Student Prediction of Drop Out Using Extreme Learning Machine (ELM) Algorithm

Muhammad Ibnu Sa'ad
Informatics Engineering Department
Amikom University Yogyakarta
Yogyakarta, Indonesia
mibnu336@gmail.com

Kusrini
Informatics Engineering Department
Amikom University Yogyakarta
Yogyakarta, Indonesia
kusrini@amikom.ac.id

M. Syukri Mustafa
Informatics Engineering Department
STMIK Dipanegara Makassar
Makassar, Indonesia
syukri@dipanegara.ac.id

2020 2nd International Conference on Cybernetics and Intelligent System (ICORIS) | 978-1-7281-7257-6/20/$31.00 ©2020 IEEE | DOI: 10.1109/ICORIS50180.2020.9320831

***Abstract*—The purpose of this study was to predict students dropping out of the Education Management Doctoral Program of FKIP Mulawarman University and to evaluate the Extreme Learning Machine in predicting student dropouts. This research uses the Extreme Learning Machine algorithm, the feedforward neural network learning method and the Support Vector Machine algorithm for comparison of the level of accuracy using the same data. The data used is as much as 110 data according to the number of students from the class of 2012 to 2018, the data is taken from the SIA Education Management Study Program of the Mulawarman University Doctoral Program and then processed. In this case, how to predict student dropouts using the variable Gender, Semester 3 IP Value, Working Status, Family Status, Age, and using two DO and NON DO Classes? And calculating the accuracy value using a confusion matrix ?. From the results of this study, it can be concluded that students drop out in the Educational Management of the FKIP Mulawarman University Doctoral Program can be predicted by Extreme Learning Machine using the training value obtained from semester 3 of the 2012-2018 class. From the results of testing the predictive accuracy of the Extreme Learning Machine is 72 %.**

Keywords—prediction drop out, extreme learning machine, rapidminer, neural network, support vector machine

I. INTRODUCTION

Educational Management Doctoral Program of FKIP Mulawarman University is a tertiary institution established in 2012. The beginning of this Study Program has been attended by as many as 20 students, and until now it continues to grow each year until 2019 reaching 120 students.

Based on the academic manual of the FKIP Doctoral Program Management Study Program, students who are eligible to graduate must have a GPA above 3.00, each course must pass with a minimum grade of B and have completed a minimum of 54 SKS including a Dissertation. The limit of the study period set by the campus is 10 Semesters or five years, and includes leave.

Based on data on the Management of Doctoral Education in FKIP UNMUL since its inception until now, there are 48 students who have been declared Drop Out. Therefore, it is necessary to study or predict student drop out so that it can be used as useful information to estimate student dropout rates in the coming years and reduce student dropout rates.

Drop out prediction can be done by a series of processes to gain knowledge or patterns from a collection of data called data mining. Data mining solves problems by analyzing data that already exists in a database. Several data mining classification algorithms have been used to predict student behavior that has the potential to drop out including decision trees, neural networks, naïve bayes, instance-based learning, logistic regression, support vector machines, K-Nearest Neighbor, Rapidminer, and Extreme Learning Machines [1].

Based on previous research conducted by Nurhayati, et al, on the STMIK AMIKOM Yogyakarta by analyzing the support vector machine (SVM) algorithm [1] for predictions

of students dropping out shows that the support vector machine (SVM) algorithm can be used to predict student dropouts using individual data input variables and student IP and GPA evaluations with dropout and non-dropout student output variables.

The method used to predict the research that will be carried out is using Extreme Learning Machine (ELM). This study uses the Extreme Learning Machine algorithm which is a new learning method of feedforward neural networks with single hidden layer feedforward neural networks (SLFNs) [2]. The ELM method has advantages in learning speed and a better level of accuracy compared to conventional methods such as moving averages and exponential smoothing so that by applying ELM that is able to produce more effective dropout student predictions [2]. The method that will be used as a comparison in this study is the Support Vector Machine (SVM) method, where the support vector machine (SVM) algorithm has also been used to predict student dropouts.

II. RELATED WORKS

A. Previous Works

According to Jasmir, et al. [3] in his research "Prediction of Drop Out Students using the Data Mining Classification Algorithm". In that study, researchers conducted research on student predictions of dropouts. Data is taken from semester one to semester six. This data will be used as a basis for calculations in predicting students who are likely to drop out using one of the data mining algorithms, namely k-Nearest Neighbor. The results of this study are the accuracy value of the predictions of students who are likely to be affected by dropouts.

According to the study conducted by Nurhayati, et al, in his research "Prediction of Drop Out Students Using the Support Vector Machine Learning Method" [1]. The research aims to produce classifications to determine whether students have dropped out or non-dropped out using a support vector machine with kernel parameters in one C parameter that provides penalties for randomly classified data points. Based on the test results, to measure the accuracy level obtained an accuracy of 98.06% and an error value of 0.0193 using 310 data sets.

Ayustina Giusti, et al [4], in his research "Prediction of Noodles Sales Using the Extreme Learning Machine (ELM) Method in Kober Mie Setan, Soekarno Hatta Branch". In that research, The amount of consumer demand for the Kober Mie Setan restaurant, the Soekarno Hatta branch, which is erratic every time, influences the remaining raw materials because the raw materials that are stored too long are not good for consumption. When demand is low and the raw materials

provided are high, the remaining raw materials from that day's sales will be discarded. So that raw materials are not wasted, then the sales prediction is needed by Kober Mie Setan, the Soekarno Hatta branch. With these sales predictions, restaurants can prioritize the purchase of certain menu raw materials that have high interest so that the remaining raw materials can be reduced. This study applied the method of Artificial Neural Network (ANN), namely Extreme Learning Machine (ELM) to predict noodle sales at the Kober Mie Setan restaurant, Soekarno Hatta branch. The process of predicting the sale of noodles at Kober Mie Setan is data normalization, training process, testing process, data denormalization, and calculation of error values using Mean Square Error (MSE). The ELM method has advantages in learning speed and a small error rate. Based on testing conducted to determine differences in the use of data features in this study produced the smallest error rate of 0.0171 by using historical data features and residual sales data features.

Dina Maulina and Rofie Sagara [5], in his research "Classification of hoax Articles Using Support Vector Machine Linear With Weighting Term Frequency – Inverse Document Frequency" in that research, The purpose of this study is to detect hoaxes by grouping weighted text, the algorithm used in the text weight classification is the Support Vector Machine (SVM), and for text weighting using the Term Frequency –Inverse Document Frequency (TF-IDF). From the results of research using 108 hoax articles and 132 non-hoax articles, the accuracy level obtained using the Cross Validation calculation with 10 Folds is 95.8333% with the vector owned by the model is 14 supporting vectors.

Lalu Abd Rahman Hakil, et al [6], in his research "K-Nearst-Based Student Graduation Prediction Application Neighbor (KNN)" in that research, The purpose of this study is to predict student graduation, this study uses the K-Nearest Neighbor algorithm to identify student graduation in new cases by adapting solutions from previous cases that have closeness to new cases. Based on tests carried out using K-Fold Cross Validation, the highest accuracy was obtained in the third model of 80% when the 4th K-Fold and 61% when the value of K = 1. While testing using the Confusion Matrix obtained the highest accuracy of 98% at K = 1 for the classification "On time", and 98% at K = 2 for the classification "Not on time".

B. Extreme Learning Machine

Extreme Learning Machine is a new learning method of artificial neural networks. This method was first introduced by Huang G.B in 2004 [7]. ELM is a feedforward artificial neural network with Single Hidden Layer Feedforward Neural Networks (SLFNs).

According to Sun, et al, the level of learning of feed forward neural networks (FFNN) consumes more time [8].

ELM learning methods are made to overcome the weaknesses of artificial neural networks, especially in terms of learning speed. Huang et al stated two reasons why feedforward neural networks have low learning speeds [9], namely:

1. Use a slow gradient based learning algorithm to conduct training.
2. All parameters on the network are determined iteratively using the learning method.

In learning by using Conventional Gradient Based Learning Algorithms such as Backpropagation (BP) and its variant Lavenberg Marquadt (LM) all parameters in feedforward ANN must be determined manually Zhu [10]. The parameters in question are the input weight and hidden bias. These parameters are also interconnected between one layer with another layer so it requires a long learning speed and is often trapped in local minima. Huang et al [7].

Whereas in ELM, the parameters meant are input weight and hidden bias which are chosen randomly, so that ELM has a fast learning speed and is able to produce good generalization performance. The ELM method has a different mathematical model from the feedforward artificial neural network. The mathematical model of ELM is simpler and more effective. Huang et al stated "in theory, this algorithm tends to provide the best generalization performance at a very fast learning speed" [7].

ELM has interesting and significant features, in contrast to popular gradient-based learning algorithms for feed-forward neural networks. The features in question are as follows [11] :

a. The speed of ELM learning is very fast in simulations reported in the literature [11], the ELM learning phase can be completed in seconds for many applications. Previously, there seemed to be a virtual speed barrier where most classical algorithms could not penetrate. It is not uncommon for training feed-forward neural networks that use classical learning algorithms requires considerable time even for simple applications.

b. ELM has better generalization performance compared to gradient-based learning [11], such as backpropagation in most cases. The classical learning algorithm based on gradients and several other learning algorithms face several problems such as local minima, incorrect learning levels, and others. To avoid this problem, several methods such as weight decay and early termination methods are often used in this classic algorithm.

c. ELM tends to achieve simple solutions without such problems. The ELM learning algorithm looks much simpler than most feed-forward neural network learning algorithms. Unlike gradient-based learning algorithms that only work for differentiated activation functions [11], ELM algorithms can be used to train SLFNs with many undifferentiated activation functions.

The general structure of ELM can be seen in Fig. 1. Extreme learning machines run on a single hidden layer feedforward neural network, which is used by the ELM method [5]. In carrying out the training process, ELM uses inverse matrix theory. The inverse matrix theory used by ELM is the Moore-Penrose pseudoinverse theory [11].

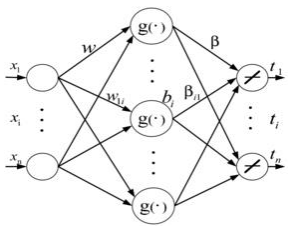

Fig. 1. General structure of ELM

III. METHOD

A. Type of Research

This type of research is quantitative research. Quantitative research is research conducted using statistical formulas and symbols [12]. This study also aims to explain the phenomenon profoundly through data collection. So in this study the emphasis is more on the number (quantity) of data.

B. Data Collection

The data obtained in this study are new student data obtained directly from the academic department data source of the Educational Management of the FKIP Doctoral Program in Mulawarman University, then library research which is secondary data both in the form of data mining books, ebooks, and journals about student predictions drop out.

The data used are the data of new students from 2013 to 2019 in the Academic Education Management Study Program at the FKIP Mulawarman University.

In each data category consisting of several variables or data attributes, there are a total of 6 input variables viz

1). IP Semester 3

2). Gender

3). Working Status

4). Family Status

6). Age

C. Research Design

This research uses data mining standardization model [2] namely CRISP-DM (Cross Industry Standard Process for Data Mining) with the following steps:

- Data Understanding.

  The data used in this study are primary data sources. The data was obtained from the Academic Education Management Department of the FKIP Mulawarman University in the form of data on semester 3 of the Achievement Index, Gender, Working Status, Family Status, and Age. Data is taken from 2012 to 2018.

- Data Preparation.

  The first step is determining the data to be processed. From the data obtained, not all data will be processed because the research conducted has data limitations that will be used. Second stage, the handling of missing value data. Missing Value is incomplete data because the attribute is not recorded or the attribute is not owned. Missing value handling is done by deleting an empty record. Third stage, determine the attributes that will be used from the first stage. Attributes used are semester 3 Grade Achievement scores, Gender, Working Status, Family Status, and Age. The fourth step is to convert data. Data with the selected attributes are then converted to facilitate the data mining process for some of the attributes, because the data will be processed using the Python programming language.

- Modelling.

  The method used in this research is Extreme Learning Machine, to measure accuracy the Python programming language will be used.

- Evaluation.

  At this stage validation and measurement of the accuracy of the results achieved by the model that is Confusion Matrix will be performed.

D. Data Analysis Method

In the Extreme Learning Machine data analysis method the Confusion Matrix is used. According to Sokolova and Marina, the classification method will be evaluated mainly on the accuracy of the classification results [13]. The accuracy of a classification affects the performance of a classification. To do the analysis, a confusion matrix can be used, which is a matrix of predictions that will be compared with the original class of input data. For example, a test (i, j) of a confusion matrix is a percentage of the time of a classifier that identifies input I as the pattern of class j. Each column of the matrix corresponds to the output classifier and each row to the input. The accuracy of a classification where i = j explains the accuracy of the classification in each class.

To show the speed of the learning process and the accuracy of the Extreme Learning Machine, the Extreme Learning Machine (ELM) performance will be compared with the Support Vector Machine (SVM) algorithm on the same data, namely the Educational Management of the FKIP Doctoral Program in Mulawarman University.

E. Research Stages

In Fig. 2, the steps of making a student prediction model drop out using the Extreme Learning Machine method are: data collection, implementation of the Extreme Learning Machine method for prediction through data sharing, Extreme Learning Machine training and testing, and analysis of the predicted results of the Extreme Learning Machine.

In the initial stages of the study begins by determining the background and objectives of the study and defining the scope. Literature study was carried out to deepen understanding of how the Extreme Learning Machine method works as well as the stages needed for predictions using the Extreme Learning Machine method. In addition, a literature study was also conducted to find out which students dropped out so that data collection had a better focus.

The second stage of this research is data collection. The third stage is the implementation of the Extreme Learning Machine method through data sharing, training and testing. The fourth stage is the analysis of predictive results, then the last stage is drawing conclusions.

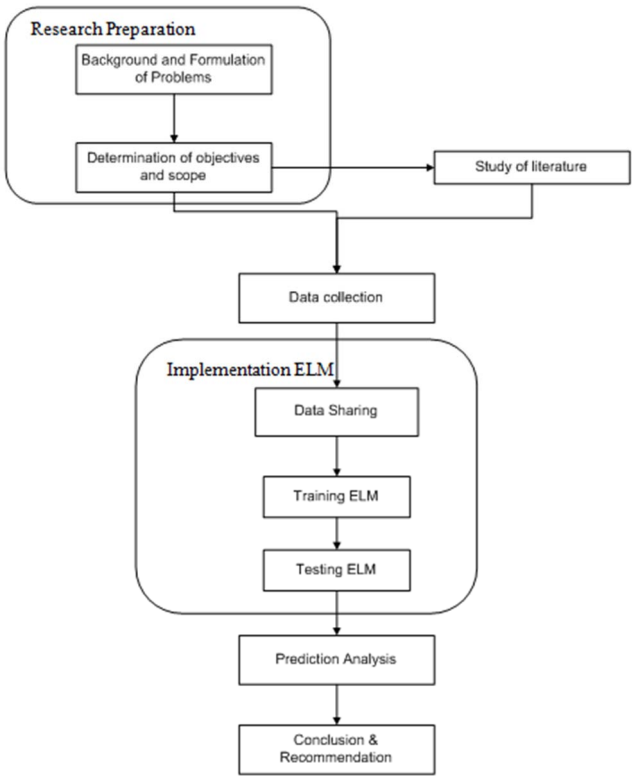

Fig. 2. Research Stages

IV. Result And Discussion

Components in the development of data mining are as follows:

- Operational Database

  Academic operational data relating to student and academic information which is a source of data on data mining variables.

- Data on UNMUL SIA

  In this Mulawarman University Academic Information System (UNMUL SIA) stored data that comes from operational databases, this data is then used in the data mining process.

- Data Mining

  Building a model of data mining Extreme Learning Machine. Then validation is performed to determine which model has the right level of accuracy in predictions.

- Python

  Prediction results can be accessed using the Python programming language.

A. Testing Vector Machine Support

Testing Support Vector Machine divides the data into two parts namely training data by 90% and testing data by 10%. The data used in this study were taken from the data of Educational Management Students of the FKIP Doctoral Program in Mulawarman University as much as 110 Student data from the class of 2012 to 2018.

TABLE I. Student Data For Class of 2012 Until 2018

No	Gender	Working Status	Family Status	Age	Grade Point Average 3	Class
0	Female	Yes	Yes	82	3.43	DO
1	Male	Yes	Yes	37	3.75	NON DO
2	Male	Yes	Yes	50	3.58	DO
3	Male	Yes	Yes	64	3.40	DO
…	…	…	…	…	…	…
107	Male	Yes	Yes	39	4.00	NON DO
108	Male	Yes	Yes	28	4.00	NON DO
109	Male	Yes	Yes	53	0.00	DO
110	Male	Yes	Yes	50	3.00	NON DO

The Vector Machine Support kernel used in this study is the RBF (Radial Base Function) kernel. RBF kernel is a kernel that has good performance with certain parameters that the results of the training did not produce large errors, Sangeetha & Kalpana [15].

Based on the results of testing prediction Students drop out using the RBF kernel on the support vector machine is 0.636364 or 63% with 110 data of Student data.

B. Testing Extreme Learning Machine

Extreme Learning Machine is tested using the sigmoid activation function, because the sigmoid activation function is a function that is often used to predict.

According to research conducted by Cao [14], the sigmoid activation function is the best activation function to be implemented in regression problems.

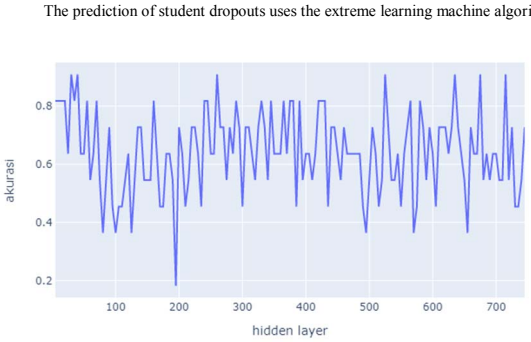

Fig. 3. Hidden layer graph and accuracy

The same data is used in ELM and SVM testing, namely 110 data divided by two parts training data as much as 99 or 90% and testing data as much as 11 or 10%.

Based on the graph shown in Fig. 3, the predicted accuracy of drop out students using Extreme Learning Machine is 0.727273 or 72%. The results and testing can be concluded that the amount of training data affects the value of the resulting error for the amount of training data 90% and

testing data 10%. The higher the amount of training data used, the lower the error rate in the test. Because the Extreme Learning Machine method is a training method. The more training data used, the better the prediction results.

C. Evaluation of the Confusion Matrix

At this stage, testing is done on the system that has been built. System testing is done by calculating the level of accuracy of the system in recognizing attributes based on data that has been created. The steps that must be taken are using a confusion matrix.

To see the most effective network architecture, it can be seen from the existing confusion matrix in the Extreme Learning Machine method and the Support Vector Machine method.

To measure the percentage of accuracy of the results of the introduction of the system for each variation the following formula is used.

$$Accuracy = \frac{the\ number\ of\ students\ is\ predicted\ to\ be\ true\ (DO + NON\ DO)}{the\ total\ number\ of\ students} x100\% \quad (1)$$

Meanwhile, to see the results of sensitivity and specificity using the following formula.

$$Sensitivity = \frac{Positive\ True}{Positive\ True + Negative\ False} x100\% \quad (2)$$

$$Specificity = \frac{Negative\ True}{Positive\ False + Negative\ True} x100\% \quad (3)$$

The following confusion matrix is on the system :

A. Variation 1 Extreme Learning Machine. Data recognition results:

a) True (Data can be recognized correctly) : 8

b) False (Data is recognized incorrectly) : 3

TABLE II. CONFUSION MATRIX EXTREME LEARNING MACHINE

	True	False
True	3	1
False	2	5

Based on the confusion matrix in table 4, then by calculating the accuracy value of what percentage of students predicted DO and NON DO of all students. The blue table column shows the data that is recognized correctly (True). While the white tables describe the number of data that is recognized incorrectly (False).

The Accuracy for Extreme Learning Machine is 72%.

The results of the calculation of the sensitivity and specificity of the Extreme Learning Machine are as follows.

$$Sensitivity = \frac{3}{3+2} x100\% = 60\% \quad (4)$$

$$Specificity = \frac{5}{1+5} x100\% = 83\% \quad (5)$$

B. Variation 1 Support Vector Machine. Data recognition results:

a) True (Data can be recognized correctly) : 7

b) False (Data is recognized incorrectly) : 4

TABLE III. CONFUSION MATRIX SUPPORT VECTOR MACHINE

	True	False
True	1	3
False	1	6

Based on the confusion matrix in table 4, then by calculating the accuracy value of what percentage of students predicted DO and NON DO of all students. The blue table column shows the data that is recognized correctly (True). While the white tables describe the number of data that is recognized incorrectly (False).

The Accuracy for Vector Machine Support is 63%.

The results of the calculation of the sensitivity and specificity of the Support Vector Machine are as follows.

$$Sensitivity = \frac{1}{1+1} x100\% = 50\% \quad (6)$$

$$Specificity = \frac{6}{3+6} x100\% = 66\% \quad (7)$$

V. CONCLUSION

The Extreme Learning Machine algorithm can be used to predict student dropouts by using the variables Gender, Working Status, Family Status, Age, Semester GPA 3.

From the data mining technique that was built, the model that has the best level of accuracy in predicting students dropping out of the Educational Management of the FKIP Doctoral Program at Mulawarman University is the Extreme Learning Machine, this technique is more accurate than the support vector machine method. Based on the evaluation using confusion matrix, Extreme Learning Machine has an accuracy rate of 72%, while Support Vector Machine has an accuracy rate of 63% with the same data in Extreme Learning Machine predictions. Thus Extreme Learning Machine is more accurate than Support Vector Machine in predicting student drop outs.

VI. RECOMMENDATION

Suggestions for further research to measure the level of accuracy of student dropouts in the future, more complete student data can be used in several study programs. Added several data mining algorithms to measure the accuracy of student drop out predictions. Adding a decision support system method Combining variables to see which variable is most influential in predicting student dropouts.

REFERENCES

[1]. N. Siti, Kusrini, L. E. Taufiq, "Prediksi mahasiswa drop out menggunakan metode support vector machine learning," Jurnal Ilmiah Sisfotenika, 2015.

[2]. P. Heny, "Prediksi kinerja ademik mahasiswa baru menggunakan extreme learning machine pada STMIK Widya Cipta Dharma Samarinda, Bina Nusantara University, 2017.

[3]. Jasmir, J. A. Pareza, D. A. Zaenal, E. Rasyeir, "Prediksi mahasiswa drop out dengan menggunakan algoritma klasifikasi data mining,"

[4]. Prosiding Annual Research Seminar, ISBN : 978-979-587-813-1, vol. 4, No. 1, 2018.

[5]. A. Giusti, A. W. Widodo, S. Adinugroho, “Prediksi penjualan mie menggunakan metode *extreme learning machine* (ELM) di Kober Mie Setan cabang Soekarno Hatta,” Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer, j-ptiik.ub.ac.id, e-ISSN: 2548-964X, vol. 2, no. 8, , pp. 2972-2978, Agustus 2018.

[6]. D. Maulina and R. Sagara, "Classification of hoax articles using pupport vector machine linear with weighting term frequency – inverse document frequency," Jurnal Mantik Penusa, vo. 2, no. 1, e-ISSN 2580-9741, p-ISSN 2088-3943, Juni 2018.

[7]. L. A. R. Hakil, et al., "K-nearst-based student graduation prediction application neighbor (KNN)," Jurnal Teknologi Informasi dan Multimedia (JTIM), vol. 1, no.1, pp. 30-36, Mei 2019.

[8]. G. B. Huang, Q.Y. Zhu, and C.K. Siew, "Extreme learning machine: a new learning scheme of feedforward neural networks*,"* Proceedings of IEEE International Joint Conference of Neural Network (IJCNN), vol.2, pp. 985-990, 2004.

[9]. Z.L. Sun, T. M. Choi, K.F. Au, Y. Yu, "Sales forecasting using extreme learning machine with applications in fashion retailing, "Decision Support System 46 (1), pp. 411-419, 2008.

[10]. G.B. Huang, D.H. Wang, and Y. Lan, "Extreme learning machines: A Survey," Springer*,* pp. 107-122, 2011.

[11]. Q. Y. Zhu, A. K. Qin, P. N. Suganthan, G. B. Huang, "Evalutionary extreme learning machine," Pattern Recognition 38 (10), pp. 1759-1763, 2005.

[12]. G. B. Huang, Q. Y. Zhu, and C. K. Siew, "Extreme learning machine: theory and application," Elsevier Science : Neurocomputing *70,* pp. 489-501, 2006.

[13]. Wahidmurni, Pemaparan Metode Penelitian Kuantitatif: UIN Maulana Malik Ibrahim Malang, Jawa Timur, 2017.

[14]. M. Sokolova, "A systematic analysis of performance measures for classification tasks*,*" Information Processing and Management 45 pp. 427-437, 2009.

[15]. W. Cao, et al., "Sone Trick in Parameter Selection for Extreme Learning Machine," IOP Conference Series: Materials Science and Engineering, 2017.

[16]. R. Sageetha and B. Kalpana, "Performance Evaluation of Kernels in Multiclass Support Vector Machine," International Journal of Soft Computing and Engineering (IJSCE), 1(5), 2011