SCHOOL SUCCESS PREDICTION USING ARTIFICIAL NEURAL NETWORK BASED ON INTERNAL AND EXTERNAL FACTORS

[1]M. MAHAPUTRA HIDAYAT, [2]RATNA NUR TIARA SHANTY

[1]Informatics Department, Faculty of Engineering, Bhayangkara Surabaya University

Jl. A. Yani 114, Surabaya, East Java

[2]Informatics Department, Faculty of Engineering, Dr. Soetomo Surabaya University

Jl. Semolowaru No. 84 Surabaya, East Java

e-mail: mahaputra@ubhara.ac.id

ABSTRACT

In relation with improving the quality of education, various attempts have been made: an increase teachers quality, complete educational facilities, increased allocation of funds for education and educational evaluation of the implementation of sustainable activities. Once observed, it seems clear that the problem is serious in improving the quality of education is the low quality of education at all educational levels, especially schools. Evaluation aims to assess the failure of schools achieving good standards of competence in effort schools improve the quality of education. Levels of school failure to improve the quality of education is influenced by several factors both from the students and of teachers and the school itself. With artificial neural network (ANN), we expect that we will be able to predict school failure which is related to several internal and external factors. So that, we can obtain valid information about some attributes which are affecting to the school failure. Then, some actions can be taken for preventing school failure as the effort to increase those school's educational quality. From the experimental results yield the number of hidden nodes configuration 10, the value of learning rate 0.15, momentum 0.6 and the tolerance value of MSE 0.0013118%.

Keywords: *Prediction, School Failure, Artificial Neural Network, Educational Data Mining*

1. INTRODUCTION

Problem of school failure has been becoming an interesting thing to investigate. Many researchers spawn great deals about the various factors which influence the low performance of students. Generally, school failure can be affected by both internal and external factors. Internal, which is coming from the students. And external, which is coming from the environment around the students.

According to Ministry of National Education, a school can be categorized as quite good, if it has eight criteria. There are: (1) high quality in student selection. Students will be considered based on their academic achievements, psychology test, and physical test, (2) fulfillment facilities for teaching and learning process, (3) ideal atmosphere for learning activity, (4) High rate in teacher professionalism, (5) improvement of curriculum for fulfilling the students who have high motivation in studying, (6) Long-run Student's learning hours, (7) better quality in learning process, (8) have benefit for the environment around that school (Depdikbud, Pengembangan Sekolah Unggul, 1994).

While according to Joan Lipsitz's book, "Successful Schools for Young Adolescent", a school can be categorized as quite good, if it has several criteria, such as: (1) student aspect; quality of graduates haveto be admitted by other institutions, (2) teacher aspect; teacher plans some adequate subjects to learn, adequate teachers to teach, (3) organization activity: schools have national holiday and religious celebration programs, (4) parents accept their children's report of study, well. Parents have choices for sending their children to the nice schools [5].

All of those criteria can be measured by an approach that uses data mining techniques. With artificial neural network, we expect that we will be able to predict school failure, so that we can obtain valid information about some

attributes which are affecting to the school failure. Then, some actions can be taken for preventing school failure as the effort to increase those school's educational quality.

Several studies have done similar topic about predicting school failure prediction. [1] has done on identifying some factors that affect the low performance of student, which affects to the student's failure, at different educational levels. There are examples about how to apply EDM techniques for predicting drop out and school failure [4]. These works have shown promising results with respect to those sociological, economic or educational characteristics that may be more relevant in the prediction of low academic performance. Plenty of examples about how to implement data mining for solving the prediction of school failure, which has been done [7]. The latest research conducted [6] on predicting student performance in a collaborative learning environment to produce output that is valid enough to predict the performance of students in learning.However, few information has been found in the research, such as the use of attributes from the student side only and less specific. Therefore the method we proposed, in addition to the students there is also an assessment of the teacher andthe school itself, which at the end of the expected results are more accurate and actual.

In this paper, we do research for predicting school failure by adding teacher and school variable which affect failure or success. We use Artificial Neural Network as the proper method for predicting, because we believe that method has ability to obtain output based on the result of data training, although we do not have the real data training as the input. System has artificial for determining output value from the inputs which are not included in the input matrix by using prediction based on the weight of each neuron.

2. ARTIFICIAL NEURAL NETWORK

2.1 Neural Network Structure

Artificial neural network (ANN) is a mathematical model to simulate the process of learning and the process of decision making of the human brain. ANN can predict the wide range from a statistical model and flexible in modelling both linear and nonlinear. ANN can be used for multivariate statistical such as multiple regression, discriminant analysis, cluster analysis. In many cases, the result which is obtained using ANN, can be compared with multivariate statistical model.

There are three main kinds of ANN, Multilayer Perceptron, Radial Basis Function, and Kohonen Network. In this paper, we use Multilayer Perceptron as the solution for predicting. Multilayer Perceptron is a model for mapping a set of data input become a set of output, by using nonlinear activation function. In Multilayer Perceptron, both independent and dependent variable have steps of metric and non metric measurement. Multilayer Perceptron is a neural network feedforward, where the information move in one way from input layer through hidden layer and output layer.

2.2 Training Algorithm

Neural network obtains weight value from a specific learning algorithm. This weight will be used for doing value transformation from input to output. Learning algorithm is an adjustment step to the weights that have been formed randomly. Renewal of the weight value is generally defined as follows :

$$w_{ij}(n+1) = w_{ij}(n) + \Delta w_{ij}(n) \quad (1)$$

which is calculated by learning algorithm as the initial weights which are randomly determined at initialization phase.

2.3 Back-Propagation Algorithm

Data input from the input node are forwarded into hidden layer, then forwarded into output node. Each relationship of unit i and unit j has weight wij, which indicate the strength of connection. The weight total, aj, for an input xij and weight wij, are defined as follow :

$$a_j = \sum_{i=1}^{n} w_{ij} x_i \quad (2)$$

where n is numbers of input in a neuron. Activation function which will be used is sigmoid logistic activation function :

$$g(a) = \frac{1}{1+e^{-a}} \quad (3)$$

Error value Ej(n), between actual output Yj(n) and output value of neuron dj(n), will be calculated through formulation :

$$E_j(n) = d_j(n) - y_j(n) \quad (4)$$

with backpropagation learning formula is :

$$\Delta w_{ij} = \eta x_i + \alpha \Delta w_{ij} = \eta x_i - \alpha \frac{\delta E_j}{\delta w_{ij}} \quad (5)$$

where, *n* is learning rate and alpha is moment factor. Those parameter determine how high influence of the old parameter with the changing of new parameter.

3. DATA PREPARATION

In data mining, data preparation consists of 3 points : data selection, preprocessing, and data transformation. The purpose of preprocessing is to transform data sets so that their information content is best exposed to the mining tool in proper format. Data preparation step can take considerable amount of resources such as human, processing time, and cost.

3.1Data Selection

In this data selection process, we will identify all of the internal and external information which are related to this research. In this step, we determine some attributes that will support input variables. The following table shows all variables which are used as input data are : student, school, and teacher.

Table 1.Student Related Variable

Student Variable	**Description**
name	Student's name
class	Student's class
sex	Student's sex (categorical: female or male)
address	Student's home address
distance	The distance between student's living and the school (numeric: kilometers)
livingStatus	Student's living status (binary: with parents or not)
Medu	Mother's education (categorical)
Mjob	Mother's job (categorical)
Fedu	Father's education (categorical)
Fjob	Father's job (categorical)
Famsize	Family size in a house (integer : people)
Famrel	Quality of family relationship (numeric: from 1=bad, to 3=excellent)
reason	Reason to choose their school
studyTime	Weekly study time (numeric: hours)
excul	Extra-curricular activities (binary: yes or no)
schoolsup	Extra educational school support (binary: yes or no)
famsup	Extra educational family support (binary: yes or no)
internetHome	Internet access at home (binary: yes or no)
internetSchool	Internet access at school (binary: yes or no)
handphone	How often the students use their handphone (categorical: from 1=seldom, to 3=often)
higherEducation	Wants to take higher education (binary: yes or no)
Romantic	Have a romantic relationship (binary: yes or no)
Freetime	Free time after school (numeric: from 1=more than 10 hours, to 3=less than 5 hours)
goOut	Going out with friends (numeric: from 1=more than 10 hours, to 3=less than 5 hours)
Health	Current health status (binary: yes or no)
Absences	Number of school absences (numeric:days)
School Variables	**Description**
Accreditation	a type of quality assurance of educational which are evaluated (letter: from A to C or not accredited)
SchoolExp	Average years of School Experience (numeric: years)
Teacher Variables	**Description**

CertificationStatus	(boolean: yes or no)
Experience	(numeric: from 0 to > 10 years)
Age	(numeric: years)

3.2 Preprocessing

We use pre-processing for handling the noisy data and missing value data. By cleaning and integrating, the noisy data and the information which are not relevant from dataset will be reduced.

3.2.1 Data Cleaning

This process is used for removing duplicate data, checking inconsistent data, handling missing value data and smoothing noisy data.

3.2.2 Data Transformation

We use data transformation for transforming data into analytic model. In addition to the transformation of this data is also useful for modeling data to fit the expected analysis and data formats required by the data mining algorithms. When doing data transformation, main data type which will be used, have to be regarded :

- Categorical : all of the possible values are limited. This categorical types such as: (1) nominal, unordered data, such as marital status, or gender. Then the ordinal form, shape size, such as customer loyalty rating.
- Quantitative : all of the possible values which can be measured their differences. This quantitative types such as : (1) Continuous : (real values : monthly hire, average of transaction in certain time period). (2) Discrete : (integer values : numbers of employee, total of transaction in one time period.

One way to transform data is by doing normalization, which can be done by classifying the data to a certain scale.

3.2.3 Data Ranking

In data ranking process, each data is classified into a certain range. We make three range for classifying the data. The following table is one of the student's variable, called distance variable which is classified into three range :

Table 2. Ranking Scale

Skala (km)	Ranking
0-10	3
11-20	2
> 20	1

Table 3. Data Ranking Atribute Distance

Jarak_tinggal	Ranking
13.8	2
2.5	3
7.1	3
11.8	2
6.6	3
32.5	1

The goal of this step is to keep each contain of data uniform, suitable with certain classification scale. So that, we will easy to normalize the data.

3.2.4 Normalization

Before we implement input data and target data into ANN, we must first normalize the data. Normalization is one of some techniques in transformation data, that converts the original data into its data range between 0 and 1 corresponding to the activation function to be used. Its function is for recognizing the data that will become weight input.

$$Normalization = (A - Amin)/(Amax - Amin) \quad (6)$$

where:

A = Attribute Input Value
Amin = Weight Minimum Value
Amax = Weight Maximum Value

4. TRAINING PROCESS

After normalizing the data, we have pattern of data input with the range 0 – 1. It will be trained using a tool from MATLAB software. The data which are used as training data, can be seen in Table 4. The target in Table Training, are obtained from processing student and teacher + school variables. We make priority for each variable : Student = 60% and Teacher+School = 40%.

Table 4. Data Training

No	Student	Teacher+School	Target
1	0	0.75	0.3
2	0.30	0.625	0.43
3	0.8	0.375	0.63
4	0	0.5	0.2
5	0	0.375	0.15
6	0.5	0.5	0.5
7	0.3	0.625	0.43
8	0.3	0.75	0.48
9	0.4	0.75	0.54
10	0.2	0.625	0.37
11	0.3	0.625	0.43
12	0.9	0.75	0.84
13	0.8	0.625	0.73
14	0.8	0.5	0.68
15	0.7	0.625	0.67
16	0.1	0.5	0.26
17	0.5	0.875	0.65
18	0.3	0.875	0.53
19	0.4	0.875	0.59
20	0.8	0.5	0.68
21	0.6	0.625	0.61
22	0.5	0.5	0.5
23	0.5	0.625	0.55
24	0.7	0.5	0.62
25	0.7	0.5	0.62
26	0.6	0.5	0.56
27	0.3	0.25	0.28
28	0.3	0.75	0.48
29	0.5	0.25	0.4
30	0.7	0.375	0.57

4.1 Testing

In this step, we do data testing on some data samples. The following Table 5 shows the result:

Table 5. Data Testing Result

No	Student	Teacher+School	Target	Output	Target Differences	Correction
1	0.7	0.5	0.62	0.624281	0.004	✓
2	0.3	0.5	0.38	0.378019	0.002	✓
3	0	0.75	0.3	0.299968	0.000	✓
4	0.1	0.5	0.26	0.263434	0.003	✓
5	0.6	0.75	0.66	0.673511	0.014	✓
6	0.4	0.625	0.49	0.5048	0.015	✓
7	0.1	0.5	0.26	0.263434	0.003	✓
8	0.8	0.625	0.73	0.732036	0.002	✓

9	0	0.875	0.35	0.376049	0.026	-
10	0.2	0.75	0.42	0.420534	0.001	✓
11	0.5	0.75	0.6	0.598322	0.002	✓
12	0.6	0.75	0.66	0.673511	0.014	✓
13	1	0.5	0.8	0.828784	0.029	-
14	0.4	0.75	0.54	0.541859	0.002	✓
15	0.6	0.625	0.61	0.607155	0.003	✓
16	0.6	1	0.76	0.754515	0.005	✓
17	0.5	0.75	0.6	0.598322	0.002	✓
18	0.6	0.75	0.66	0.673511	0.014	✓
19	1	0.625	0.85	0.832349	0.018	-
20	0.5	0.625	0.55	0.551545	0.002	✓
21	0.5	0.75	0.6	0.598322	0.002	✓
22	0.5	0.75	0.6	0.598322	0.002	✓
23	0.6	0.5	0.56	0.586145	0.026	-
24	0.7	0.625	0.67	0.646407	0.024	-
25	0.4	0.25	0.34	0.340711	0.001	✓
26	0.5	0.625	0.55	0.551545	0.002	✓
27	0	0.5	0.2	0.197477	0.003	✓
28	0.6	0.875	0.71	0.733332	0.023	-
29	0.6	0.75	0.66	0.673511	0.014	✓
30	0.7	1	0.82	0.825443	0.005	✓

5. PERFORMANCE EVALUATION

5.1 Mean Square Error

In this step, we use Mean Squared Error (MSE) for assessing the ANN modeling. MSE of an estimator is an expected value from quadratic error. Mean square total error shows the difference between the model output and the output process. The smaller the MSE value (close to zero value), the greater the success rate of training, whereas the greater the MSE value, the smaller the training success rate. Its value equation can be written as follows :

$$MSE = \frac{1}{N}\sum_{t=h}^{N}(y_t - \overline{y_t})^2 \quad (7)$$

where:

MSE = Mean Squared Error
N = numbers of sample
y_t = Index Actual Value
$\overline{y_t}$ = Index Prediction Value

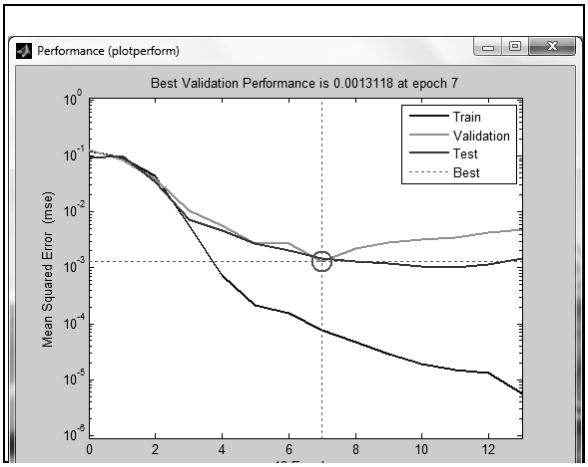

Figure 1.Means Square Error

Based on Figure 1 above, shows that best validation condition performance is 0.0013118 at epoch 7.

6. CONCLUSION

We implement Artificial Neural Network as a method for predicting school success. From comparing the prediction result with the target value which have been assumed, we obatain some conclusions :

a. An approach using ANN, can be used as the solution for predicting school success rate from a school in Indonesia.
b. While implementing, the output is a value of school success prediction, which is currently derived from the attributes of student, school and teacher.
c. Configuration parameters for the training of the prediction system using ANN are best obtained from the experimental results yield the number of hidden nodes configuration 10, the value of learning rate 0.15, momentum 0.6 and the tolerance value of MSE 0.0013118%.

REFRENCES

[1] Araque F., Roldan C., Salguero A. (2009). *Factors Influencing University Drop Out Rates*. Computers & Education, 53, 563–574.

[2] Croninger, Robert., King Rice, Jennifer., Rathbun, Am., Nishio, Masako. (2007). *Teacher qualifications and early learning: Effects of certification, degree, and experience on first-grade student achievement*. Journal of Sciencedirect, 26, 312-324.

[3] H. Trias Rahmadya and W. Pudjo Prabowo, (2009). *Penerapan Soft Computing Dengan Matlab*. Rekayasa Sains.

[4] Kotsiantis S. (2009). *Educational Data Mining: A Case Study for Predicting Dropout – Prone Students*. Int. J. Knowledge Engineering and Soft Data Paradigms, 1(2), 101–111.

[5] Lipsitz, J. (1984). *Successful schools for young adolescents*. New Brunswick, NJ: Transaction.

[6] Olsen J., Aleven V. and Rummel N. (2015). *Predicting Student Performance In a Collaborative Learning Environment.*Proceedings of the 8th International Conference on Data Mining, 211-217.

[7] Vera, C. Márquez., Romero, C., Ventura, S. (2011).*Predicting School Failure Using Data Mining*. Proceedings of the Fourth International Conference on Data Mining, 271-275.