Predicting the Final result of Student National Test with Extreme Learning Machine

Eka Mala Sari Rochman[1], Aeri Rachmad[1], Fitri damayanti[1]

[1]University of Trunojoyo, Madura, Indonesia
email: ekamala.sari@yahoo.com

ARTICLE INFO

ABSTRACT

The level of student achievement is a benchmark of the quality assessment of a school. This student's assessment is based on the national final exam scores every year. When the national exam score increases, it will affect the number of students who will enroll in a school. It affects the number of classes to be opened in the registration of new student candidates. This study aims to predict student achievement based on the value of subjects that become the focus on the final national examination. One method of forecasting in the Artificial Neural Network (ANN) is the Extreme Learning Machine (ELM). The working principle in this method is basically the same as ANN method in general. Namely, there are input layer, hidden layer and output layer. By randomly assigning the input parameters, the ELM generates good generalization performance. By using 20-20-1 network architecture, this research has a result in a small RMSE value of 0.314.

INTRODUCTION

Every year the government always holds the National Examination (UN) for each level of school that produces a final result of the value of Pure Ebtanas (NEM). NEM is one of the educational quality benchmarks for a student to get at school. The better and higher the NEM values which were obtained by a student, it will affect the level of acceptance of new students in the school and vice versa. If the NEM produced is low, it is not close the possibility of decreasing the number of enthusiasts. So indirectly, it can

help the school to promote to the public about the education system that occurred in the school. The prediction of student achievement using the NEM value is a conventional method that cannot be known for certain accuracy [1]. Artificial network is used to predict the student's NEM according to average parameters of the subjects tested in one year. This prediction is used to find the relationship between NEM with the amount of value of each student. So, later it will emerge a pattern where the input pattern of subject value will be able to give predictive output pattern according to data trained.

Many applications are used to support in modeling forecasts such as stock forecasting, weather and so on [2]. Artificial Neural network (ANN) is able to recognize an event based on past data. ANN studied past data so that it can make decisions on data that has not been studied. The Extreme Learning Machine (ELM) is part of ANN which is the Single Hidden Layer of Feedforward Neural Networks (SLFNs) [3]. ELM has advantages in obtaining accuracy as well as good learning speed so that this method is widely used by researchers [4]; because of its universal ability to classify data. The output result is able to approach the optimal completion, and computation time is relatively short. ELM performance can produce good generalizations in many cases and can do training faster than the current popular conventional learning algorithm [5].

In this study predicts the value of student achievement based on the value of local subjects. Because the most important thing in the exam is to know the potential possibility of failure that will occur, then the preparation should start early with more effectively. It takes a prediction in order to achieve a passing rate.

METHODS

The data source in this study was obtained from SMAN 1 Batuan Sumenep. The data used the average value of report cards of class XII students during the last year in the academic year 2014-2015 with a total of 89 data. Therefore in this study, the data is divided into two data training and testing; 62 training data. In November-December, it used data testing with a total of 27 data.

Table 1. The Value of Student Report Cards Subjects in the Tested

National Student Registration Number	Indonesian	Field Of Study		Science	Amount	Average
		English	Math			
9988941814	68	84	72,5	67,5	1265	79,0625
9994795793	76	90	72,5	87,5	1299	81,1875
9995635230	68	82	87,5	50	1288	80,5
9982288653	76	66	67,5	52,5	1276	79,75
9986325331	78	72	42,5	80	1342	83,875
9999094166	68	80	65	77,5	1271	79,4375
9995636705	70	70	65	75	1300	81,25
9983199461	64	66	72,5	70	1270	79,375
9982301152	68	70	75	80	1283	80,1875
9982249129	52	66	72,5	77,5	1266	79,125
9987089144	72	78	80	85	1255	78,4375
9995615613	60	66	70	72,5	1276	79,75
9996000341	68	72	82,5	70	1262	78,875

National Student Registration Number	Indonesian	Field Of Study		Science	Amount	Average
		English	Math			
9986240260	76	72	82,5	77,5	1274	79,625
9995617917	66	76	82,5	80	1295	80,9375
4505810	86	62	65	75	1280	80
9998092773	74	84	85	80	1290	80,625
9994846716	56	86	40	52,5	1278	79,875
9983441278	66	76	70	75	1302	81,375
9988846847	78	70	70	82,5	1256	78,5
9985009579	78	92	80	85	1250	78,125
9999346006	76	72	90	87,5	1261	78,8125
9997365081	72	70	75	70	1267	79,1875

Extreme Learning Machine (ELM)

The form of neural architecture in the human brain is modeled by a computational approach, ANN. ANN is used as an approach to solve various application problems such as pattern recognition, signal processing, process control and time series forecasting. ANN is a collection of interrelated processing elements which is called as units or nerves.

Extreme Learning Machine is a new learning method of artificial neural network. ELM learning method is made to overcome the weakness of ANN that is in learning speed process. In this method, it cannot be determined what the expected results during the learning process. During the learning process, the weight value is arranged in a certain range depending on the given input value [6]. The design of the network architecture can be seen in Figure 1, which is the parameter of the input weight. The bias is chosen randomly which is resulting in good generalization performance and fast learning speed [7]. The learning process in the ELM is divided into three namely preprocessing data, training, and testing.
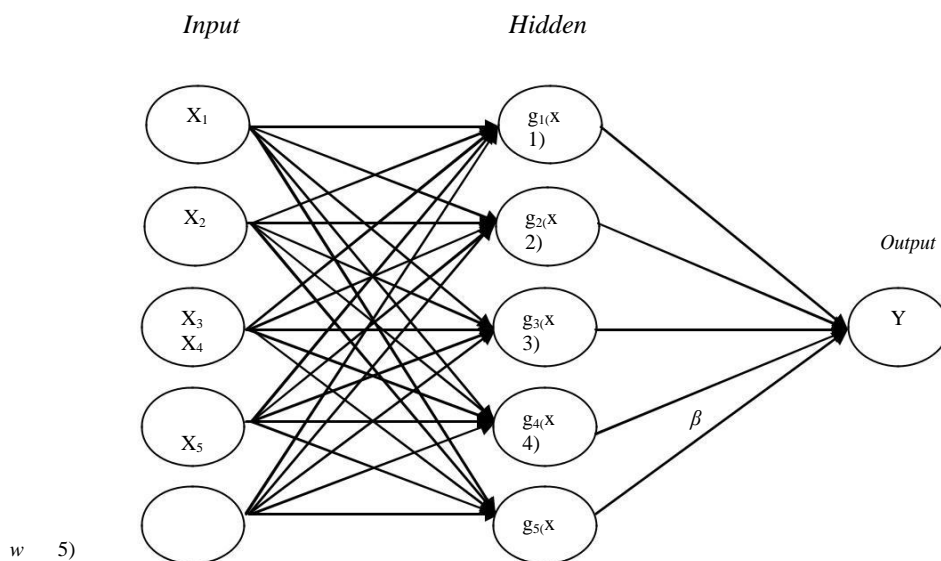
Figure 1. Architecture *ELM*

Changing the value of a data into a value with a certain range is called normalization. Prior to processing, the average data value should be normalized first, because ANN can only accept the range value [0,1] or [-1,1]. The formulation of normalization is shown by equation (1).

$$X = 2 \times (xp - \min\{xp\}) / (\max\{xp\} - \min\{xp\}) - 1 \tag{1}$$

X is the normalized value, whereas xp is the original value, for min (xp) is the minimum value in the data set, whereas max (xp) is the maximum value in the data set.

After the input process is done, the data has been normalized into input value for input neuron, then the next process is training. Which one first determines the number of hidden layers as well as its neurons and activation functions to be used. In this study, sigmoid function used because the data is stationary which many researchers use this function in solving forecasting problems. For the number of neurons used in the hidden layer refers to the study [8] which mentions the number of stable neurons in the range of 0-30. In this study, the architecture used is 20-20-1 i.e. 20 input neurons, 20 neurons in the hidden layer, and 1 output generated.

The testing process is to forecast based on the weight of input and output and bias is derived from the hidden layer by measuring the low error rate using RMSE. The output weight is obtained from the inverse matrix of the hidden layer and output, while the input weight is determined randomly. The equation can be seen in equation (2).

$$\beta = H^T T$$

$$\tag{2}$$

The error rate depends on the learning algorithm, the quality of the data and the type of network used [9]. After doing the forecasting process, the next step is to do the process of denormalization. Denormalization is the process of returning the output result to its true value. While denormalization is shown by equation (3).

$$X = 0.5 \times (xp + 1) \times (\max\{xp\} - \min\{xp\}) + \min\{xp\} \tag{3}$$

Size Errors Forecasting

This research uses Root Mean Square Error (RMSE) calculation in evaluating performance of ELM method. The accuracy of forecasting result that calculates error value from calculation process by system which have been matched with actual data. The smaller the error rate is generated in a forecast then the resulting error rate is smaller.

$$RMSE = \sqrt{\frac{\sum_{t=1}^{n}(X_t - F_t)^2}{n}}$$

$$\tag{4}$$

Xt is the expected value of forecasting with period t. While the value of Ft is the value of forecasting results of the system in the period t squared. Then, the averaging is as much as n the amount of data

RESULTS AND DISCUSSION

Trial at this stage is doing some things that determine the parameters of the forecasting model and the use of different data sets as a comparison. Training and testing ELM use parameters with the number of iterations 1000 times, the learning rate of 0.1. The architecture is used on the input layer nodes, the hidden layer and the output layer are 20-20-1, with the RMSE value of 0.314. The graph of training results can be seen in Figure 2, while the graph of the test results can be seen in Figure.3.

Figure 2 is the training data result which shows changes and shapes that are less stable because of the varying student grades. The data is a preparation of a previously normalized student score with a range of values [0,1]. While in Figure 3, the target output that has resulted from training data process. The RMSE value is obtained based on the comparison of real data with the targeted output.
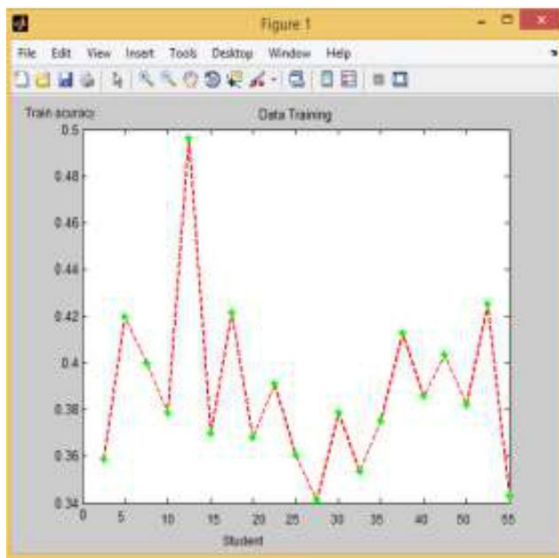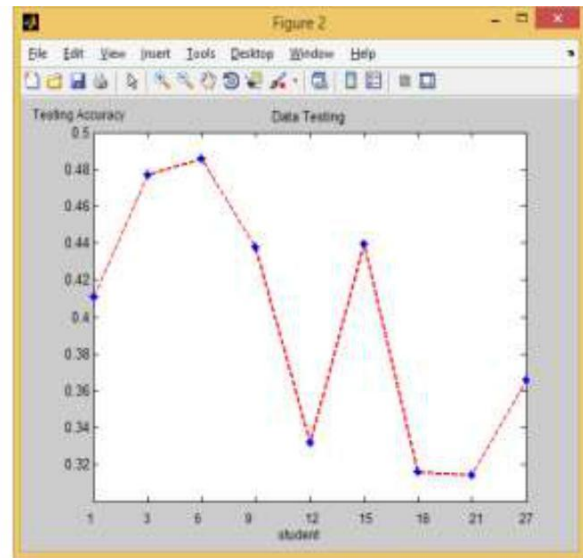

Figure 2. Training Accuracy   Figure 3. Testing Accuracy

CONCLUSION

From the results of trials that have been done using 20-20-1 architecture, it finally gets the following conclusions:

1. ELM produces a low RMSE of 0.341.
2. Learning speed which is required by ELM is very short, which has the average of 0.0312 seconds.
3. The amount of data greatly affects the RMSE value; the more data is processed, then the possibility of RMSE produced can be smaller
4. A suggestion of development of this research, it would be nice to use training data and testing data more than one year. So it is expected that the training process will produce a more stationary and stable value because the processed data more and more varied.

REFERENCES

[1].   J. Lu, J. Huang and a. F. Lu, ―Time Series Prediction Based on Adaptive Weight Online Sequential Extreme Learning Machine,‖ *applied sciences,* vol. 7, no. 217, pp. 1-14, 2017.

[2].   N. Liu, S. Ren, T.-M. Choi, C.-L. Hui and S.-F. Ng, ―Sales Forcasting for FashionRetailing Service Industry: A Review,‖ Hindawi Publishing Corporation Mathematical Problems in Engineering, p. 9, 2013.

[3].   H. Guang-Bin, Z. Hongming, D. Xiaojian and Z. Rui, ―Extreme Learning Machine for Regression and Multiclass Classification,‖ IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS—PART B: CYBERNETICS, pp. 513-528, 2012.

[4].   B. Yadav, Sudheer, S. MATHUR and J. ADAMOWSKI, ―Assessing the suitability of extreme learning machines (ELM)for groundwater level prediction,‖ JOURNAL OF WATER AND LAND DEVELOPMENT, Vols. I-III, no. 32, pp. 103-112, 2017.

[5].   S. Sajjadi, S. Shamshirband, M. Alizamir, P. L. Yee, Z. Mansor, A. A. Manaf, T. A. Altameem and A. Mostafaeipour, ―Extreme learning machine for prediction of heat load in district heating systems,‖ *Energy and Buildings 122 ,* p. 222–227, 2016.

[6].   S. Kusumadewi, Artificial Intellegence, Jogjakarta: Graha Ilmu, 2003.

[7].   Huang, G.B., Zhu, Q.Y., dan Siew, C.K. 2005. Extreme Learning Machine : Theory and applications. Elsevier science : Neurocomputing 70 (2006) 489-501