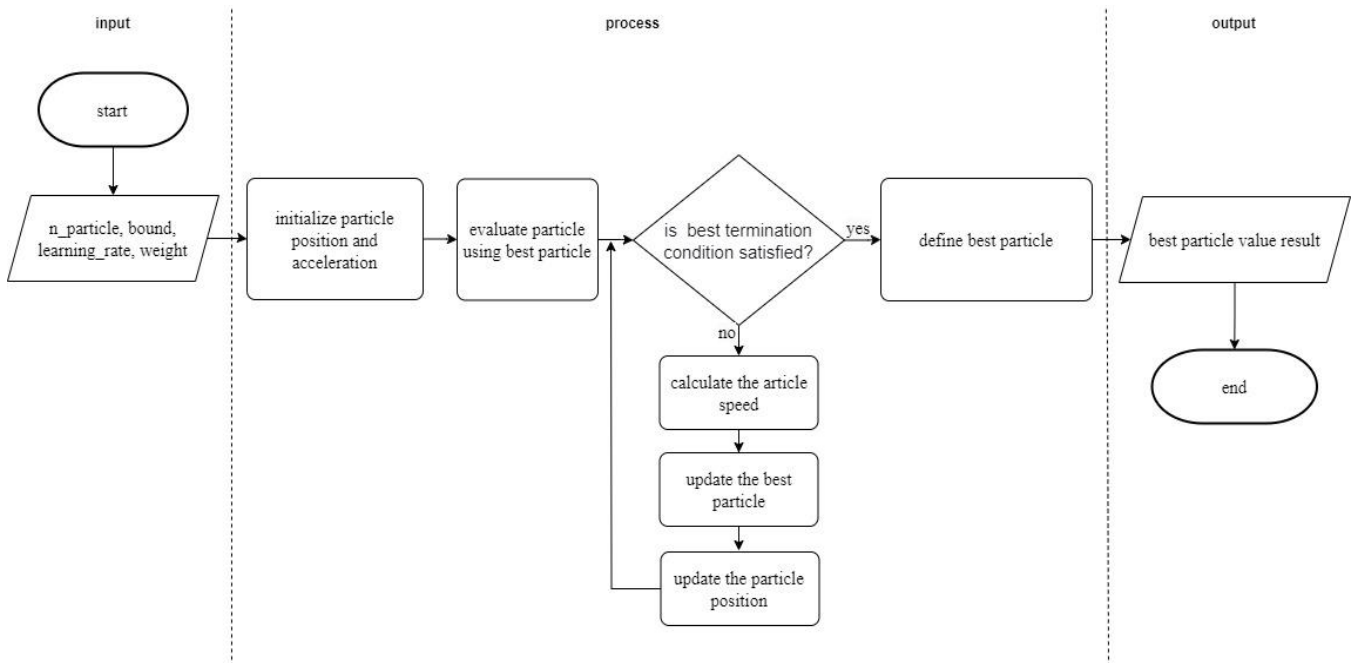


Fig. 2 Flowchart PSO

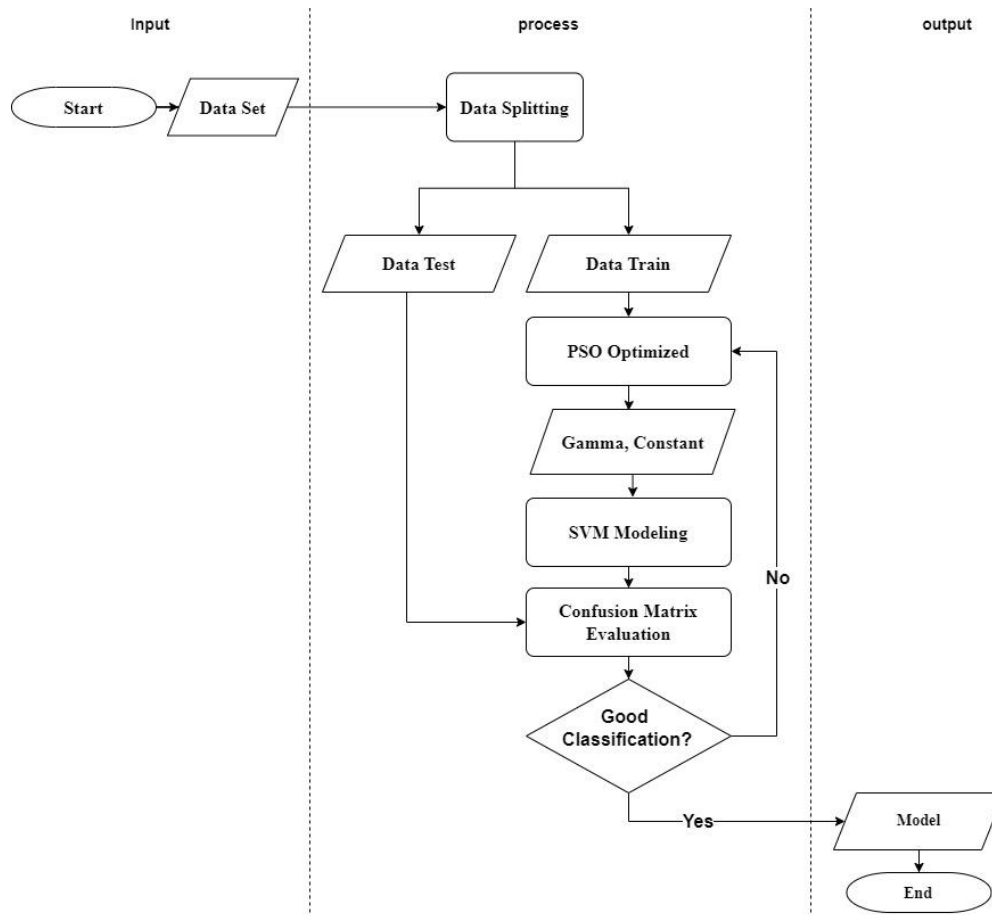


Fig. 3 Flowchart SVM with PSO

D. Evaluation

After modeling, the next stage included evaluating how well the resulting model performed. When assessing the machine learning model, it was important to consider accuracy and precision. Furthermore, accuracy was the ratio of correct prediction to every possible prediction. Precision metric measured the ratio of actual positive predictions to the number of actual positives in a dataset [24]. The evaluation value of this model was conducted through K-fold validation, which also helped determine when the model was overfitting, best fitting, or underfitting [25]. During evaluation, the model's performance was assessed using AUC values and ROC curves. The ROC curve plots True Positive Rate (TPR) against False Positive Rate (FPR) on a two-dimensional graph. Key points include (0,0), indicating no false positives and only true positives, and (1,1), indicating a positive condition. The TPR and FPR values are inversely related, meaning as TPR increases, FPR decreases. Random classification results in a diagonal line on the ROC curve, known as Random Performance [26]. AUC was an area that showed the accuracy of

prediction model and an area of a square, with its value always between 0 and 1 [27]. The values obtained when assessing the performance of the model using AUC were divided into several categories, as shown in Table III.

III. RESULT AND DISCUSSION

A. Data Preprocessing

After data cleaning, including handling missing values and converting categorical data to numeric, the data was prepared for modeling. The training and testing data were separated, with 3,224 lines from the 2015–2018 class used for testing and 1,436 from the 2019 class.

TABLE III
AUC VALUE CATEGORY

Grade AUC	Category
0.90 – 1.00	Excellent Classification
0.80 – 0.90	Good Classification
0.70 – 0.80	Fair Classification
0.60 – 0.70	Poor Classification
0.50 – 0.60	Failure

Training data was labelled: Label 1 for students who graduated in 3.5 to 4 years, and Label 0 for those taking longer than 4 years. Examples of the processed data are shown in Table IV.

B. Modeling

The application of modeling to classify student data using SVM algorithm was compared with the results of PSO parameter optimization. The study used support from the Python programming language with its open-source library. When classifying data, SVM had to determine the exact kernel to use. In this study, RBF kernel was selected and implemented with parameters C and Gamma. This kernel was selected because it could classify data very well [28], [11]. Following the selection in determining SVM parameters using PSO method, several efforts were needed to determine the parameters using PSO to find the optimal SVM parameter values. In addition, several experimental parameters were used for PSO as shown in Table V. C1, C2, W, Number of particles, Number of iterations, and Fold were the initial parameters for performing PSO. The Gamma and C columns represented the optimal values resulting from PSO optimization (Table V).

After testing several parameters in Table V, the optimal SVM parameters were found to be Gamma 0.0085 and C 0.4196. The tuning of PSO parameters for

SVM was achieved through an experiment with 100 iterations, 100 particles, 10 K-folds, c1 0.3, c2 0.5, and w 0.9. Following this, a comparison of modeling results was conducted, including the evaluation of SVM with the determined parameters, SVM optimized by PSO, and the model's accuracy and precision values.

1) *Accuracy*: Accuracy described how accurately the model could classify correctly. To determine the accuracy value, the cross-fold validation value was determined 10 times by comparing the accuracy produced by SVM model with the parameter values obtained, where C = 1 and Gamma = 1. These C and Gamma value was considered good and received a good accuracy value in the study [11]. The results of SVM with parameters optimized by PSO were also presented in Table VI.

Table VII compares the average training and validation accuracy of the SVM model with randomly determined parameters against those optimized using PSO. The SVM model with specified parameters had a high training accuracy of 99.75 but a low validation accuracy of 66.63, indicating overfitting [29] [30]. In contrast, the PSO-optimized SVM showed more balanced results, with training and validation accuracies of 82.58 and 81.22, respectively, indicating a best fitting as seen on Fig. 5.

TABLE IV
DATA AFTER THE PREPROCESSING

No	Study Programme	Cumulative GPA 1	Cumulative GPA 2	Employment Status	Label
1	1	3.3	3.66	...	1	1
2	1	3.15	2.92	...	0	1
...
4660	1	3.22	2.75	...	0	-

TABLE V
PSO PARAMETER TUNING RESULTS

C1	C2	W	Number of Particles	Iteration	Fold	Gamma	C	Accuracy
0,3	0,5	0,9	100	10	1	0.0378	1.4694	81.54%
0,3	0,5	0,9	100	10	2	0.0543	3.3972	80.93%
0,3	0,5	0,9	100	10	3	0.0549	0.2947	81.24%
0,3	0,5	0,9	100	10	4	0.1320	-3.0954	84.62%
0,3	0,5	0,9	100	10	5	0.0161	0.8359	78.16%
0,3	0,5	0,9	100	10	6	0.0085	0.4196	86.16%
0,3	0,5	0,9	100	10	7	0.0501	0.6860	79.33%
0,3	0,5	0,9	100	10	8	0.0671	0.7884	81.18%
0,3	0,5	0,9	100	10	9	0.0041	3.3651	83.03%
0,3	0,5	0,9	100	10	10	0.0734	4.9142	80.25%

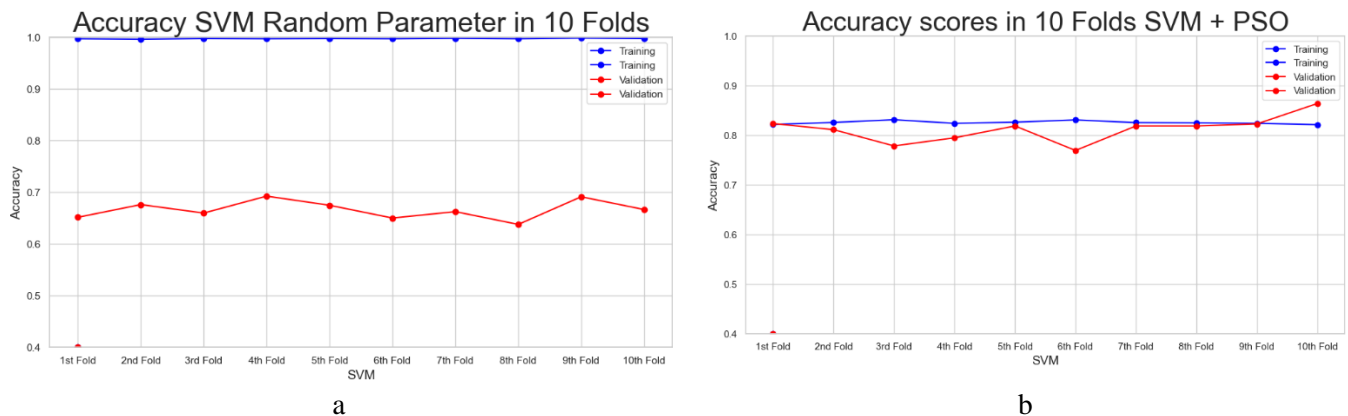


Fig. 5 Accuracy comparison (a) SVM random parameter model (b) SVM PSO optimization

Fig.5 compares the SVM model with random parameters (Fig. 5(a)) to the PSO-optimized SVM model (Fig. 5(b)). The graph in Fig. 5(a) shows a significant gap between the training and validation lines, indicating a poor fit. In contrast, Fig. 5(b) shows that the PSO-optimized SVM model has much closer training and validation lines, suggesting a better-fitting model.

2) *Precision*: Precision could be described as the level of accuracy between the requested data and prediction results provided by the model. The results of comparing the precision values of the random SVM parameters with SVM parameters resulting from PSO optimization were shown in Table VII.

The average precision value of the training process was 0.997, while the validation value was 0.667, indicating an overfitting model. In contrast, the PSO-optimized SVM model achieved a training precision of 0.802 and a validation precision of 0.793, suggesting a better fit. Figure 6 illustrates the precision values for both the randomly parameterized SVM model (Fig. 6(a)) and the PSO-optimized model (Fig. 6(b)).

The graph (Fig. 6(a)) showed the training and validation lines were very different, compared to the graph (Fig. 6(b)) which had the training and validation lines close to each other. The comparison concluded that the graph (Fig. 6(a)) generated an overfitting model, and the graph (Fig. 6(b)) produced a best-fitting model.

**TABLE VI
COMPARISON OF ACCURACY VALUE**

K-Fold	Accuracy Training		Accuracy Validation	
	SVM	SVM + PSO	SVM	SVM + PSO
K-1	0.997	0.821	0.651	0.823
K-2	0.996	0.826	0.676	0.811
K-3	0.997	0.831	0.659	0.778
K-4	0.997	0.824	0.692	0.795
K-5	0.997	0.826	0.674	0.818
K-6	0.997	0.831	0.650	0.769
K-7	0.998	0.825	0.662	0.818
K-8	0.997	0.825	0.637	0.818
K-9	0.998	0.824	0.691	0.823
K-10	0.997	0.821	0.666	0.864
Average Accuracy	99.75	82.58	66.63	81.22

**TABLE VII
COMPARISON OF PRECISION VALUES**

K-Fold	Precision Training		Precision Validation	
	SVM	SVM + PSO	SVM	SVM + PSO
K-1	0.997	0.799	0.657	0.788
K-2	0.995	0.805	0.669	0.788
K-3	0.997	0.805	0.662	0.764
K-4	0.997	0.800	0.680	0.777
K-5	0.997	0.804	0.679	0.790
K-6	0.997	0.806	0.656	0.759
K-7	0.997	0.801	0.667	0.817
K-8	0.997	0.800	0.650	0.806
K-9	0.998	0.801	0.682	0.796
K-10	0.997	0.799	0.666	0.841
Average Accuracy	0.997	0.802	0.667	0.793

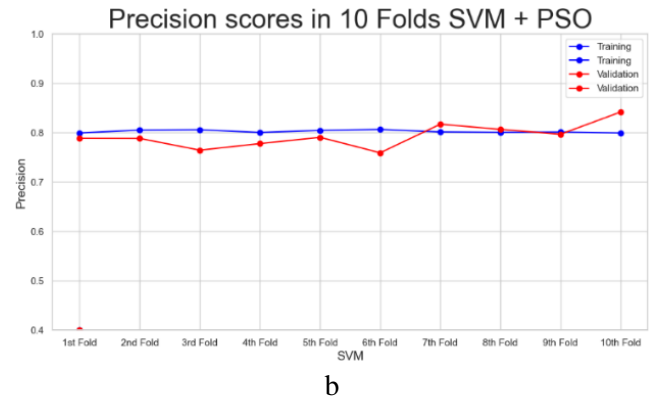
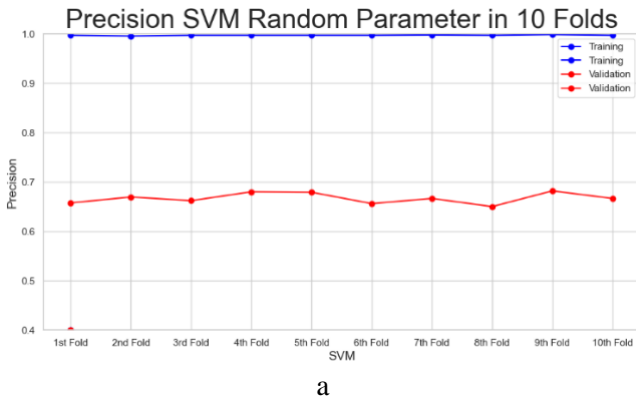


Fig. 6 Comparison of the precision of the random parameter SVM model with PSO optimization SVM

3) *AUC ROC Curve*: ROC graph was a two-dimensional graph of the relationship between TPR or Sensitivity (Y-axis) and FPR or 1-Specificity (X-axis). The graph showed the level of probability or accuracy of the model. The ROC value was concluded to be good (Perfect Classification) when the ROC curve approached the value 1 at the top left point [1]. Meanwhile, the AUC value was an area that showed the level of accuracy of prediction model [27].

Fig.7 compares the ROC curves and AUC values of the SVM model with random parameters (Fig. 7(a)) and those optimized by PSO (Fig. 7(b)). The random SVM model had an AUC of 0.748, indicating fair classification, while the PSO-optimized model achieved an AUC of 0.824, indicating best classification. This shows that the PSO-optimized SVM model outperformed the one with random parameters. After developing the optimal model, it was deployed to a web platform using the Flask framework, as shown in Fig . 8.

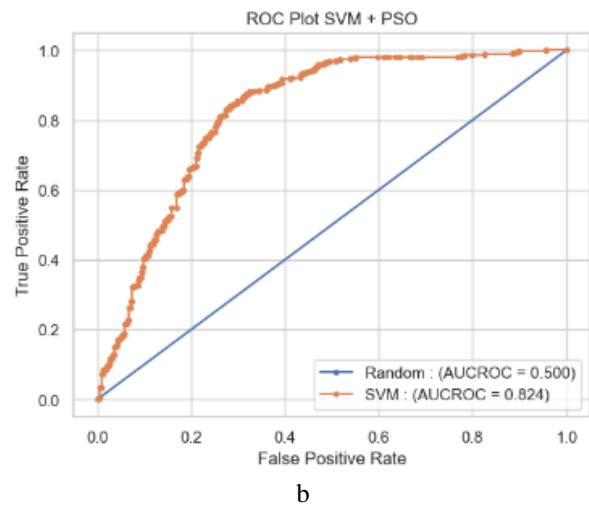
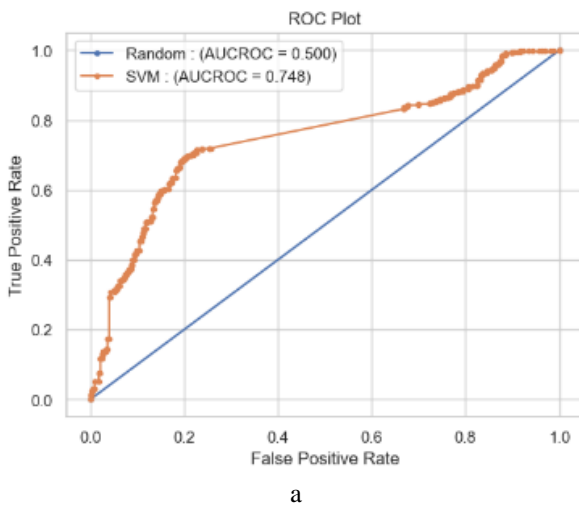


Fig. 7 Comparison of the AUC ROC value of SVM random parameter model with PSO optimization SVM

The screenshot shows a web application titled "Student Completion Prediction System" with the INSTIKI logo. It features a "Choose File" button, a "Predict" button, and a table of student data. The table has columns for Student ID, Name, Major, GPA 1, GPA 2, GPA 3, GPA 4, and Result. The results are categorized as "Not On Time" or "On Time".

Student ID	Name	Major	GPA 1	GPA 2	GPA 3	GPA 4	Result
19102109	Abdul Kharisman	SK	0.0	0.0	0.0	0.0	Not On Time
19102164	Adi Sanjaya	SK	2.06	2.75	2.94	2.93	Not On Time
19102047	Afuwa Ghofur	SK	3.6	3.74	3.67	3.67	On Time
19102154	Ahmad Khaerozi	SK	3.5	3.3	3.15	3.18	On Time
19104212	Anak Agung Istri Yustisia	TI	3.27	3.26	3.32	3.28	On Time
19104156	Anak Agung Made Arya	TI	3	0.0	0.0	0.0	Not On Time

Fig. 8 Deployment model results

Fig. 8 showed the results of the deployed model saved and tested using the test data created in the previous data preparation stage. Consequently, this process allowed the model to classify existing data and predict whether student would complete the study without delay. The web application facilitated data-driven decision-making to increase student’s retention and graduation rates.

IV. CONCLUSION

In conclusion, timely program completion is crucial for assessing university quality. To ensure students finish on time, early progress monitoring is essential, especially given the challenge of supervising large student populations. Machine learning, specifically data classification using SVM with PSO optimization, offers a solution by predicting timely completion. This study applied SVM for classification and PSO for parameter optimization, achieving a well-fitting model with good accuracy. Evaluation through cross-fold validation yielded training and validation accuracy of 85.32% and 83.16%, respectively, with an AUC value of 0.851, indicating good classification. The study also highlighted that SVM models with predetermined parameters led to overfitting, as evidenced by significant discrepancies between training and validation metrics.

REFERENCES

[1] O. Maimon and L. Rokach, *Data mining and knowledge discovery handbook*, vol. 48, no. 10. 2011.
 [2] M. S. Zulfiker, N. Kabir, A. A. Biswas, P. Chakraborty,

and M. M. Rahman, “Predicting students’ performance of the private universities of Bangladesh using machine learning approaches,” *Int. J. Adv. Comput. Sci. Appl.*, vol. 11, no. 3, pp. 672–679, 2020, doi: 10.14569/ijacsa.2020.0110383.
 [3] I. Romli and A. T. Zy, “Penentuan Jadwal Overtime Dengan Klasifikasi Data Karyawan Menggunakan Algoritma C4.5,” *J. Sains Komput. Inform. (J-SAKTI)*, vol. 4, no. 2, pp. 694–702, 2020.
 [4] N. Priyasadie and S. M. Isa, “Educational Data Mining in Predicting Student Final Grades on Standardized Indonesia Data Pokok Pendidikan Data Set,” *Int. J. Adv. Comput. Sci. Appl.*, vol. 12, no. 12, pp. 212–216, 2021, doi: 10.14569/IJACSA.2021.0121227.
 [5] M. Y. Cho and T. T. Hoang, “Feature Selection and Parameters Optimization of SVM Using Particle Swarm Optimization for Fault Classification in Power Distribution Systems,” *Comput. Intell. Neurosci.*, vol. 2017, 2017, doi: 10.1155/2017/4135465.
 [6] X. Yao, “Application of Optimized SVM in Sample Classification,” *Int. J. Adv. Comput. Sci. Appl.*, vol. 13, no. 6, pp. 540–547, 2022, doi: 10.14569/IJACSA.2022.0130666.
 [7] I. Hidayat, “Prediksi kelulusan tepat waktu menggunakan algoritma svm dan k-nearest neighbour berbasis particle swarm optimization di STIMK Eresha,” *J. ESIT (E-Bisnis, Sist. Inf., Teknol. Informasi)*, vol. 16, no. 01, pp. 55–65, 2021.
 [8] Z. Zhao, Yuying Song, F. Cui, J. Zhu, Z. Xu., “Point Cloud Features-Based Kernel SVM for Human-Vehicle Classification in Millimeter Wave Radar,” *IEEE Access*, vol. 8, pp. 26012–26021, 2020, doi:

- 10.1109/ACCESS.2020.2970533.
- [9] G. Lantzanakis, Z. Mitraika, and N. Chrysoulakis, "X-SVM: An Extension of C-SVM Algorithm for Classification of High-Resolution Satellite Imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 5, pp. 3805–3815, 2021, doi: 10.1109/TGRS.2020.3017937.
- [10] M. Azimi-Pour, H. Eskandari-Naddaf, and A. Pakzad, "Linear and non-linear SVM prediction for fresh properties and compressive strength of high volume fly ash self-compacting concrete," *Constr. Build. Mater.*, vol. 230, p. 117021, 2020, doi: 10.1016/j.conbuildmat.2019.117021.
- [11] H. Al Azies, D. Trishnanti, and E. Mustikawati P.H, "Comparison of Kernel Support Vector Machine (SVM) in Classification of Human Development Index (HDI)," *IPTEK J. Proc. Ser.*, vol. 0, no. 6, p. 53, 2019, doi: 10.12962/j23546026.y2019i6.6339.
- [12] A. P. Gopi, R. N. S. Jyothi, V. L. Narayana, and K. S. Sandeep, "Classification of tweets data based on polarity using improved RBF kernel of SVM," *Int. J. Inf. Technol.*, 2020, doi: 10.1007/s41870-019-00409-4.
- [13] S. Yenaeng, S. Saelee, and W. Samai, "Automatic Medical Case Study Essay Scoring by Support Vector Machine and Genetic Algorithms," *Int. J. Inf. Educ. Technol.*, vol. 4, no. 2, pp. 132–137, 2014, doi: 10.7763/ijiet.2014.v4.384.
- [14] F. Budiman, "SVM-RBF parameters testing optimization using cross validation and grid search to improve multiclass classification," *Sci. Vis.*, vol. 11, no. 1, pp. 80–90, 2019, doi: 10.26583/sv.11.1.07.
- [15] N. A. Al-Thanoon, O. S. Qasim, and Z. Y. Algamal, "A new hybrid firefly algorithm and particle swarm optimization for tuning parameter estimation in penalized support vector machine with application in chemometrics," *Chemom. Intell. Lab. Syst.*, vol. 184, no. December 2018, pp. 142–152, 2019, doi: 10.1016/j.chemolab.2018.12.003.
- [16] N. B. Korade and M. Zuber, "Stock Price Forecasting using Convolutional Neural Networks and Optimization Techniques," *Int. J. Adv. Comput. Sci. Appl.*, vol. 13, no. 11, pp. 378–385, 2022, doi: 10.14569/IJACSA.2022.0131142.
- [17] R. Eberhart and J. Kennedy, "New optimizer using particle swarm theory," *Proc. Int. Symp. Micro Mach. Hum. Sci.*, pp. 39–43, 1995, doi: 10.1109/mhs.1995.494215.
- [18] S. Panda and N. P. Padhy, "Comparison of particle swarm optimization and genetic algorithm for FACTS-based controller design," *Appl. Soft Comput. J.*, vol. 8, no. 4, pp. 1418–1427, 2008, doi: 10.1016/j.asoc.2007.10.009.
- [19] F. Nzanywayingoma and Y. Yang, "Analysis of Particle Swarm Optimization and Genetic Algorithm based on Task Scheduling in Cloud Computing Environment," *Int. J. Adv. Comput. Sci. Appl.*, vol. 8, no. 1, pp. 19–25, 2017, doi: 10.14569/ijacsa.2017.080104.
- [20] Yoga Religia and A. Amali, "Perbandingan Optimasi Feature Selection pada Naïve Bayes untuk Klasifikasi Kepuasan Airline Passenger," *J. RESTI (Rekayasa Sist. dan Teknol. Informatika)*, vol. 5, no. 3, pp. 527–533, 2021, doi: 10.29207/resti.v5i3.3086.
- [21] I. Irawan, M. R. Qisthiano, M. Syahril, and P. M. Jakak, "Optimasi Prediksi Kelulusan Tepat Waktu: Studi Perbandingan Algoritma Random Forest dan Algoritma K-NN Berbasis PSO," *J. Pengemb. Sist. Inf. dan Inform.*, vol. 4, no. 4, pp. 26–35, 2023, doi: 10.47747/jpsii.v4i4.1374.
- [22] S. Santoso and R. Nurmalina, "Perencanaan dan Pengembangan Aplikasi Absensi Mahasiswa Menggunakan Smart Card Guna Pengembangan Kampus Cerdas," *J. Integr.*, vol. 9, no. 1, p. 84, 2017, doi: 10.30871/ji.v9i1.288.
- [23] J. Han, M. Kamber, and J. Pei, *Data mining: Data mining concepts and techniques*. 2011.
- [24] O. A. Alrusaini, "Deep Learning Models for the Detection of Monkeypox Skin Lesion on Digital Skin Images," *Int. J. Adv. Comput. Sci. Appl.*, vol. 14, no. 1, pp. 637–644, 2023, doi: 10.14569/IJACSA.2023.0140170.
- [25] D. Berrar, "Cross-validation," *Encycl. Bioinforma. Comput. Biol. ABC Bioinforma.*, vol. 1–3, pp. 542–545, 2018, doi: 10.1016/B978-0-12-809633-8.20349-X.
- [26] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognit. Lett.*, vol. 27, no. 8, pp. 861–874, 2006, doi: 10.1016/j.patrec.2005.10.010.
- [27] P. F. Gorunescu, *Data Mining Concepts, Models and Techniques*, vol. 59. 2011.
- [28] E. S. Putri and M. Sadikin, "Prediksi Penjualan Produk Untuk Mengestimasi Kebutuhan Bahan Baku Menggunakan Perbandingan Algoritma LSTM dan ARIMA," *Format J. Ilm. Tek. Inform.*, vol. 10, no. 2, p. 162, 2021, doi: 10.22441/format.2021.v10.i2.007.
- [29] S. Pothuganti, "Review On Over-fitting and Under-fitting Problems In Machine Learning and Solutions," *Int. J. Adv. Res. Electr. Electron. Instrum. Eng.*, vol. 7, no. 9, pp. 3692–3695, 2018, doi: 10.15662/IJAREEIE.2018.0709015.
- [30] A. Montesinos López, Osva Antonio and J. Crossa, *Multivariate Statistical Machine Learning Methods for Genomic Prediction*. 2022.

