International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-8 Issue-12, October 2019

Features Selection Based ABC-SVM and PSO-SVM in Classification Problem

Mochamad Wahyudi, Sfenrianto Sfenrianto, Weiskhy Steven Dharmawan

Abstract: ***Feature selection can be used to improve the performance of classification algorithms. This study aims to use algorithms for feature selection in classification problems. The method in feature selection use to the Artificial Bee Colony (ABC) algorithm and Particle Swarm Optimization (PSO) based on Support Vector Machine (SVM). The ABC-SVM algorithm functions as a method of feature selection to choose the optimal subset according to the objectives set to provide the results of the classification. Then, the PSO-SVM as a comparison method in other feature selection. The results of the classification conducted by PSO-SVM with the vowel dataset is good classification (AUC 0.873), when compared with the ABC-SVM with the classification result is superior (AUC 0.996). The results of the Precision Recall and F-Measure calculations on the PSO-VSM algorithm have good classification results for sonar and wavefrom data sets. Meanwhile, the results of tests conducted by ABC-SVM get superior value from the classifier quality efficiency in the vowel data set.***

Keywords : ***Feature Selection, Attribute Selection, Artificial Bee Colony, Support Vector Machine***

I. INTRODUCTION

Humans need information quickly and up to date, in the form of data and documents. But the obstacle that arises is how to organize a large collection of data or information with large dimensions to be well organized so that it is easy to use it. High data dimensions affect the identification of dimensions or variables in a data or information. Thus the process of reducing the dimensions of the data is needed accurately and efficiently to produce low-dimensional data from high dimensions.

One type of solution that can be done to deal with the problem of large data dimensions is to perform feature selection.

Revised Manuscript Received on October 30, 2019.
***** Correspondence Author
Mochamad Wahyudi, Master of Computer Science - Postgraduate Programs STMIK Nusa Mandiri, Indonesia, Email: wahyudi@bsi.ac.**id.**
Sfenrianto Sfenrianto*, Information Systems Management Department, BINUS Graduate Program – Master of Information Systems Management, Bina Nusantara University, Jakarta 11480 Email : sfenrianto@binus.edu
Weiskhy Steven Dharmawan, Master of Computer Science - Postgraduate Programs STMIK Nusa Mandiri, Indonesia, Email: weiskhy.wvn@bsi.ac.id**.**

It is a process that reduces the dimension of features by selecting important attributes and eliminating irrelevant, redundant and noisy attributes to get a more accurate data classification. Thus, feature selection is an important step in text classification and directly affects the classification performance.

Some research has been done in features selection with better results. For example, the impact of using Ant Colony Optimization to perform features selection shows better performance than other algorithms in the aspects of precision and recall [1]. In Zahran's research using Particle Swarm Optimization (PSO) in the pre-processing stage in the categorization of text documents. The results of trials conducted showed PSO is better on average in the aspects of precision and recall compared to the others [2].

In order to enhance the features selection effect, many studies have tried to add intelligent optimization algorithms to the feature's selection method. Particle Swarm Optimization algorithm is the simplest and fastest algorithm in the application process to find the value of optimization [3]. In terms of other research classifications Support vector machine (SVM) works better than Multi Layer Perceptron (MLP), this is due to the ability to generalize Support Vector Machine in high dimensional space [4]. Support Vector Machine is used for pattern classification, mapping in the input space with nonlinear change to a high dimensional space, where the linear classification problem is optima [5].

SVM performance is highly dependent on an adequate choice of parameter values, kernel, and regularization parameters. The choice of SVM parameters is generally an optimization problem where search techniques are used to find parameter configurations that maximize SVM performance [6]. There are many optimization techniques that have been used to optimize parameters in machine learning, such as genetic algorithms (GA) and Particle Swarm Optimization (PSO) [7] [8]. From this description, SVM is very good in classification, SVM is generally an optimization problem where search techniques are used to find parameter configurations that optimize SVM performance [6]. However, SVM has a weakness in optimizing the prediction of dimensional data which is quite large. In other studies, the Artificial Bee Colony (ABC) algorithm which is included in the genetic algorithm (GA) is proposed to solve the numerical problem optimization which is inspired by the behaviour of honey bees in searching for food [9].

In this research, Artificial Bee Colony (ABC) will be applied to Feature Selection by applying feature selection / attributes on large datasets to optimize the performance of the Vector Machine (SVM) in classification problems.

II. LITERATURE REVIEW

Artificial Bee Colony (ABC) is an optimization algorithm [9]. The use of ABC is expected to attribute the selection in SVM to improve accuracy and classification. The problem of the many attributes will reduce accuracy and increase the complexity of the Support Vector Machine (SVM) algorithm. For this reason, Artificial Bee Colony is used to solve this problem. Then a 10-Fold Cross validation test was performed on the performance of the two methods, the results shown by the Confusion matrix and the ROC Curve to get the Area Under Curve (AUC) value.

While, the Feature selection is an expensive process, and also contradicts the initial assumption that all information or attributes are needed in order to achieve maximum accuracy [10]. It is requiring high computational costs when dealing with large data sets. Reducing the dimension of the number of attributes of a data set or group of attributes can effectively cut these costs. Feature Selection is carried out by minimizing losses that can occur due to loss of some information. The purpose of Feature Selection in the data mining domain is to identify the smallest cost at which data mining algorithms can keep error rates below the efficiency line border. Cost referred to is a function of the theoretical complexity of the data mining algorithm derived from the model, and correlates with the time required for the algorithm to run the model, as well as the size of the data set [11].

The purpose of attribute selection is to reduce the attributes of a dataset to eliminate variables that are considered irrelevant. Feature selection methods can be classified into three main categories [12], namely: (1) Filter method for selecting relevant attributes before moving on to the next learning phase, the attributes that are considered the most important are chosen for learners, while the rest are excluded; (2) The wrapper method for assessing a group of variables using the Artificial Intelligence classification to find the best features with an iterative searching process; (3) The embedded method for the attribute selection process is located within the learning algorithm, so that the selection of the optimal set of attributes is directly made during the generation phase of the model

Research conducted by Yonghe Lu, Minghui Liang Ye, Lichao Cao in [3], regarding features selection using Particle Swarm Optimization (PSO) and methods for text classification using the KNN method. The dataset used is Reuters-21578. Preprocessing data using standard analyzer in lucene 3.0. The dataset was processed using PSO feature selection by conducting 6 experiments, namely: Fixed inertia weight, Improved inertia weight, Fixed constriction factor, Improved constriction factor, Improved synchronously inertia weight and constriction factor, Improved asynchronously inertia weight and constriction factor. The experimental results show that the most stable accuracy is the 6th trial. Evaluation in this study uses the Convergence Curve and MacF1. Then a significant test was done using a T-test and the results in program 6 showed significant differences to other programs except for programs 3 and 4. The accuracy value in experiment 6 was 80.64%. Figure 1, shows the model proposed in the study [8].

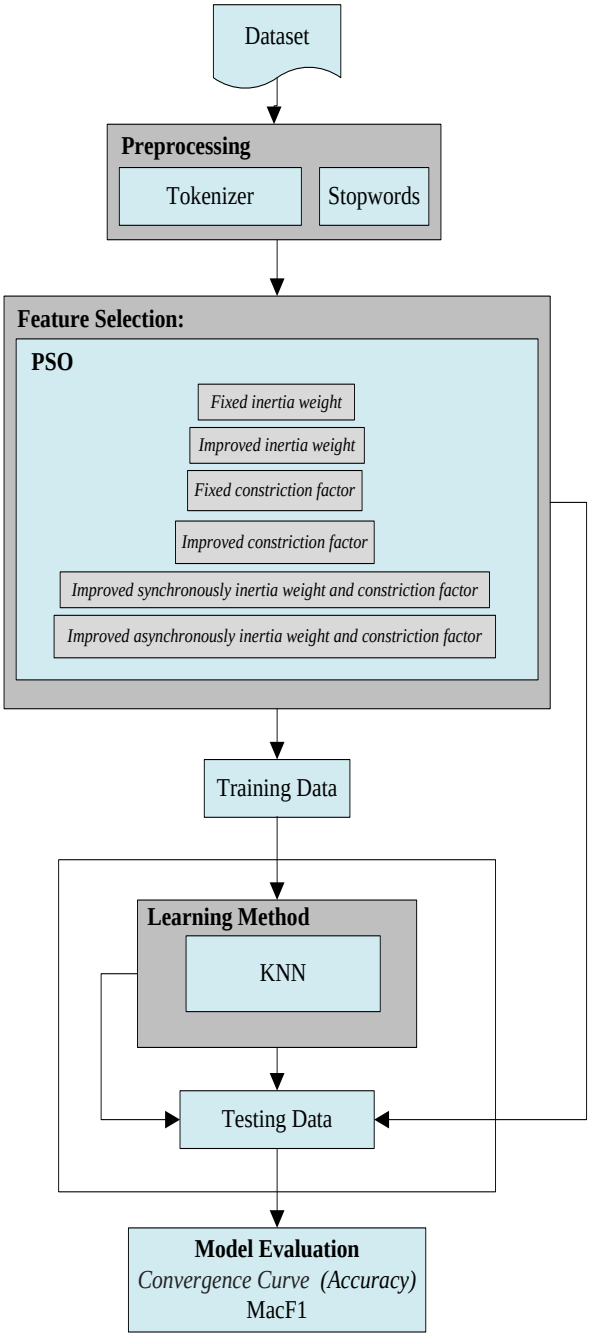


Fig.1 The model proposed by Lu [3]

III. PROPOSED MODEL

The data used in this study uses UCI data repositories including Vowel, Sonar Wavefrom and Heart statlog. Vowel dataset is a speech data pattern in eleven English vowels, when the speaker speaks each speech will be recorded in an input array in ten formats. Sonar data is a variety of signal data that is painted to classify sonar signal types. While the Wavefrom data is a data generator wave database that has 2 wave waves class, and Heart statlog is a heart database that has variables that must be predicted whether or not to have a heart disease event. The data can be obtained from https://archive.ics.uci.edu/ml/datasets.

Then, the proposed model in this study uses a model in the training data that has been selected variable. The training data was tested using a model that is based on Artificial Bee Colony Algorithm and the classification of data will use Support Vector Machine to produce an increase in average Precision, Recall and F measure, and Area Under Cover (AUC) values compared to the method the comparison is PSO-SVM. Tests conducted on training data and Testing with this method are done with 10 fold cross validation technique. This is done with the aim of producing the highest accuracy in the t-test results of the method. The results of this classification can be used as the best value after the 10 fold cross validation technique. All datasets are tested by the proposed method / model. Experiments are carried out one by one on the dataset and the model used. Figure 2 is the proposed method flow of this research.

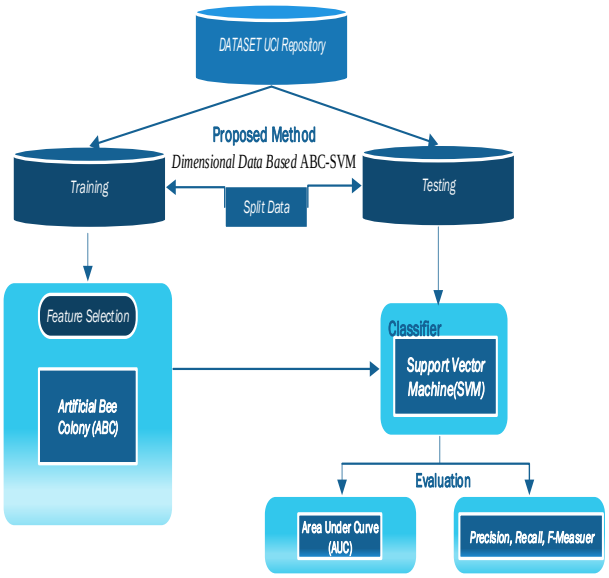


Fig 2 A Proposed model

IV. EXPERIMEN RESULT AND ANALYSIS

Artificial Bee Colony is used for features selection. it is the process of reducing the dimensions of features by selecting important features and eliminating irrelevant, redundant and noise features to get a more accurate data representation. Three approach models in feature selection are: (1) Filter Method; (2) Wrapper Method; and (3) Embedded Metohd. The Filter Metohd evaluates features by ranking features independently of classes in the traning set. Whereas the Wrapper Method uses artificial intelligence methods to find the best features with an interactive searching process. Embedded Method uses a simultaneous linear prediction approach to increase goodness-of-fit and decrease the number of input features.

Based on the three approaches above, the approach used in this study belongs to the second approach, the Wrapper Method. Figure 3 shows the application of Feature Selection in the vowel dataset resulting from the application of the Artificiall Bee Colony algorithm (ABC) using the Wrapper Method approach and Embedded Method.

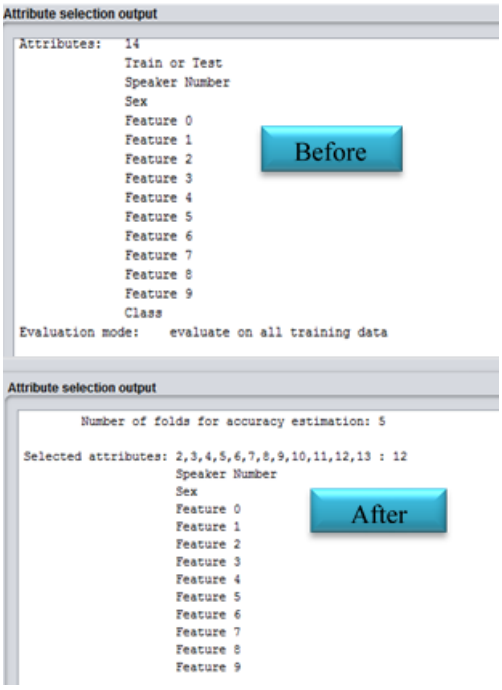
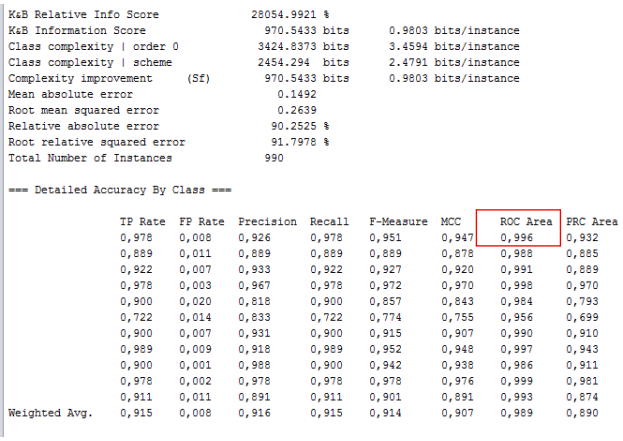


Fig. 3 A Feature selection results in the vowel dataset

Based on Figure 3, the experimental results of the Vowel dataset in attribute selection on the Artificial Bee Colony, in the Vowel dataset there are 14 attributes selected into the remaining 12 attributes.

Then, the calculation results on the Vowel dataset using Artificial Bee Colony and Support Vector Machine (ABC-SVM) are shown in figure 4. In the ROC area there is a Wighted Average value of 0.996 called the AUC (Area Under Curve) value.



K&B Relative Info Score 28054.9921 %
K&B Information Score 970.5433 bits 0.9803 bits/instance
Class complexity | order 0 3424.8373 bits 3.4594 bits/instance
Class complexity | scheme 2454.294 bits 2.4791 bits/instance
Complexity improvement (Sf) 970.5433 bits 0.9803 bits/instance
Mean absolute error 0.1492
Root mean squared error 0.2639
Relative absolute error 90.2525 %
Root relative squared error 91.7978 %
Total Number of Instances 990

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area
	0,978	0,008	0,926	0,978	0,951	0,947	0,996	0,932
	0,889	0,011	0,889	0,889	0,889	0,878	0,988	0,885
	0,922	0,007	0,933	0,922	0,927	0,920	0,991	0,889
	0,978	0,003	0,967	0,978	0,972	0,970	0,998	0,970
	0,900	0,020	0,818	0,900	0,857	0,843	0,984	0,793
	0,722	0,014	0,833	0,722	0,774	0,755	0,956	0,699
	0,900	0,007	0,931	0,900	0,915	0,907	0,990	0,910
	0,989	0,009	0,918	0,989	0,952	0,948	0,997	0,943
	0,900	0,001	0,988	0,900	0,942	0,938	0,986	0,911
	0,978	0,002	0,978	0,978	0,978	0,976	0,999	0,981
	0,911	0,011	0,891	0,911	0,901	0,891	0,993	0,874
Weighted Avg.	0,915	0,008	0,916	0,915	0,914	0,907	0,989	0,890

Fig. 3 Calculation results on the Vowel dataset

For data mining classification, AUC values can be divided into five groups [13]: 0.90-1.00 = very good classification, 0.80-0.90 = good classification, 0.70-0.80 = sufficient classification, 0.60-0.70 = bad classification, and 0.50-0.60 = wrong classification. For the comparison method used Particle Swarm Optimization and Support Vector Machine (PSO-VSM). Figure 5 displays the results of testing the comparison method based on Particle swarm optimization and Support Vector Machine using the Cross-Validation method.

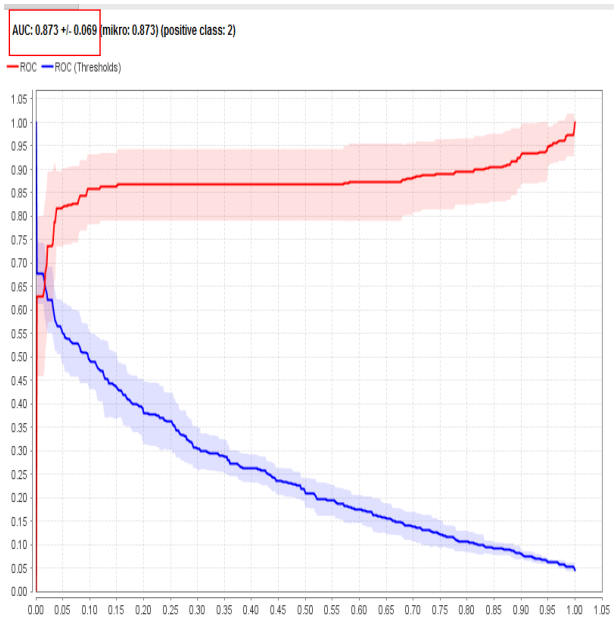


Fig. 5 PSO-SVM Test Results

From the PSO-SVM test results obtained with the AUC 0.873 results and it will be used in this study. Thus, the results of the classification conducted by PSO-SVM with the Vowel dataset have an AUC value between 0.80-0.90. This means that the classification results are good (good classification). When compared with the proposed method, ABC-SVM, the classification results are superior (AUC 0.996). Table 1 summarizes the AUC value from the calculation results of 4 test comparisons between PSO-SVM and ABC-SVM.

Table- I: Comparison of AUC Values

Dataset	Algorithm	
	PSO-SVM	***ABC-SVM***
Vowel	0.873	0.996
Sonar	0.784	0.785
Wavefrom	0.913	0.918
Heart Statlog	0.823	0.847

The results of the performance evaluation of the proposed method are done by calculating the test parameters in the form of precision, recall and f-measure. Table 2 and Figure 4 show the results of precision, recall and f-measure for the PSO-SVM Algorithm. Meanwhile Table 3 and Figure 5 show the results of precision, recall and f-measure for the ABC-SVM Algorithm.

Table- II: Precision, Recall and F-Measure for PSO-SVM

Dataset	PSO-SVM		
	Precision	***Recall***	***F-measure***
Vowel	0.8857	0.8670	0.8762
Sonar	**0.7698**	**0.8411**	**0.8038**
Wavefrom	**0.9217**	**0.8438**	**0.8810**
Heart Statlog	0.8059	0.8868	0.8442

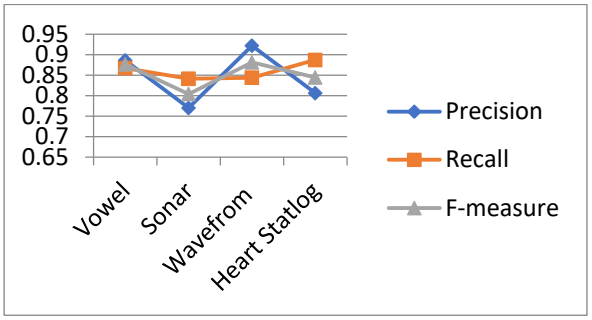


Fig. 4 Precision, Recall and F-Measure for ABC-PSO

Table- III: Precision, Recall and F-Measure for ABC-SVM

Dataset	ABC-SVM		
	Precision	***Recall***	***F-measure***
Vowel	**0.926**	**0.978**	**0.951**
Sonar	0.731	0.732	0.754
Wavefrom	0.880	0.823	0.851
Heart Statlog	**0.848**	**0.893**	**0.870**

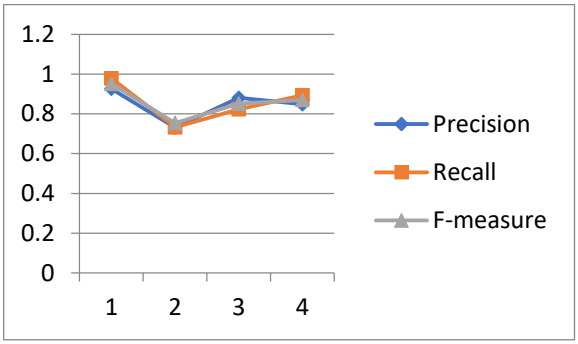


Fig. 5 Precision, Recall and F-Measure for ABC-SVM

Precision, Recall, and F-Measure show the results of the classification efficiency evaluation. The results of the Precision, Recall and F-Measure calculations on the PSO-SVM algorithm are featured in the sonar dataset with an average of 0.7698, 0.8411, 0.8038 and Wavefrom 0.9217, 0.8438, 0.8810 (see Table 2 and Figure 4). This result means that the sonar and wavefrom data set processed by PSO-SVM has quite good classification results. Meanwhile, the test results conducted by ABC-SVM get superior value of the quality efficiency of classifers and quality attributes of 2 datasets, namely vowels with an average of 0.926, 0978, 0.951 and heart statlogs 0.848, 0.893, 0.870 (see Table 3 and Figure 5). This result means that the vowel and heart statlog show quite good results on ABC-SVM.

V. CONCLUSION

Based on the results of experiments and data analysis in this study, it can be obtained the highest Area Under Curve (AUC) value comparison in the ABC-SVM model testing. There are advantages of the proposed method against the PSO-SVM comparison method, after parameter optimization has been performed with the testing phase.

The trial results show ABC-SVM on average has a better performance compared to the PSO-SVM comparison method in the aspects of Area Under Curve (AUC) values and some Precision, Recall and F-Measure values. Feature selection using the Artificial Bee Colony (ABC) algorithm method produces features with better weights which have a significant influence on the classification problem in the Support Vector Machine (SVM) so that better results are also obtained.

In this study Support Vector Machine (SVM) is used because it is known from the results of previous studies that in SVM as a single method has a very good generalization ability to solve problems even with limited samples. From these results it is known that the success of the Support Vector Machine (SVM) is greatly influenced by the selection of appropriate attributes. The more attributes and information used will result in the amount of time and cost being sacrificed, even reducing the higher level of accuracy and complexity.

REFERENCES

1. M. H. Aghdam, N. Ghasem-aghaee, and M. E. Basiri, "Expert Systems with Applications Text feature selection using ant colony optimization," Expert Syst. Appl., vol. 36, no. 3, pp. 6843–6853, 2009.
2. B. M. Zahran, G. Kanaan, and F. Sciences, "Text Feature Selection using Particle Swarm Optimization Algorithm," vol. 7, pp. 69–74, 2009.
3. L. Yonghe, L. Minghui, Y. Zeyuan, and C. Lichao, "Improved particle swarm optimization algorithm and its application in text feature selection," *Appl. Soft Comput. J.*, pp. 1–8, 2015.
4. J. V. E. Frias-Martinez, A. Sanchez, "Support vector machines versus multi-layer perceptrons for efficient off-line signature recognition," vol. 19, pp. 693–704, 2006.
5. J. Ren, "Knowledge-Based Systems ANN vs . SVM : Which one performs better in classification of MCCs in mammogram imaging," *Knowledge-Based Syst.*, vol. 26, pp. 144–153, 2012.
6. T. A. F. Gomes and R. B. C. Prud, "Combining Meta-Learning and Search Techniques to Select Parameters for Support Vector Machines," 2010.
7. I. Ilhan and G. Tezel, "A genetic algorithm – support vector machine method with parameter optimization for selecting the tag SNPs," vol. 46, pp. 328–340, 2013.
8. F. Khoshahval, H. Minuchehr, and A. Zolfaghari, "Performance evaluation of PSO and GA in PWR core loading pattern optimization," *Nucl. Eng. Des.*, vol. 241, no. 3, pp. 799–808, 2011.
9. D. Karaboga and B. Basturk, "Artificial Bee Colony (ABC) Optimization Algorithm for Solving Constrained Optimization," pp. 789–798.
10. L. Rokach and O. Maimon, *Data Mining With Decision Trees - Theory and Applications*. 2015.
11. C. Vercellis, *Business Intelligence: Data Mining and Optimization*[11] *for Decision Making*.
12. C. W. Dawson, "An artificial neural network approach to rainfall-runoff modelling," vol. 43, no. February, pp. 47–66, 1998.
13. F. Gorunescu, *Data mining: concepts and techniques*. 2011.

AUTHORS PROFILE

Mochamad Wahyudi is a Lecturer of Master of Computer Science - Postgraduate Programs, STMIK Nusa Mandiri, Indonesia. (e-mail: wahyudi@bsi.ac.id). Research interest in e-Learning, business intelligence, and Computer Science.

Dr. Sfenrianto, S,Kom, M.Kom is a Faculty Member of the Information Systems Management Department, BINUS Graduate Program – Master of Information Systems Management, Bina Nusantara University, Jakarta 11480, Indonesia. (e-mail: sfenrianto@binus.edu). With lecturing subject: Digital Business and E-Commerce Management. Research interest in Digital Business, e-Commerce, business intelligence, E-Learning and Information System.

Weiskhy Steven Dharmawan is a Lecturer of Master of Computer Science - Postgraduate Programs, STMIK Nusa Mandiri, Indonesia. (e-mail: weiskhy.wvn@bsi.ac.id). Research interest in Information system, e-Commerce, business intelligence, and Computer Science