Research Article

Neural Network Model for Mathematic Scores Prediction: Case Study in SMK Negeri Pakis Aji, Jepara, Indonesia

Adi Sucipto[1,*], Joko Minardi[2]

[1] Program Study Teknik Informatika, FST, Universitas Islam Nahdlatul Ulama Jepara
[2] Program Study Sistem Informasi, FST, Universitas Islam Nahdlatul Ulama Jepara
[*] Corresponding author: adisucipto@unisnu.ac.id

Abstract: Aim of this research is to apply Neural Network Algorithm to predict score of mathematic in the national exam. During the time, the teacher only provided national exam materials and additional tryout tests without knowing how to predict the exam scores in mathematics subject. Data mining neural network algorithm obtained \Root Mean Square Error (RMSE) values which were used as basic improvement and clustering class By conducting research using data mining neural network algorithm, it proved that this model can be used to predict scores of Mathematics subject at SMK Negeri 1 Pakis Aji.. The result of this research by using data mining neural network algorithm found RMSE 0138 +/- 0.092. The lower the RMSE values the more accurate the neural network to predict mathematics scores of SMK Negeri 1 Pakis Aji.

Keywords: data mining, mathematics score, neural network, prediction, RMSE

Introduction

Mathematics scores in vocational schools have never been carried out by master teachers so far the teacher only observes according to the results of daily learning, mid-semester and end of semester, so the teacher cannot predict how graduation level from the math score in the 12th National Examination. The results of this prediction are very important to be able to be known early so that the teachers can anticipate the final grade of 12th grade students not to fail Samsuri (2013) [1]. Neural Network (NN) The algorithm based on Rapid Miner's form plate is able to provide an answer as the aim of this research is Larose (2005) [2]. using NN using Multiplayer Perception Approach (MLA) and Linear Discriminant Analysis (LDA) models, Quadratic Discriminant Analysis (QDA), and Logistic Regression (LR) to predict how much credit will not be distributed to financial institutions Ian H Witten (2011) [3] uses NN with hybrids on Central Banks to predict credit allocations to consumers so that credit risk can be predicted early. These four algorithms are usually only used to predict banking customers with very large amounts of data from a country, until now we have not found an algorithm that is used to predict the value of mathematics lessons in a school and the results we obtained turned out to be able to predict math scores before exam [4].

The purpose of this study was to predict the mathematical grades of 12th grade students before taking the national examination. The importance of the results of these predictions is expected that teachers can anticipate preparing mathematics learning since students are in grades 10, 11 and 12 so that students feel ready to take the national exam later [5].

The main problem that has been faced is that there is no way to predict the score of National exam in mathematics subject experienced by students of SMK Negeri 1 Pakis Aji. One way to predict the score of national exam in math subject is by using Data mining neural network algorithm which can make

predictions of national exam results in mathematics subject in SMK Negeri 1 Pakis Aji as a basis to perform further necessary actions.

Benefit using Data mining neural network algorithm, the researchers hope to solve the existing problem that there is no prediction of the students' scores in mathematics subject; hence, this prediction is expected to be earlier overcoming the students who get the low score based on the prediction before the national examination is conducted. We has been able how accurate is Data mining neural network algorithm in Rapid Miner series 7.2 can predict the students' scores in mathematics subject of the national [6].

Initial data collection began by collecting data in the form of students' scores, obtained from the Vice Principal of curriculum section of SMK Negeri 1 Pakis Aji, in the form of 1, 2, 3, 4 and 5, the prospective of national exam (UN) participants in year 2013-2017 for Mathematics subject, which are used as a basis for predicting the scores of National Exam [7].

Observation

Observation is one of data collection techniques which not only measures the attitudes of respondents (interviews and questionnaires) but also can be used to record various phenomena that occur (situation, condition). This technique is used when a research is aimed at studying human behavior, work processes, natural phenomena and performed to not too large number of respondents. In this observation, researchers directly got involved in the daily activities of the objects or situations observed as the data source [8].

This section describes the initial processing (pre-processing) of the data, comprise: first, sorting the data by checking it accurately whether it is already in numerical form, making sure that the data of students' scores have been in numerical form. Second, the outlier data should be removed or discarded because it will interfere with the research process. The purpose of pre-processing is to eliminate Missing value [9].

Materials and Methods

Methods

Figure 1. This stage describes the method used for comparing data mining classification algorithm. The process is done gradually, started from data processing pre-processing data namely integration, selection and cleansing. Furthermore, doing comparison to classification model used, that is the data mining algorithm *Neural Network* [10].
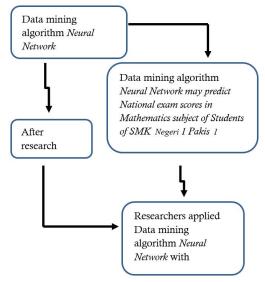

Figure 1. Research Method

Experiment and Testing Method

Research variable; Variable used in this study consisted of 5 input variables and 1 target variable, namely: X = Marks 1st semester report, X2 = Marks 2nd semester report, X3 = Marks 3rd semester report, X4 = Marks 4th semester report, X5 = Marks 5th semester report, and as the target is Y = Scores of National Exam. There were several stages to obtain the best configuration performance of Neural Network that was by determining the *neural network* parameters, namely: *neural network* configuration performance model.

Evaluation and Result Validation

The model proposed in this study has been tested using *confusion matrix* to determine the accuracy level. *Confusion matrix* will describe accuracy result, from correct positive predictions, incorrect positive predictions, correct negative predictions, and incorrect negative predictions. The accuracy has been calculated from all the correct predictions (both positive and negative predictions) divided to all data testing. The higher the accuracy value, the better the model created.

The testing was measured by using RMSE (Root Mean Square Error). The RMSE would describe the positive class in curve form. Testing was done by calculating the value of RMSE (Root Mean Square Error), the lower the RMSE value in *Linear Regression*, the better the classification model formed [5].

Results and Discussions

The initial data of this research used mathematics scores of Students of SMK Negeri 1 Pakis Aji Jepara which had been in the form of data including students' registration number (NIS), students' name and their mathematics report marks from 1st to 5th semester, the average scores of daily tasks from 1st to 5th semester. This study used the data of mathematics scores of SMK Negeri 1 Pakis Aji Jepara in 2013 until 2017 with total of 620 students. (Initial data attached).

Parameter is one of the accuracy determinants. Therefore, the attributes used in this study come to 17 and the data used was the mathematics scores of students of SMK Negeri 1 Pakis Aji Jepara, with Data mining neural network algorithm with the amount of 620 students [11].

In the early experiment, the researchers did parameter trial using neural network algorithm. There were 7 parameters used, including: national students registration number (NISN), students' name, report marks of 1st semester, report marks of 2nd semester, report marks of 3rd semester, report marks of 4th semester, and report marks of 5th semester. The experiment resulted RMSE 4.013 +/- 0.705. The following is the table Performance Vector.

Table 1. Performance vector using 7 parameters

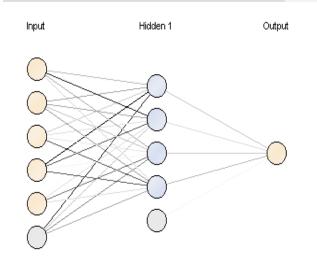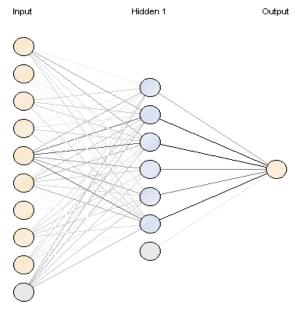| No.	Total Data	RMSE
1.	417	4.787+/-0.834
2.	620	4.013+/-0.705

Figure 2. Improved *Neural Network*

Figure 2. is a node-shaped graph connected together like a neural network from rules result the researchers have got. Then from the first experiment, the neural network algorithm used was optimized by using more parameters, that are 12 parameters, include: NISN, students' name, report marks of 1st semester, report marks of 2nd semester, report marks of 3rd semester, report marks of 4th semester, report marks of 5th semester, daily tasks of 1st semester, daily tasks of 2nd semester, daily tasks of 3rd semester, daily tasks of 4th semester, daily tasks of 5th semester. The data analysis delivered RMSE 0.138 +/- 0.092. Table 2: The following is the table Performance Vector.

Table 2. Performance vector using 12 parameters

No.	Total Data	RMSE
1.	417	0.167 +/- 0.121
2.	620	0.138 +/- 0.092

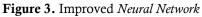**Figure 3.** Improved *Neural Network*

Figure 3. is a node-shaped graph which is connected together like a neural network from rules result the researchers found out. After the first and second experiments with parameters which corresponded to Wicaksana [6], one of respondents in this research, then the third experiment using more parameters, are 17 parameters, include: NISN, students' name, report marks of 1st semester, report marks of 2nd semester, report marks of 3rd semester, report marks of 4th semester, report marks of 5th semester, daily tasks of 1st semester, daily tasks of 2nd semester, daily tasks of 3rd semester, daily tasks of 4th semester, daily tasks of 5th semester, average score of tests of 1st semester, average score of tests of 2nd semester, average score of tests of 3rd semester, average score of tests of 4th semester, average score of tests of 5th semester. From the third experiment for neural network algorithm delivered RMSE0.072 +/- 0.032. Table 3: The following is the table Performance Vector.

Table 3. Performance vector using 17 parameters

No.	Total Data	RMSE
1.	417	0.224 +/- 0.140
2.	620	0.261 +/- 0.127

Figure 4. Improved Neural Net (*Neural Network*)

The figure 4. above is a node-shaped graph connected together like a neural network from rules result the researchers found out.

RSME Calculation

Here are the results of RMSE calculation using RapidMiner and Microsoft Excel.
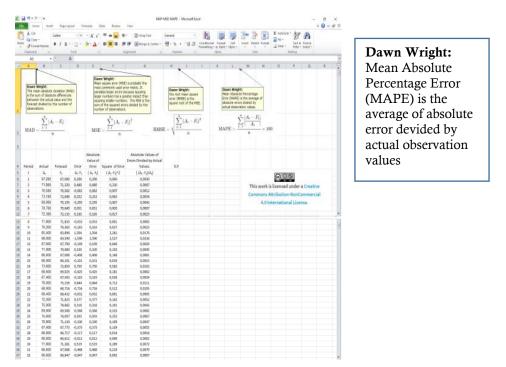

Dawn Wright:
Mean Absolute Percentage Error (MAPE) is the average of absolute error devided by actual observation values

Figure 5. Result of RMSE Calculation using RapidMiner

Dawn Wright

The mean absolute deviation (MAD) is the sum of absolute differences between the actual value and the forecast divided by the number of observation, with mathematical formulas

$$MAD = \frac{\sum_{t=1}^{n}|A_t - F_t|}{n}$$

Dawn Wright

Mean square error (MSE) is probably the most commonly used error metric. It penalizes langer errors because squaring langer numbers has a greater impact than squaring smaller numbers. The MSE is the sum of the squared errors devided by number of observations.

$$MSE = \frac{\sum_{t=1}^{n}(A_t - F_t)^2}{n}$$

Dawn Wright

The not mean square error (RMSE) is the square root of the MSE

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

Dawn Wright

Mean Absolute Persentage Error (MAPE) is the everage of absolute error by actual observation value

$$MSE = \frac{\sum_{t=1}^{n}(A_t - F_t)^2}{n}$$

Performance Vector

Root_mean_squared_error: 0.532 +/= 0.416 (mikro: 0.696 +/= 0.000)
Aquared_error: 0.456 +/= 0.844 (mikro 0.484 +/= 1.848

Performance Vector above is the result of RMSE calculation using application rapidminer 7.2 with 12 parameters, and with total data 36 found RMSE

Figure 5. is the result of RMSE calculation using Microsoft Excel with 12 parameters, and with total data 36 delivered RMSE 0.528 +/- 0.412.

Testing method using RapidMiner obtained is presented Table 4.

Table 4. Experiment result

Experiment	Total Parameter	Total Data	RMSE
1	7	417	4.787 +/- 0.834
	7	620	4.013 +/-0.705
2	12	417	0.167 +/-0.121
	12	620	0.138 +/-0.092
3	17	417	0.224 +/-0.140
	17	620	0.261 +/-0.127

In the first experiment applied 7 parameters, namely NISN, students' name, report marks of 1st semester, report marks of 2nd semester, report marks of 3rd semester, report marks of 4th semester, and report marks of 5th semester, according to neural network algorithm, it delivered high value. It is RMSE 4.013 +/- 0.705.

In the second experiment applied 12 parameters, from previously 7, include NISN, students' name, report marks of 1st semester, report marks of 2nd semester, report marks of 3rd semester, report marks of 4th semester, and report marks of 5th semester, daily tasks scores of 1st semester, daily tasks scores of 2nd semester, daily tasks scores of 3rd semester, daily tasks scores of 4th semester, daily tasks scores of 5th semester, found RMSE 0.138 +/- 0.092. This proved that by adding parameters, the data processing worked better as well as the RMSE number.

In the third experiment applied 17 parameters, from formerly 12 parameters, include NISN, students' name, report marks of 1st semester, report marks of 2nd semester, report marks of 3rd semester, report marks of 4th semester, report marks of 5th semester, daily tasks of 1st daily tasks of semester, daily tasks of 2nd semester, daily tasks of 3rd semester, daily tasks of 4th semester, daily tasks of 5th semester, average score of tests of 1st semester, average score of tests of 2nd semester, average score of tests of 3rd semester, average score of tests of 4th semester, average score of tests of 5th semester, from previously in the second experiment using 12 parameters, include NISN, students' name, report marks of 1st semester, report marks of 2nd semester, report marks of 3rd semester, report marks of 4th semester, and report marks of 5th semester, daily tasks scores of 1st semester, daily tasks scores of 2nd semester, daily tasks scores of 3rd

semester, daily tasks scores of 4th semester, daily tasks scores of 5th semester, found RMSE 0.261 + / - 0.127. This proved that by adding parameters do not always produce a good RMSE value.

In the first and second experiments, the accuracy value of neural network by adding parameters was better with RMSE4.013 +/- 0.705 become 0.138 +/- 0.092, however for the accuracy of neural network in the third experiment with 17 parameters including NISN, students' name, report marks of 1st semester, report marks of 2nd semester, report marks of 3rd semester, report marks of 4th semester, report marks of 5th semester, daily tasks of 1st daily tasks of semester, daily tasks of 2nd semester, daily tasks of 3rd semester, daily tasks of 4th semester, daily tasks of 5th semester, average score of tests of 1st semester, average score of tests of 2nd semester, average score of tests of 3rd semester, average score of tests of 4th semester, average score of tests of 5th semester, previously in the second experiment using 12 parameters, including NISN, students' name, report marks of 1st semester, report marks of 2nd semester, report marks of 3rd semester, report marks of 4th semester, and report marks of 5th semester, daily tasks scores of 1st semester, daily tasks scores of 2nd semester, daily tasks scores of 3rd semester, daily tasks scores of 4th semester, daily tasks scores of 5th semester, found decrease value of RMSE from 0,138 +/- 0,092 became 0,261 +/- 0.127.

Conclusion

The results were varied and difficult to predict. During this time, the teacher only provided national exam materials and additional tryout tests without knowing how to predict the scores of national exam of mathematics subject. By conducting research using data mining neural network algorithm, it proved that this model can be used to predict scores of Mathematics subject at SMK Negeri 1 Pakis Aji. Data mining neural network algorithm obtained RMSE (Root Mean Square Error) value which was used as basic improvement and clustering class. Experimental approach using data mining Neural Network algorithm properly determined Root Mean Square Error (RMSE). The result found the smallest value of RMSE was 0.138 +/- 0.092 with validation number was 5, hidden layer was 3, learning rate was 0.2, momentum was 0.1, training of cycles was 500.

Acknowledgement

References

[1] Samsuri, Prediksi Nilai Ujian Nasional Pada Sekolah Menengah Kejuruan Dengan Metode Neural Network. Teknik Komputer Dan Jaringan(Tkj) Sekolah Menengah Kejuruan (SMKN 3) Palu, Dinas Pendidikan Palu (2013).
[2] D.T. Larose, Discovering Knowledge in Data: An Introduction to Data Mining, Wiley2005.
[3] I.H. Witten, E. Frank, Data mining: practical machine learning tools and techniques with Java implementations, Acm Sigmod Record 31(1) (2002) 76-77.
[4] J.S.D. Raharjo, Model Artificial neural network berbasis Particle swarm optimization untuk Prediksi laju inflasi, Jurnal Sistem Komputer 3(1) (2013) 10-21.
[5] D.K.D. Susanto, M. Bettiza, N. Nikentari, Prediksi Nilai Uas Siswa Smk Menggunakan Algoritma Levenberg-Marquardt, (2015).
[6] D. Aprilla, D.A. Baskoro, L. Ambarwati, I.W.S. Wicaksana, Belajar data mining dengan rapidminer, Jakarta: Gramedia Pustaka Utama (2013).
[7] W. Utomo, Prediksi nilai ujian nasional produktif sekolah menengah kejuruan menggunakan metode Neural Network, Techno. Com 14(1) (2015) 33-41.
[8] Sugiyono, Metode penelitian pendidikan:(pendekatan kuantitatif, kualitatif dan R & D), Alfabeta2008.
[9] P.V.V. Theja, L. Vanajakshi, Short term prediction of traffic parameters using support vector machines technique, IEEE, 2010, pp. 70-75.
[10] X. Wu, V. Kumar, The top ten algorithms in data mining, CRC press2009.
[11] E. Prasetyo, Data Mining konsep dan Aplikasi menggunakan MATLAB, Yogyakarta: Andi (2012).