Indonesian Journal of Artificial Intelligence and Data Mining (IJAIDM)
Vol. 3, No. 2, September 2020, pp. 95 – 102
p-ISSN: 2614-3372 | e-ISSN: 2614-6150

Comparison of Data Mining In E-Learning Learning Based On Log Aktivity On PSO-Based Nural Network Algorithms With PSO-Based SVM

[1]Elin Panca Saputra, [2]Supriatiningsih, [3]Indriyanti
[1,2,3] Information Systems, University of Bina Sarana Informatika
Email: [1]elin.epa@bsi.ac.id , [2]supriatiningsih.stq@bsi.ac.id, [3]indriyanti.iyt@bsi.ac.id

Article Info
Article history:
Received Aug 28th, 2020
Revised Sept 15th, 2020
Accepted Oct 05th, 2020

Keyword:
AUC
E-learning
Neural Network
NN-PSO
SVM PSO

ABSTRACT
Our research objective is to obtain comparison results from several algorithms, we made a comparison the Neural Network method combination of Particle Swarm Optimization and Vectore Machine support algorithm with Particle Swarm Optimization, from evaluating student learning outcomes with e-learning lectures conducted by students. Accuracy of results in learning. And the purpose and objective of this research is to select several attributes that influence during student learning process. The attributes we use from students who undergo e-learning lessons are taken from the activity log during learning where there are activities. The logs we use are like gender, Exercises, Forums, Chats, Discussions, Uploaded Tasks, Messages, Practice Quizzes and Total Notes. Student activity logs are stored in the LMS database. The data used in this study, from the calculation results obtained by the Neural Network (NN) algorithm combined with PSO, obtained a percentage of 97.35%, and the results of the AUC step. Then the second trial was carried out using the support vectore machine (SVM) algorithm combined with PSO with a proportion of 88.47% and AUC proportion of 93.80%. Then the results using the neural network algorithm combined with PSO have a higher yield compared to the spport vector machine method combined with PSO with a percentage ratio of 8.88% while the proportion value on AUC is 4.8%.

Corresponding Author:
Elin Panca Saputra,
Departement Information Systems,
University of Bina Sarana Informatika,
Jl. Kramat Raya No. 98 Senen, Jakarta Pusat, Indonesia
Email:elin.epa@bsi.ac.id

DOI: http://dx.doi.org/10.24014/ijaidm.v3i2.10519

1. INTRODUCTION

Data mining is a process to find relationships that have meaning, patterns, and refer to by examining a set of big data stored in a storage using techniques from pattern recognition such as statistical and mathematical techniques [1]. Data mining has become one of the effective tools for data analysis and knowledge management systems, therefore there are many fields that have adapted approaches to data mining to solve problems [2]. The learning process based on e-learning is implemented in higher education institutions using LMS Moodle version 3.8. Moodle's LMS is a free Learning Management Systems section, largely in response to a large community Management Systems Course [3]. LMS Moodle is currently widely used by schools and colleges to support e-learning based learning systems because it has many features that support it. That is our reason for analyzing because from various countries using the Moodle LMS system, which is done by providing lecture materials, presenting lecture material by lecturers, as well as practicum assignments and ending with the results of the exams that have been provided, because Moodle is one of the most widely used platforms internationally. [4].

LMS is an activity theory [5], Among them are class interactions or in-person meetings in class and online support so that convergence occurs [6], which has the aim of improving the learning development of students and lecturers. Students' collaborative learning environment Can develop abilities that encourage the formulation of questions, explain and justify opinions, articulate a reasoning, develop and reflect on the knowledge acquired [5]. LMS must be supported by the latest information technology which is one of the pillars of temporary society and is a normal part of everyday life for most people. without exception is the world of education. Until now, consumption for technology in education varies in each country and type of educational institutions, as well as between other institutions[4].

To get e-learning user activity, we extract knowledge from the expected e-learning data assuming that a user action path can be taken to identify certain information about the user. That e-learning users can be obtained from activity activities and the data used to form a model that predicts student behavior, or explains its validity in carrying out activities on e-learning [1]. With that, the management system in student e-learning learning is not designed with data analysis because usage data is not made systematically. because thorough analysis takes a long time which can be tedious [7], The LMS system has generate statistical reports. However, this report cannot assist the instructor in making conclusions that are useful to either the potential of the course or the capacity of the students and only have purposes for administrative purposes on the platform. Based on the discussion of this background, the problem we discuss in this research is how to implement data mining by comparing between PSO-based Neural Network algorithms and PSO-based Support Vector Machines.

Our research objective is to compare the level of accuracy of an algorithm that we tested, namely implementing an artificial neural network and support vector machine with each algorithm combining with Particle swarm optimization (PSO) and for the determination of the attribute selection to compare the level of accuracy based on the results of e-learning based learning activities. log activity contained in the learning management system. However, in choosing a selection feature and increasing the level of accuracy, in the success of e-learning learning, the author has difficulty determining these features. selection of features is defined as the process used in search engines where the feature subset that contains data is processed in a learning algorithm application. [8]. The reason we use Neural Network and Support Vector Machine is the difference with neural network algorithm sense which tries to find the hyperplane from one of the classes, meanwhile Support Veectore Machine (SVM) to try to find the best hyper fields in the space. The basic principle of SVM is classification on a linear, and the following will be developed in order to do non-linear exercises. by entering the concept of tricks on the kernel in a workspace that has high dimensions [18] .

Algorithm using Particle Swarm Optimization (PSO) which can assume a very potential solution, Refers to observation - it's like a flock of birds. Foraging is represented in an algorithm as a continuous updating of a parameter that accounts for the elements of the set [9]. This article begins with a brief Optimization description uses on metaheuristic algorithms, typical PSO applications, and the possibility for the addition of a basic algorithm. Then, the PSO algorithm and details of the line model are presented in a few. This section presents two examples of neuron tuning coefficients using PSO - two cases of neural networks and a support vector engine for analysis.

The results of our trials using several methods that we apply to the neural networks method have the highest accuracy, because the algorithm is an artificial intelligence method. So that the results of the comparison of the level of accuracy that can be obtained are greater than the In supporting vector engine algorithms, because the supporting vector engine or SVM has several drawbacks or weaknesses, the weakness in the supporting vector engine algorithm is the need to select a good kernel function for good performance, using a separator function that separates the data into two different classes [12]. In addition, by getting the results of a comparison between Neural Networks with PSO-based support vector machines, the research that we do can also find out what attributes can determine the level of learning success using LMS. So that by using the PSO Algorithm, besides being able to increase the accuracy results, it can also determine the selected attributes.

2. RESEARCH METHOD

The research method used to compile the research that is being carried out includes research:

2.1 Understanding Data Mining

The term understanding of data mining is a term used to search hidden knowledge in a database. Data Mining (DM) aims to perform deep or high-level excavation of raw data [13].

2.2 Understanding Artificial Neural Network Algorithm (ANN)

Explanation of a Neural Network is an abstraction and simulation of the structure and function of biological neural systems. The main network that has developed independently but over time appears new contexts and applications that cover a wider area covering the neural network. in fact, the technique of an

artificial neural network is naturally connected to one another to form a computational set with such a strong theoretical basis and with unquestionable solve a problem in various fields of information technology. [14]. Therefore, the researcher chose one of the Neural Network Algorithms because It has had success applied to various fields of research and industry to perform tasks such as forecasting, data classification, and regression analysis [15]. Neural network is a computerized program designed to simulate the learning function of the human brain, and has the advantage of being good at generalizing, as well as drawing from the experiences that can be obtained. [8]. The following is a model that describes the Neural Network Algorithm.

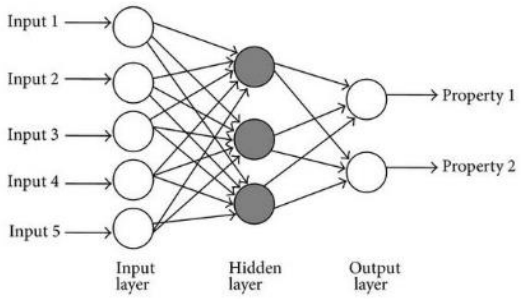


Figure 1. Neural Network Model [13]

2.3 Understanding Support Vector Machine Algorithm

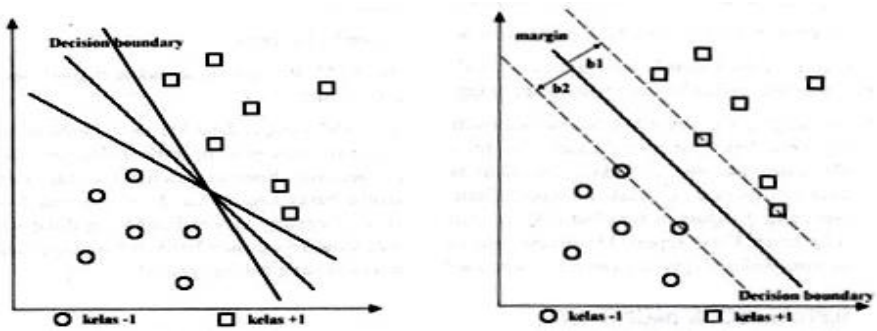
an explanation of a Support Vector Machine Algorithms are new algorithms in the field of intelligence made by humans that have a theory of learning about statistics in users. It has proven success applied to many fields and has gained a lot of interest. It was first introduced by Boser et al in 1992. besides that the notion of a Support Vector Machine (SVM) Algorithm is a suitable classification and better for classification shows better application clarification performance, and has generalization performance, which can avoid local minimum structure and problems for small sample cases in machine learning, besides being absent in human ability and knowledge and suitable for solves a nonlinear and higher dimensional problem [16]. Here, we will be limited to solving a constrained optimization problem usually solved by using a multiple problem which is defined as follows [17]:

$$-\frac{1}{2}\sum_{i,j=1}\alpha_i\alpha_j < x_i x_j > -\varepsilon \sum_{i=1}\alpha_i + \sum_{i=1} y_i\alpha_i \quad (1)$$

αi = βi − β* i and βi, βi* are obtained by completing a quadratic program and being a Lagrangian multiplier. Once this optimization problem is resolved, the parameter vector w is obtained by:

$$w = \sum_{i=1}(\beta_i^* - \beta_i)\varphi(X_i) \quad (2)$$

the explanation of the Support Vector Machine is to maximize hyperplane limit (maximal margin hyperplane), as illustrated in Figure 2 (a) there are a number of possible hyperplane options for the dataset, and 2 (b) is a hyperplane with the maximum margin**.** Although 2 (aactually can use any hyperplane, hyperplane with maximum margin will give better generalization to the classification method.



a. Possible Boundary Drawings b. Decision boundary with maximum margin

Figure 2. Draw the possible boundaries for the dataset

2.4 Explanation Particle Swarm Optimization Algorithm (PSO)

explanation on Algorithm Particle swarm optimization or (PSO) is a technique used to search for optimized network configuration through a particle moving in a limited search space capable of producing a globally optimal solution in the search space [10].

3. RESULT AND ANALYSIS

The results of calculating and testing our data use the Artificial Neural Network (ANN) method which compares the SVM method, each of which uses the Particle Swarm Optimization algorithm, which is as follows.

3.1 Calculations Neural Network method combined with PSO

To see the results of processing and examiners in calculating data, Neural Network (ANN) method is combined PSO, which we do with the help of the Rapidminer Version 5.3 software,which is as follows.

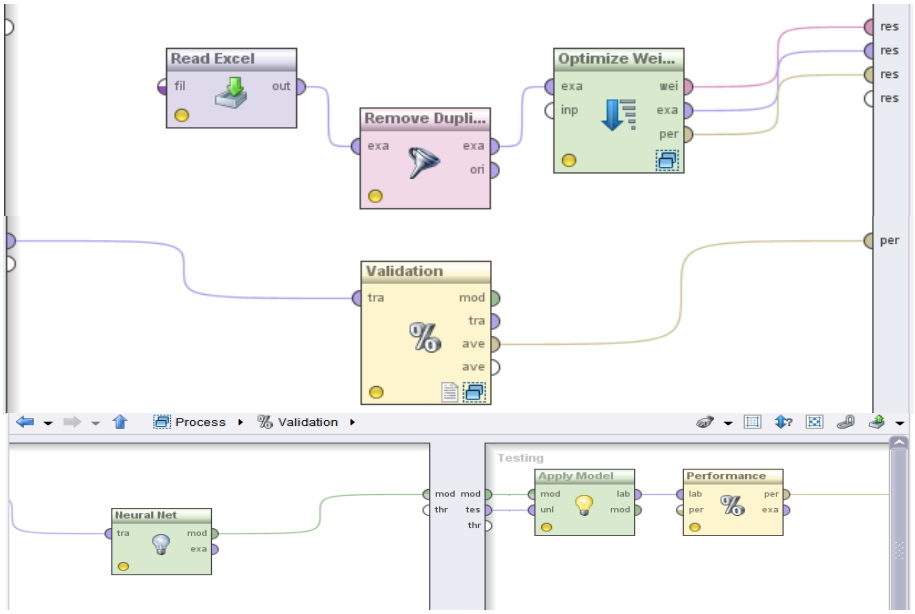


Figure 3. Examination The results of the K-Fold Cross Validation Test with Neural Network (NN) combined with Particle Swarm Optimization (PSO)

After calculating the data with calculations on the Neural Network algorithm, the configuration matrix to see it is in table 1.The results obtained by the value of True Positive (TP) with a score of 407 classified with 1 made with the Neural Network algorithm for prediction, and the calculation obtained, the number of False Negative (FN) values has a total of 12, the data will be predicted to be 1 but become 2, after that the calculation of True Negative or (TN) has a total of 217 as predicted, and False Positive (FP) has a total of 5, the predicted data will be 2 as it turns out is 1. Therefore, the level of accuracy generated by the algorithm on the Neural Network (NN) combined with PSO the proportion value is 97.35%, while the resulting AUC percentage value is 0.986 and can be summed. calculations obtained, sensitivity results, specificity results, and PPV results can be shown in the table 1:

Table 1. Results of Testing and Data Processing on the Method

Classification	Predicted C	
	C = 1	C = 2
C = 1	407	12
C= 2	5	217

C = Class

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} = \frac{407 + 217}{407 + 217 + 5 + 12} = 0{,}9735$$

$$Sensitivity = \frac{\text{TP}}{\text{TP + FN}} = \frac{407}{407 + 12} = 0{,}9713$$

$$Specificity = \frac{\text{TN}}{\text{TN+FP}} = \frac{217}{217+5} = 0{,}9774$$

$$PPF = \frac{\text{TP}}{\text{TP + FP}} = \frac{407}{407 + 5} = 0{,}9878$$

$$NPV = \frac{\text{TN}}{\text{TN + FN}} = \frac{217}{217 + 12} = 0{,}9475$$

From the calculation of the data above, it is obtained by applying the Neural Network (NN) algorithm and PSO to produce a proportion with an accuracy value of 97.35% and the calculation has been carried out by algorithm for the level of accuracy, specificity, and npv results. While the results of the weight attribute value are as follows:

Table 2. Results of weighting attribute value data

Atribute	Weights
gen	0
Practice Questions	1
Forum	0
Chatting	1
Diskusi	0.526
Upload	1
Message	1
Quiz	0.148
Total Log	1

3.2 Test Results with the PSO combination Support Vector Machine (SVM)

The results of processing and checking data in data computing using a SVM to be combined with the Particle Swarm Optimization (PSO) algorithm which is carried out using the Rapidminer Version 5.3 software are as follows.

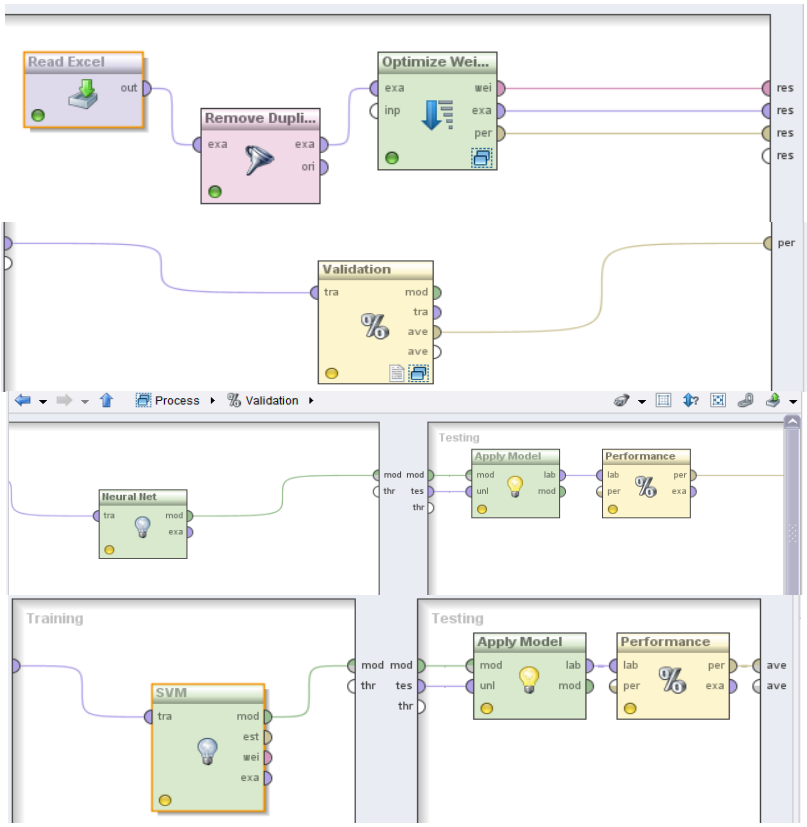


Figure 4. Testing on K-Fold Cross Validation Support Vector Machine (SVM) Based on Particle Swarm Optimization (PSO)

After calculating using the SVM algorithm combined with PSO algorithms and configuration matrices can be shown in detail in table 3.With the number of True Positive or (TP) in the amount of 402 which will be classified as 1 according to the prediction given. Created with the Support Vector Machine (SVM) algorithm, and the result of False Negative (FN) with a total of 64 data will be predicted to be 1 but it is predicted to change to 2, then the result is TN with a value of 165 data prediction 2, is in accordance with the prediction , and as many as 10 False Positive (FP) 2 prediction data turns out to be prediction 1.Therefore, the level of accuracy produced by the PSO-based Support Vector Machine (SVM) algorithm has a percentage value of 88.47%, while the AUC percentage value is 0.938%. and can be calculated to produce values, results from sensitivity, results from specificity, results from ppv and npv per count The results can be seen as follows.

From the results of the presentation produced with the Support Vector Machine (SVM) algorithm combined with PSO will produce a value by producing a proportion of 88.47%, and it has been processed to produce value, specificity results, ppv results, and results. npv results.

Table 3. Testing Results and Data Processing

Classification	**Predicted Class**	
	C = 1	C = 2
C = 1	402	64
C= 2	10	165

C :Class

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} = \frac{402 + 165}{402 + 165 + 10 + 64} = 0{,}8845$$

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} = \frac{402}{40264} = 0{,}8626$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TP} + \text{FN}} = \frac{165}{165 + 10} = 0{,}9428$$

$$\text{PPF} = \frac{\text{TP}}{\text{TP} + \text{FP}} = \frac{402}{402 + 10} = 0{,}9757$$

$$\text{NPV} = \frac{\text{TN}}{\text{TN} + \text{FN}} = \frac{165}{165 + 64} = 0{,}7205$$

From the results of the above data processing obtained by combining the Support Vector Machine (NN) algorithm and PSO resulting in a percentage value of 97.35% and these have been computed to produce a value the correct presentation value, specificity results, ppv results, and npv results. While the results of the weight attribute value are as follows:

Table 4. The results of the data weights attribute value

Atribute	**Weights**
Gen	0.80
Practice Questions	0.73
Forum	0.20
Chatting	0.20
Diskusi	0.93
Upload	1
Message	0
Quiz	1
Total Log	0

3.3 Comparison results of the Neural Network algorithm with Support Vectore Machine Combined with Particle Swarm Optimization

From the results of these tests we calculated the Neural Network algorithm combined with Particle Swarm Optimization and the Support Vector Machine algorithm combined with Particle Swarm Optimization. for details in table 5:

Table 5. Comparation of accuracy and AUC

Algorithm	Accuracy	AUC	Comparison
Neural Network algorithm And PSO	97,35%	98,60%	Accuracy 8,88%
Support Vector Mechine algorithm And PSO	88,47%	93,80%	AUC 4,80%

4. CONCLUSION

The calculations we have done in testing use the Neural Network algorithm combined with (PSO) and the Support Vector Mechine (SVM) method combined with (PSO), data we get from logs or student activity activities from learning styles students with the elearning learning model. So it can be seen that the percentage value generated from the calculation produces an accurate value and AUC results of each algorithm that has been tested the Neural Network (NN) combined with (PSO) produces, the proportion value as much as 97.35% while the AUC value with a proportion value of 98.60%. Meanwhile, the calculations carried out on the SVM algorithm combined with PSO obtained an accuracy percentage value of 88.47% and the AUC percentage result of 93.80%. So it can be proven from testing data on student activities using the Neural Network, algorithm combined with (PSO), so that the results are obtained that this method is superior or higher in its accuracy value compared to the SVM. combined with (PSO), based on an increase in the proportion of 8.88% and the percentage value is AUC of 4.80%, it is clear that this result is included in the (very good) category. However, in reality, there are still shortcomings in using a neural network algorithm, namely, the training process that requires a long time if the data used is very large. and the following are some of the attributes selected in determining the level of accuracy of elearning learning based on of student activity using the NN algorithm, combined with (PSO) of the 9 selected attributes, such as Exercise, Chat, Discussion, Upload Total Tasks, Messages, and log totals. Whereas with the Combined Support Vector Machine (SVM) method (PSO), there are 9 selected attributes including gender, discussion, collecting assignments, and Quiz Exercise.

REFERENCES

[1]. D. T. Larose and C. D. Larose, *Discovering Knowledge in Data: An Introduction to Data Mining: Second Edition*, vol. 9780470908. 2014.

[2]. A. M. Ahmed, A. Rizaner, and A. H. Ulusoy, "Using data Mining to Predict Instructor Performance," *Procedia Comput. Sci.*, vol. 102, no. August, pp. 137–142, 2016.

[3]. M. Jovanovic, M. Vukicevic, M. Milovanovic, and M. Minovic, "Using data mining on student behavior and cognitive style data for improving e-learning systems: A case study," *Int. J. Comput. Intell. Syst.*, vol. 5, no. 3, pp. 597–610, 2012.

[4]. N. Kerimbayev, J. Kultan, S. Abdykarimova, and A. Akramova, "LMS Moodle: Distance international education in cooperation of higher education institutions of different countries," *Educ. Inf. Technol.*, vol. 22, no. 5, pp. 2125–2139, 2017.

[5]. M. Jones and N. H. Hashim, "Activity theory: a framework for qualitative analysis," *4th Int. Qual. Res. Conv.*, pp. 3–5, 2014.

[6]. M. Vásquez Astudillo, "Aplicación de modelo pedagógico Blended Learning en educación superior," *Rev. DIM Didáctica, Innovación y Multimed.*, no. 35, p. 1, 2017.

[7]. A. Krüger, A. Merceron, and B. Wolf, "A data model to ease analysis and mining of educational data1," *Educ. Data Min. 2010 - 3rd Int. Conf. Educ. Data Min.*, pp. 131–140, 2010.

[8]. E. P. Saputra and A. Hamid, "Fitur Seleksi Atribut Hasil Kelulusan Mahasiswa Elearning Berdasarkan Log Dengan Neural Network," *J. Kaji. Ilm.*, vol. 19, no. 1, p. 24, 2019.

[9]. M. Kaminski, "Neural network training using particle swarm optimization - A case study," *2019 24th Int. Conf. Methods Model. Autom. Robot. MMAR 2019*, pp. 115–120, 2019.

[10]. M. S. Hasibuan, L. E. Nugroho, and P. I. Santosa, "Model detecting learning styles with artificial neural network," *J. Technol. Sci. Educ.*, vol. 9, no. 1, pp. 85–95, 2019.

[11]. K. M. G. Noama, A. Khalid, A. A. Muharram, and I. A. Ahmed, "Improvement of E-learning Based via Learning Management Systems (LMS) Using Artificial Neural Networks," *Asian J. Res. Comput. Sci.*, vol. 4, no. 1, pp. 1–9, 2019.

[12]. P. Janardhanan, L. Heena, and F. Sabika, "Effectiveness of support vector machines in medical data mining," *J. Commun. Softw. Syst.*, vol. 11, no. 1, pp. 25–30, 2015.

[13]. J. Djuriš, D. Medarević, M. Krstić, I. Vasiljević, I. Mašić, and S. Ibrić, "Design space approach in optimization of fluid bed granulation and tablets compression process," *Sci. World J.*, vol. 2012, 2012.

[14]. A. Prieto *et al.*, "Neural networks: An overview of early research, current frameworks and new challenges," *Neurocomputing*, vol. 214, pp. 242–268, 2016.

[15]. X. Liu, "Full-Text Citation Analysis : A New Method to Enhance," *J. Am. Soc. Inf. Sci. Technol.*, vol. 64, no. July, pp. 1852–1863, 2013.

[16]. W. Gong and W. Wang, "Application research of support vector machine in E-learning for personality," *CCIS2011 - Proc. 2011 IEEE Int. Conf. Cloud Comput. Intell. Syst.*, pp. 638–642, 2011.
[17]. M. Sarhani, A. El Afia, and R. Faizi, "Hybrid approach-based support vector machine for electric load forecasting incorporating feature selection," *Int. J. Big Data Intell.*, vol. 4, no. 3, p. 141, 2017.
[18]. Suci Aulia, Sugondo Hadiyoso, and Dadan Nur Ramadan, "Comparative Analysis of NPC with SVM for Classification of Diabetes Retinopathy based on Exudate and Microaneurysm Images," *Int. J. Elkomika.*, vol. 1, no. 3, p. 75, 2015.

BIBLIOGRAPHY OF AUTHORS

Elin Panca Saputra as a lecturer from the Bina Sarana Informatika University, currently the leader of the Elearning team on the campus. In 2020, it has passed the ICSID Conference indexed by Scopus which was held on 6 & 7 August 2020. As well as being active as a member of APTIKOM.

Supriatiningsih as a Lecturer from the Bina Sarana Informatika University. Currently, She joins the Elearning Team at the campus as a developer of online-based learning systems. In 2020 this has passed the ICSID Conference indexed by Scopus which was held on 6 & 7 August 2020.

Indriyanti as a Lecturer from Bina Sarana Informatika University. Currently, She joins the Elearning Team at the campus as a developer of online-based learning systems. In 2020 this has passed the ICSID Conference indexed by Scopus which was held on 6 & 7 August 2020.