HYBRID OPTIMIZATION METHOD BASED ON GENETIC ALGORITHM FOR GRADUATES STUDENTS

Ridwansyah[1]; Ganda Wijaya[2]; Jajang Jaya Purnama[3]

[1] Informatics Engineering Study Program, [2] Information Systems Study Program
STMIK Nusa Mandiri, Jakarta, Indonesia
www.nusamandiri.ac.id
[1]rdwansyah@gmail.com , [2]ganda.gws@nusamandiri.ac.id

[3] Computer Technology Study Program
Universitas Bina Sarana Informatika, Jakarta, Indonesia
www.ubsi.ac.id
[3]jajangja2412@bsi.ac.id

Abstract— Graduation is a target that must be achieved by students, especially graduating on time will be very important. To determine students who graduate on time or cannot be determined before students reach the final semester and hold a trial, many students who fail to graduate on time cause delays and affect the quality assurance of a tertiary institution. The problem in this research is how to optimize student graduation in order to graduate on time. Therefore, to determine this decision, we conducted a graduation data trial using the SVM method with GA optimization. SVM with accurate learning skills and good generalizations in classifying non-linear data, but SVM is weak in terms of parameter optimization it requires optimization using GA. GA is a method that has evolved to produce a more optimal data. From the results of processing using SVM and GA, we get more optimal results with 86.57%. Then from these results can help students to graduate on time.

Keywords: Genetic Algorithm, Student Graduation, Hybrid Optimization

***Abstrak*—** *Kelulusan merupakan suatu target yang harus dicapai oleh mahasiswa terutama lulus dengan tepat waktu akan menjadi hal yang sangat penting. Untuk menentukan mahasiswa lulus tepat waktu atau tidak dapat ditentukan sebelum mahasiswa mencapai semester akhir dan melakukan sidang, banyak mahasiswa yang tidak lulus tepat waktu mengakibatkan terhambatnya dan berpengaruh terhadap jaminan mutu sebuah perguruan tinggi. Masalah dalam penelitian ini adalah bagaimanan mengoptimalkan kelulusan mahasiswa agar dapat lulus tepat waktu. Oleh karena itu untuk menentukan keputusan tersebut kami melakukan uji coba data kelulusan dengan metode SVM dengan optimasi GA. SVM dengan kemampuan belajar yang akurat dan generalisasi yang baik dalam pengklasifisian data non-liner, tetapi SVM lemah dalam hal optimasi parameter maka diperlukan optimasi menggunakan GA. GA merupakan suatu metode yang berevolusi untuk menghasilkan suatu data yang lebih optimal. Dari hasil pengolahan menggunakan SVM dan GA maka didapatkan hasil yang lebih optimal dengan 86,57%. Maka dari hasil tersebut dapat membantu mahasiswa untuk lulus tepat waktu*

Kata Kunci*: Algoritma Genetika; Optimasi Hibrida; Kelulusan Mahasiswa.*

INTRODUCTION

Graduation is a very important part for every student who is studying and one of the main goals for students who are pursuing education is including students who take diploma education. Graduation can be achieved by each student who has agreed to meet all graduation requirements. There are several requirements for students and this can be different for each level of Education and Higher Education. Graduation is not only for the benefit of students, but also for the benefit of lecturers, and for guardians and also universities in general. It is very important for students so that it is very necessary to be well considered and approved as early as possible by all parties who have an interest. The most basic thing that can be done is to implement and apply policies related to student graduation. Policies related to student graduation will be very beneficial for all parties concerned. With the presence of the application, the implementation and prediction of students carried out optimally are expected to provide a positive for all parties related to student graduation. Then we need a method for predicting or classifying student data by selecting different and organizing different data, and consequently, the process of evaluating and making decisions becomes important.

The problem raised was that many students could not graduate on time due to various factors which were an obstacle for a tertiary institution. Students who graduate on time are very helpful in terms of accreditation and quality assurance in a tertiary institution. Some methods for analyzing data that have been carried out by researchers include using Decision Tree (DT), Naive Bayes (NB), Neural Network (NN) and Support Vector Machine (SVM). Among these methods have been carried out to check the data of postgraduate students, namely SVM, DT and NN, the best results have been obtained using SVM (Riyanto, Hamid, and Ridwansyah 2019).

SVM can minimize risk by inserting vectors into higher dimensional spaces (Bin and Min 2012), a subset of the selected features and kernel types, resulting in good detection performance for highly unbalanced datasets (Wang, Yang, and Li 2017), and is a method that can overcome the problem of classification of small samples (Liu et al. 2019) and find a pretty good solution, with statistical learning theory and data sets can be analyzed all types of objects (Yu and Hung 2017). but it is very difficult to choose the appropriate SVM parameters. SVM is still limited because it has to set parameters in advance and can cause less fitting and overfitting (Ye et al. 2017), An algorithm is needed to find a solution to the difficulty of choosing SVM parameters (Li and Zhang 2009).

Feature selection is used to select the optimal subset of the original subset. With feature selection, excessive and irrelevant or unimportant data is deleted (Jiang, Tang, and Zhang 2010). Genetic algorithm is more suitable for choosing the right SVM parameters because SVM is very difficult to choose the right parameters. GA can help classify and can avoid excessive local optimization and is often used to look for problems (Yu and Hung 2017), The process of searching for genetic algorithms is iterative: evaluating, selecting, and rejoining strings in the population during each iteration until they reach several termination conditions (Gao, Yang, and Hu 2010), the purpose of using GA is to improve the quality and accuracy of modeling. With the GA-SVM method, students are expected to achieve graduation on time.

The aim of this study is to make students able to achieve graduation on time, and for universities as one that determines higher education accreditation and can optimize weak SVM methods in terms of parameter optimization. time.

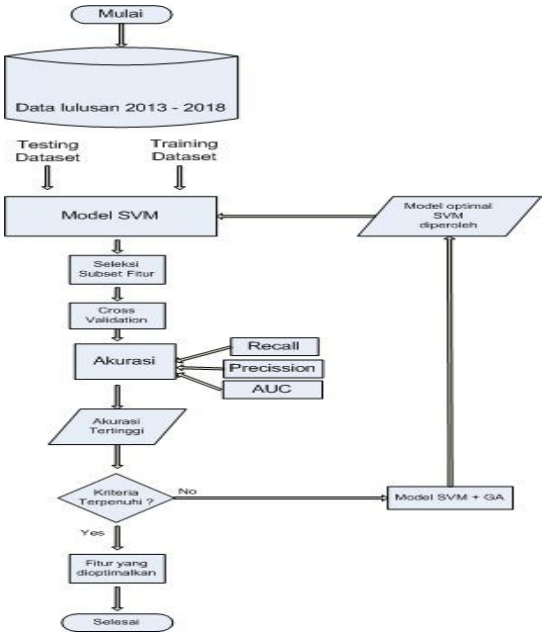
MATERIALS AND METHODS

Previous studies using GPA factors have been analyzed and insignificant results were obtained from student graduation (Freitas and Leonard 2011). Research on student graduation has been carried out with several neural networks (NN), decision trees (DT) and naïve Bayes (NB) methods and produces the highest accuracy, NN (Ashok and Apoorva 2016), even other studies have been carried out using these three methods and adding linear regression to these results. NB Get the best accuracy (Devasia, P, and Hegde 2008). Penelitian yang telah dilakukan dengan metode NN, DT dan SVM pada kumpulan data dari universitas dengan menghasilkan SVM sebagai akurasi tertinggi (Riyanto, Hamid, and Ridwansyah 2019). From the highest evaluation results SVM is optimized using PSO (Particle Swarm Optimization) from the research results that PSO can improve the SVM model (Suhardjono, Wijaya, and Hamid 2019). From research related to student graduation, it will be tested using hybrid optimization using a support vector engine based on genetic algorithms.

The research that was raised was experimental research, the experiment tested the truth in the hypothesis raised from statistics and related to the research problem. Experimental research is divided into two parts. First the absolute and comparative experiment, absolute experimental research has an impact on experimental results, comparative experiential comparing two different objects.

The research that was raised was comparative experimental research. The model used designs the classification of GA hybrid optimization algorithms with Support Vector Machine (SVM). To find out the best model designed between GA hybrid optimization on SVM and without GA hybrid optimization on SVM. This study uses a dataset from one of the private universities in Indonesia.

Data collection techniques use primary data models, where primary data is collected from direct sources rather than second parties. And the dataset was taken from one of the private universities in Indonesia through interviews and observations. While the study of literature or supporting data obtained from books, journals, and other publications. We briefly explain the SVM and GA methods which can be seen in Figure 1 explaining that the diagram of the hybrid optimization activity uses a genetic algorithm model with a support vector machine. Existing parameters greatly affect the accuracy of the SVM classification. GA is used to find better parameter combinations in SVM train classifiers in larger training sets.



Source: (Ridwansyah, Wijaya, and Purnama 2020)

Figure 1. Activity Diagram of The Optimation Hybrid of Genetic Algorithm and Support Vector Machine

The existing dataset is trained by the classifier on a larger training set based on a subset of the selected features and the type of kernel. If all classifiers are complete, they are tested using a support vector machine algorithm and the selection of feature subsets is chosen to get the best data. After that, it is validated using cross validation and measuring validation accuracy in testing data sets through the resulting model. We use accuracy, the higher the level of accuracy, the greater the results achieved. After achieving the highest accuracy, choosing a subset of characteristic features, namely to achieve the use of a small number of the same or better classification results, and must consider the number of characteristics, giving the same accuracy for two characteristics of a subset, which has a lower number of characteristics will be higher. And we observe the validation accuracy curve and stop training while getting the best validation accuracy during the training process. After getting the highest accuracy and observing the accuracy curve we also take into account the results of recall and precission. If the results of the criteria are not met then do the optimization using a genetic algorithm after the SVM-GA hybrid is done to get the maximum optimization results, so as to get the optimized features.

RESULTS AND DISCUSSION

The proposed model can be seen in Figure 1. This model consists of 1) graduation student data sets, 2) feature selection, 3) test data with support vector machine models, 4) model validation tests, 5) model evaluations and, 6) models comparison with genetic algorithms, 7) resulting from models with optimized features. With the platform that we use with Intel® Core ™ i7-8565U CPU @ 1.80 GHz 1.99 GHz, 8.00 GB RAM for software in data processing using Rapid Miner 9.3.

Student graduation data used in this study were collected from one of the private tertiary institutions in Indonesia. Data from 2013 to 2018 with 796 parameters. And the parameters that we use Jk (Gender), Department of Middle School, High School and GPA from semester 1 to semester 6 can be seen in Table 1.

Table 1. Student Sample Data for 2013-2018

JK	JRS_SLTA	ASAL_SLTA	IPK1	IPK2	IPK3	IPK4	IPK5	IPK6	ON TIME
P	IPA	SMA N 1 KEDUNGREJA	3,55	3	3,3	3,17	3,19	3,2	YES
P	ADMINISTRASI PERKANTORAN	SMK PGRI 1 TANGERANG	2,91	2,83	2,98	3,15	3,16	3,22	YES
P	IPA	SMA DAAN MOGOT TANGERANG	3,27	3,1	3,13	3,14	3,09	3,18	NO
L	REKAYASA PERANGKAT LUNAK (RPL)	SMKN 1 PADAHERANG	3,36	3,34	3,34	3,31	3,28	3,27	YES
P	IPS	SMA 15 TANGERANG	2,73	2,61	2,51	2,49	2,68	2,84	NO
L	MULTIMEDIA	SMK BHAKTI ANINDYA	2,36	2,71	2,7	2,83	2,96	2,97	NO
P	ADMINISTRASI PERKANTORAN	SMK AL-IJTIHAD	3,14	3,24	3,28	3,25	3,16	3,12	YES
L	IPS	SMA 4	3,41	3,34	3,31	3,4	3,45	3,49	YES
L	MULTIMEDIA	SMK N 1 KAB.TANGERANG	2,64	2,56	2,54	2,59	2,66	3,04	NO
P	ADMINISTRASI PERKANTORAN	SMK AL-HIKMAH CURUG	3,23	2,9	2,89	2,59	2,7	2,75	NO
L	IPS	YADIKA 1 TANJUNG DUREN	2,91	2,95	2,72	2,8	3,02	3,05	YES
L	MULTI MEDIA	SMK N 6 TANGERANG	2,27	2,39	2,54	2,63	2,68	2,96	NO
P	TEKNIK KOMPUTER DAN JARINGAN	SMK NEGERI 1 SLAWI	3,55	3,51	3,51	3,57	3,6	3,61	YES
P	ADMINISTRASI PERKANTORAN	NUSANTARA 1	3,23	3,34	3,08	3,17	3,02	3,04	YES
L	OTOMOTIF	SMK PGRI 2	3,09	3,07	3,02	3,07	3,14	3,28	YES
L	IPS	SMAN 1 LARANGAN	2,32	2,41	2,49	2,52	2,91	2,99	YES
L	TEKNIK MESIN	SMK BINAWIYATA	3,09	2,98	2,84	3,01	3,11	3,13	YES
L	TEKNIK PEMELIHARAAN MEKANIK INDUSTRI	SMK YUPPENTEK 1	1,73	1,51	1,34	1,19	1,98	2,47	NO
L	TEKNIK PERMESINAN	SMK AR-RAHMAH	2,36	2,22	2,52	2,79	2,82	2,93	YES

Source: (Ridwansyah, Wijaya, and Purnama 2020)

The data contained in table 1 is data that has been through the cleanup stage of cleaning incomplete data and data that has no value or is missing. This process can do the processing quickly at the time of data classification and the process is done using the SVM method.

1. Support Vector Machine
   After selecting a feature subset, cross validation is performed using the SVM method. From these results obtained the confusion matrix listed in table 2

Table 2. SVM Matrix Confusion Results

Information	On time (Yes)	On time (No)	Class Precission
Prediction Yes	615	94	86.74%
Prediction No	19	68	78.16%
Class Recall	97.00%	41.98%	

Source: (Ridwansyah, Wijaya, and Purnama 2020)

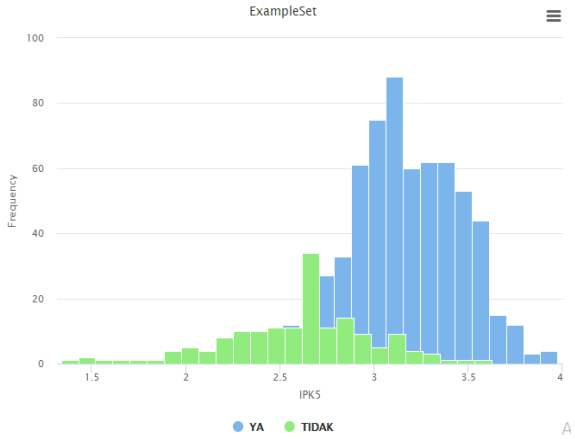
From the confusion matrix svm results can be calculated with the formula confusion matrix as follows.

$$\text{Accuracy} = \frac{(TN+TP)}{(TN+FN+TP+FP)} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (1)$$

$$\text{Accuracy} = \frac{(615+68)}{(615+19+68+94)}$$

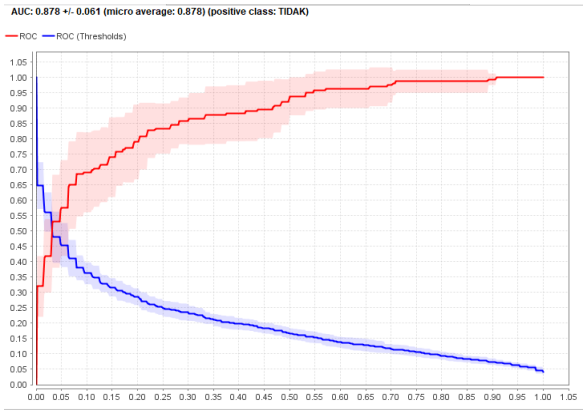
$$\text{Accuracy} = \frac{(683)}{(796)} = 0{,}85804$$

From the calculation formula, accuracy has obtained a value of 0.85804 or 85.84% with the rule that the prediction of timely graduation is 94 students but the SVM results are predicted not to be on time. With the diagram that can be seen in Figure 2.



Sumber: (Ridwansyah, Wijaya, and Purnama 2020)

Figure 2 Student Graduation Diagram with SVM Method



Source: (Ridwansyah, Wijaya, and Purnama 2020)

Figure 3 *Area Under Curve* (AUC) SVM

From Figure 3, the AUC of 0.877 shows that the results obtained are very good classification

2. Support Vector Machine dengan Genetic Algorithm
   From these results we process student data with hybrid optimization with Genetic Algorithm obtaining an accuracy of 86.57% with an average of 86.56% with the confusion matrix of the SVM model with GA which can be seen from table 3.

Table 3. Confusion Results of the Hybrid Genetic Algorithm Matrix

Information	On time (Yes)	On time (No)	Class Precission
Prediction Yes	615	88	87.48%
Prediction No	19	74	79.57%
Class Recall	97.00%	45.68%	

Source: (Ridwansyah, Wijaya, and Purnama 2020)

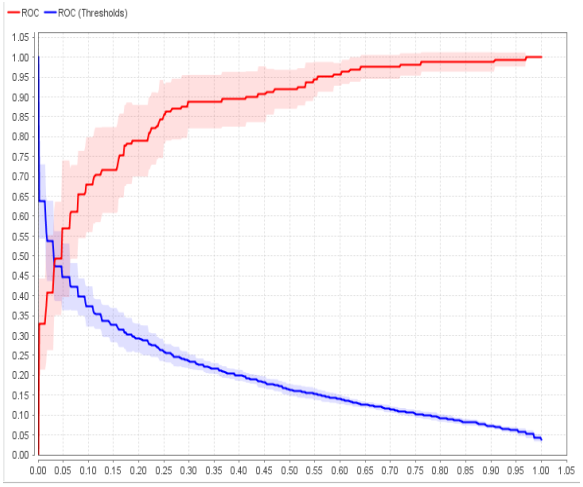
From table 2 it can be seen that the prediction of students who graduate on time is 615 and the results of the SVM algorithm are in accordance with what was predicted but the results of students who predicted to graduate on time did not graduate on time. The number of students as many as 19 students is estimated to graduate not on time to graduate on time. From the results of predictions using the SVM method, the accuracy formula can be calculated as follows:

$$\text{Accuracy} = \frac{(615+74)}{(615+19+74+88)}$$

$$\text{Accuracy} = \frac{(689)}{(796)} = 0{,}86577$$

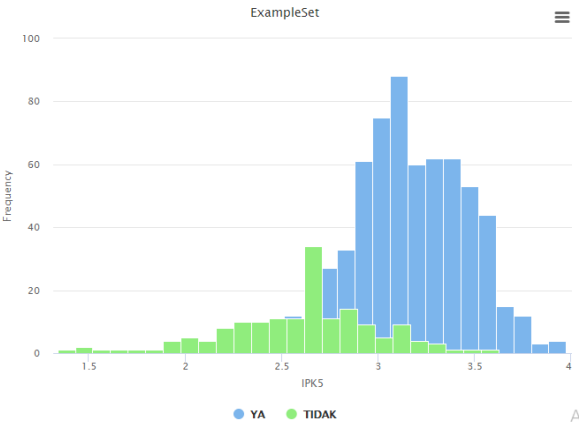
With the calculation of accuracy, the accuracy of student graduation is 0.86577 with a value of 86.57% percent. And obtained the ROC curve with

AUC of 0.878 which can be seen in Figure 4. With the graduation diagram of students in Figure 5.



Source: (Ridwansyah, Wijaya, and Purnama 2020)

Figure 4 *Area Under Curve* (AUC) Optimasi Hybrid GA



Source: (Ridwansyah, Wijaya, and Purnama 2020)

Figure 5. Student Graduation Diagram with GA Hybrid Optimization.

CONCLUSION

Not only for graduate students is also an important part for all parties related to education so it is very necessary to conduct further research to predict the time of graduation of students..

We apply SVM-GA hybrid optimization to predict student graduation. In the GA-SVM method, GA is used to select parameters that are appropriate for the classification of the SVM method. The experimental results show that the accuracy of the GA-SVM classification is higher than the accuracy of the SVM classification. Thus the graduation of students can be reached on time and as much as possible

REFERENCE

Ashok, M. V., and A. Apoorva. 2016. "Data Mining Approach for Predicting Student and Institution's Placement Percentage." *2016 International Conference on Computation System and Information Technology for Sustainable Solutions, CSITSS 2016*: 336–40.

Bin, Li, and Yang Min. 2012. "Analysis Model of Drilling Tool Failure Based on PSO-SVM and Its Application." *Proceedings - 4th International Conference on Computational and Information Sciences, ICCIS 2012* (8): 1307–10.

Devasia, Ms.Tismy, Ms.Vinushree T P, and Mr.Vinayak Hegde. 2008. "Prediction of Students Performance Using Educational Data Mining." *International Journal of Cognitive Therapy* 1(3): 266–79.

Freitas, Frances Anne, and Lora J. Leonard. 2011. "Maslow's Hierarchy of Needs and Student Academic Success." *Teaching and Learning in Nursing* 6(1): 9–13.

Gao, Xiang Ming, Shi Feng Yang, and Yu Hu. 2010. "Leakage Forecasting for Water Supply Network Based on GA-SVM Model." *Proceedings of the 2010 Symposium on Piezoelectricity, Acoustic Waves and Device Applications, SPAWDA10*: 206–9.

Jiang, Huiyan, Fengzhen Tang, and Xiyue Zhang. 2010. "Liver Cancer Identification Based on PSO-SVM Model." *11th International Conference on Control, Automation, Robotics and Vision, ICARCV 2010* (December): 2519–23.

Li, Hua, and YongXin Zhang. 2009. "An Algorithm of Soft Fault Diagnosis for Analog Circuit Based on the Optimized SVM by GA." In *2009 9th International Conference on Electronic Measurement & Instruments*, China: IEEE.

Liu, Han et al. 2019. "Effective Data Classification via Combining Neural Networks and SVM." *Proceedings of the 31st Chinese Control and Decision Conference, CCDC 2019*: 4006–9.

Ridwansyah, Ridwansyah, Ganda Wijaya, and Jajang Jaya Purnama. 2020. *Laporan Akhir Penelitian Mandiri*. Jakarta.

Riyanto, Verry, Abdul Hamid, and Ridwansyah. 2019. "Prediction of Student Graduation

Time Using the Best Algorithm." *Indonesian Journal of Artificial Intelligence and Data Mining* 2(2): 1–9.

Suhardjono, Ganda Wijaya, and Abdul Hamid. 2019. "PREDIKSI WAKTU KELULUSAN MAHASISWA MENGGUNAKAN SVM BERBASIS PSO." *Bianglala Informatika* 7(2): 97–101.

Wang, Gui Ping, Jian Xi Yang, and Ren Li. 2017. "Imbalanced SVM-Based Anomaly Detection Algorithm for Imbalanced Training Datasets." *ETRI Journal* 39(5): 621–31.

Ye, Xuehui, Yuxia Li, Ling Tong, and Ling He. 2017. "Remote Sensing Retrieval of Suspended Solids in Longquan Lake Based on GA-SVM Model." *International Geoscience and Remote Sensing Symposium (IGARSS)* 2017-July: 5501–4.

Yu, Ting Chun, and Jui Chung Hung. 2017. "Forecasting MLB Playoff Teams Using GA-SVM." *Proceedings of the 2017 IEEE International Conference on Applied System Innovation: Applied System Innovation for Modern Technology, ICASI 2017*: 446–48.