ATLANTIS PRESS
Advances in Social Science, Education and Humanities Research, volume 242
2nd International Conference on Vocational Education and Training (ICOVET 2018)

Identify Flood Disaster and Mitigation Using Neural Network Learning Vector Quantization in Malang City

Muhammad Ashar [1]
State University of Malang
Semarang st. 5, Malang,
East Java, Indonesia

Kartika Devi Suraningtyas [2]
State University of Malang
Semarang st. 5, Malang,
East Java, Indonesia

I Made Wirawan [3]
State University of Malang
Semarang st. 5, Malang,
East Java, Indonesia

Farhan Afzal [4]
State University of Malang
Semarang st. 5, Malang,
East Java, Indonesia

Abstract; Flood is the most common disaster in Indonesia and certainly harmful to society in the form of material or psychical. Therefore, it's necessary to identify the potential and flood mitigation earlier to reduce the potential losses suffered by the society after the occurrence of disaster. This is difficult to do with conventional methods so that in this research proposed "Neural Network Learning Vector Quantization as Identification Method of Potential and Mitigation of Flood Disaster". With this algorithm specified four nodes input layer, one hiden layer with two neurons and two output layers where four nodes input layer are elevation, drainage, rainfall and flood events are derived from data of BPS Malang, BMKG Karangploso, and data of BPBN. Data processing and testing will generate two outputs, they are identification of flooding potential area and no flooding potential area in every villages in Malang. The test results by using confution matrix showed the accuracy value at 95.34%, sensitivity value at 100%, specification value at 95.29%, and error rate at 4.68% on 1710 dataset that composed of 70% training data and 30% testing data with learning rate at 0.1, decrement learning rate at 0.01, maximum epoch at 10 and minimum epoch at 0.0000001.

I. INTRODUCTION

Flood is at first rank as the most common disaster that occurs in Indonesia with a frequency of 31% from 1815 to 2017 with total occurrence is 969 in 2017 [1-3]. East Java is the most province from 33 provinces that has frequent occurrence of flood with frequency of 36% [3]. In Malang, there are 33 flood prone locations in 2017 in five sub-districts and caused losses in the form of material and psychological [4]. Therefore, need several efforts to reduce the losses such as prevention, preparedness, early warning, and mitigation [5-6]. Several studies have developed an algorithm to handle such problems as research on the Citarum River flood prediction by using fuzzy methods which is the results of particle swam optimization algorithm that produces output with an accuracy rate of 73% [7]. Furthermore, research on the design of the potential for flooding using Sugeno fuzzy logic zero node has the disadvantage of requiring a lot of combinations when the parameters used much so inefficient and wasteful of resources [8]. Further research on the prediction using the C4.5 method has an advantage in determining the pattern but in processing large data, required the training data that has been processed in advance so that takes time and huge resources [9]. From previous studies can be understood that research on flood mitigation can be developed.

Neural network has advantages in terms of predictions based on a time series, high tolerance of noise, can be trained into large database, has an accuracy that approaching actual data. However Neural network has the disadvantage in limitations for the training data pattern of the coating, numeric operation, logic, arithmetic, symbolic, and the accuracy rate are determined by the amount of training so it can be wasteful of time in the training process if the data that are processed is large [10-12] , This problem can be solved by providing additional policy by providing learning vector quantization training algorithms [11, 13]. Therefore, the identification of potential and mitigation of flood is accomplished by using neural network learning vector quantization method.

A. Previous Research

The previous research that had been conducted on the design model of the potential for flood in an artery in Malang using fuzzy logic which aimed to calculate the flood potential with the outcome in the status form such as flooding status, no flooding status, and potentially to be flooding but didn't have calculation error value for the result that achieved, in addition, the research result had not been able to be applied to determine the potential for flooding in real time [8]. The research on the prediction of customer complaints to the apartment using the C4.5 algorithm with an accuracy of 75% [9]. On the rainfall prediction research with fuzzy logic could generate predictions with an accuracy rate of 82.19% [14]. Other studies on the prediction of the availability of rice in the community by using fuzzy logic and neural network in an effort to improve food security predictions with accuracy with an error value of 0.3873 [15]. Then another research on the prediction of total tempuyung flavoid level by using

IR spectroscopy combined with partial least squares regression with the results of the SEC and RMSEC was 0.0.23, RMSEP was 0074, and SEP was 0078 [16]. Further research on the application of neural networks to predict the number of unemployed in the province of East Kalimantan using back propagation learning algorithm with the test results obtained in this research was 139 830 but the accuracy had not been calculated [11]. Based on the previous researches, could be understood that prediction was able to be conducted by using several methods and the prediction method was still able to be developed due to theirs disadvantages..

II. METHODOLOGY

Neural Network was one of the artificial intelligence algorithms that used concepts such as neural network biology in human for clustering and recognizing pattern, thus, information processing occurred in simple elements called neurons via synopsis [10-13]. In determining the results, each neuron required activation function [17]. Neural Network had the advantage of case prediction based on a time series, performing non-linear prediction, performing parallel processing, noise high tolerance, could be trained for large databases, and had an accuracy that approaching the actual data [10-11], This algorithm also had the disadvantage in obtaining high accuracy should require the large training data, so it was wasteful of time and resources, beside of that, this algorithm had weaknesses in the operation of numerical, symbolic, arithmetic, logic and weaknesses in the training pattern of the coating [11-12, 20]. These problems could be solved by providing additional policy by providing learning vector quantization training algorithms [11-13].

Learning Vector Quantization was a training method for large data by classifying on the input by searching the shortest distance to the value and eliminating the noises which could potentially interfere with the process of convergence in the forecasting system. The algorithm testing was conducted by calculating the value of the closest output vector to the value of the input vector [19] [23]. In this research, the inputs vector used were the elevation and drainage of 57 villages in Malang, rainfall in November 2017, and the incidence of flooding in November 2017. The datasets used were obtained by by selecting and integrating the data of elevation and drainage that obtained from the Central Bureau of Statistics Malang, the rainfall data obtained from the Meteorology and Geophysics Karangploso, and flood occurrence data obtained from the National Disaster Management Agency. The stages of training data processing by using neural network learning vector quantization was described as follows:

a. Declaration of alpha (learning rate) had a value of 0.1, decAlfa (deduction learning rate) had a value of 0,001, minAlfa (minimum learning rate) had a value of 0.000000, and maxEpoch or maximum value allowed epoch was 10.
b. In the epoch valued 0, the program initialized the initial value that consisted of two data represented the number of identification of strengths, namely flooding (w1) and no flooding (w2). Each of these data included the input vectors x1, x2, x3, and target.
c. The program read data from the dataset would calculate the eucliedean distance from the first training data to the last training data using Equation 1.
d. $$d(x, w_j) = \sqrt{(x_1 - w_{1j})^2 + (x_2 - w_{2j})^2 + (x_3 - w_{3j})^2} \quad (1)$$
e. Specifying the output class category (cj) based on the results of the calculation of the smallest eucliedean distance.
f. Calculating the change in the values that would be used in the calculation of the next epoch. If the value of cj was not equal to the target value then updated the value of the values with Equation 2, if the value of cj equal to the target value then the updated values with Equation 3.
g. $$(w_{update}) = \sqrt{w_j(old) + \alpha(x - w_j(old)} \quad (2)$$
h. $$(w_{update}) = \sqrt{w_j(old) - \alpha(x - w_j(old)} \quad (3)$$
i. Reducing the value of alpha (learning rate), Epoch + 1, repeating the process c to minimal epoch or minalfa (learning rate).

III. RESULT AND DISCUSSION

Testing was conducted to determine the performance of neural network learning vector quantization method in identifying the potential flood that could be understand its efficiency. Therefore, the analysis of the results of research was conducted by doing gradually test to used variable, comparison of the testing data and training data, learning rate and decrease learning rate (decAlfa) using the Confusion Matrix. Each of the most optimal test results would be used as parameters in subsequent testing.

The test results were presented in Figure 2 and 7. Parameters that were used as a reference test that was learning rate of 0.1, deduction learning rate of 0.01, the minimum learning rate of 0.0000001, and maxEpoch of 10 at 70% of training data and 30% of testing data.

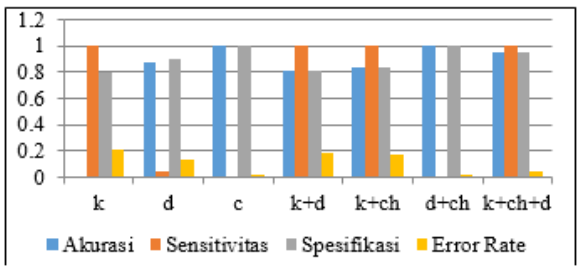


Figure 1. Comparison Graph of Test Result of Affected Variable on Algorithm Performance

Based on the results of testing the affected variable on the algorithm performance in Figure 1, explained that the most optimal performance results occurred in the composition of variable elevation (k), drainage (d), and rainfall (ch) with accuracy of 95.32%, sensitivity of 100% , specificity of 95.29% and error rate of 4.68%. While the results of the performance of the composition of the other variables were considered less than optimal because there

was an imbalance between the accuracy, sensitivity, specificity, and error rate.

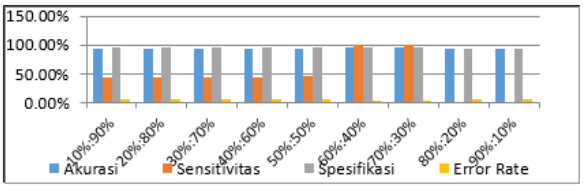


Figure 2. Evaluation Result of The Comparison Effect of Training Data and Testing Data Against The Algorithm Performance

Based on the results of comparative evaluation of the effect training data and testing data on the algorithm performance in Figure 2, the maximum performance of the algorithm was obtained at ratio of 70% of testing data and 30% of training data or 1197:513 with a value of 95.32% of accuracy, sensitivity of 100%, specificity of 95.29%, and 4.67% of error rate.

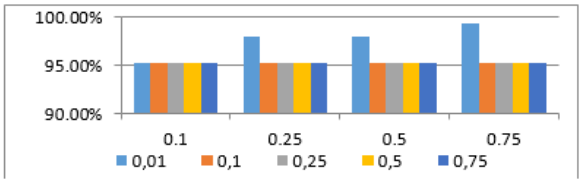


Figure 3. Evaluation Results of The Comparison Effect of Training Data and Testing Data Against Accuracy

Based on the Figure 3 there was an increment accuracy value on the data with a greater learning rate and deduction rate learning with the value of 0.01. While the data with the learning rate of 0.25, up to 0.75 and deduction learning rate was greater, the accuracy decreased and reached the value of convergent value that was 95.32%.

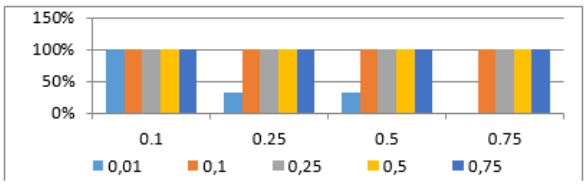


Figure 4. Evaluation Result of The Comparison Effect of Training Data and Testing Data Against Sensitivity

Based on the Figure 4 there was a significant decrement sensitivity value on the data with greater learning rate and deduction learning rate with a value of 0.01. While the data with the learning rate of 0.25, to 0.75 and a greater deduction learning rate, sensitivity values tended to increase and reached a value converges on a value of 100%.

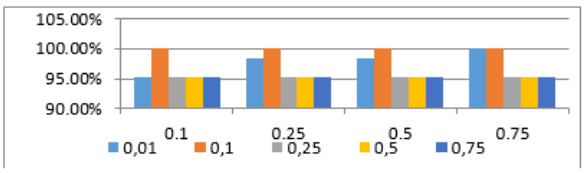


Figure 5. Evaluation Result of The Comparison Effect of Training Data and Testing Data Against Specificity

Based on the Figure 5 there was an increment specificity on the data which using deduction learning rate of 0.01 with a greater learning rate. Then the specification had increased to 100% on the data using deduction learning rate of 0.1 with learning rate progressively enlarged to 0.5 and then reached converging on the data with a deduction of 0.7 learning rate. Specificity then declined in the value of 95.29% and experienced a convergence in the data value with a deduction learning rate of 0.25 to 0.75.

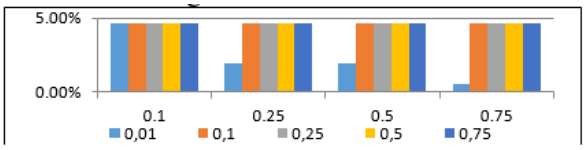


Figure 6. Evaluation Result of The Comparison Effect of Training Data and Testing Data Against Error Rate

Based on the Figure 6 there was a decrement in error rate on data which using deduction learning rate of 0.01 learning rate progressively enlarged. The decrement came from 4.67% to 1.94% to the least reached 0.58%. Then, the error rate increased on data that used a deduction learning rate of 0.1 and experienced the value of convergent till the deduction learning rate was 0.75.

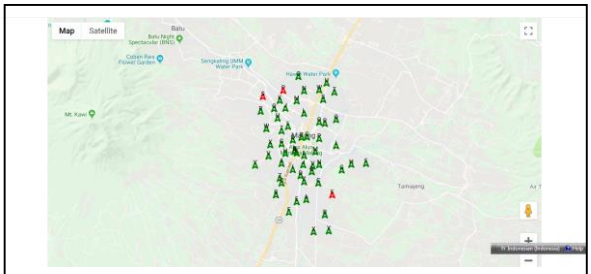
It could be concluded that the value of learning rate and deduction learning rate affected the performance of the algorithm. The higher the value of learning rate that ws used the more it would affect the performance of the algorithm. This was caused when the value of learning rate that used was greater, then the update value obtained the greater value. But the influence did not mean obtain the optimal performance of the resulting changed that the greater the learning rate and deduction learning rate with a value of 0.01, obtained the value of accuracy, specificity, and accuracy were improved, but the value of the sensitivity decreased that causing the performance was not stable. While the greater the deduction learning rate value the results obtained would converge. That was due to the greater the deduction learning rate value, the smaller the update value, it didn't have much effect on the performance of the algorithm. So the value of learning rate and deduction learning rate should not be too big and too small to obtain the optimal results.

B. Implementation

The results of identification with the best performance, displayed in a mapping simulation website so it can be used as basis information of flood disaster warning and emergency response of flood mitigation so the public knowledge about flood can be improved and reduced the potential loss. This web consists of home menu, map, and accuracy. Simulation website is described in Figure 8, Figure 9, and Figure 10.

Figure 7. Home Page of Simulation Website

Based on the Figure 7, home page of simulation website contains the title. In addition there is a button that directs users directly to map page.



STATUS/24 jam	Kelurahan Terdentifikasi Berpotensi Banjir Hari Ini
Waspada	kedungkandang
Waspada	tunggulwulung
Waspada	tlogomas

*Tidak ada pelebaran maupun penyempitan daerah teridentifikasi berpotensi banjir

Figure 8. MAP Page of Simulation Website

Based on the Figure 8, map page shows the districts location that have been identified. Green markers on map represent the areas that identified as area which has no potential of flood disaster, and red markers on map represent areas that identified as area which has potential of flood disaster. The users also can see the details by enlarging the map and see the information of district that identified as location that have potential of flood disaster.

Figure 9. Accuracy Page of Simulation Website

Based on the Figure 9, accuracy page shows the evaluation result of potential identification and mitigation of flood disaster with best parameters.

IV. CONCLUSION

Based on the research that had been conducted, could be concluded the following conclusions:

a. To identify potential and flood mitigation, The Neural Network Learning Vector Quantization Method was performed by calculating the distance between the nearest eucliedean input of drainage, elevation, and rainfall with a target value of identification.
b. To understand the performance of Neural Network Learning Vector Quantization Method which was used in the research using the confusion matrix with the best performance results were obtained in the form of value accuracy of 95.32%, sensitivity of 100%, the specification of 95.29% and an error rate of 4.68% in 1710 datasets that composed of 70% of training data, and 30% of testing data with the learning rate of 0.1, deduction learning rate of 0.01, maximum epoch was 10 and minimum epoch was 0.0000001.

REFERENCE

[1] Somantri, L. Utilization of Remote Sensing Techniques for Identifying Vulnerability and Flood Risk, Journal of Geography Gea, Vol 8, No. 2 (2008)
[2] Elza, S., Eleonara R., and L. Irsal. Efforts Agricultural Sector in the Face of Climate Change, Journal of Agricultural Research (2010)
[3] BPBN website, Indonesian Disaster Data and Information (2017). [Online] .www.bpbn.go.id, Retrieved on May 27 (2018)
[4] BPBDs Malang website, Disaster Risk Map. [On line].www.bpbd.malangkota.go.id, Retrieved on March 3 (2017)
[5] Law of the Republic of Indonesia Number 34 of Article 1 (2007)
[6] Malang Regional Regulation No. 11 (2014.)
[7] Maulina, Phitsa. Citarum River Flood Prediction with Fuzzy Logic Algorithms Results ParticleSwarm Optimization, Journal of Information Technology, Vol. 3 ISSN: 2353-6579. September (2016)
[8] Morrow, WK., Muslim MA., and Sunaryo. Construction of Flood Potential Models on Arterial Malang Using LogicFuzzy, EECCIS Journal, Vol. 7, No. 1 (2013)
[9] Irfiani, Eni. Prediction of Customer Complaints in Apartment Using Algorma C4.5, Journal of paradigms, Vol XVI, No. 2 (2014)
[10] Apriliyah, Mahmudy WF, and Widodo AW. Estimated Expenses Sales Litrik Resilent Using Backpropagation Neural Network (RPROP), Cursor, Vol 4, No. 2, pp. 41-47 (2008)
[11] Andrijasa, MF, & Mistianingsih. Application of Neural Network to Predict Unemployment in the province of East Kalimantan Using Backpropagation Learning Algorithm, Journal of Information Mulawarman, Vol 5 No. 1 (2010)
[12] Endang, T. Stock Price Forecasting Model with Feedback Propagation Neural Network, unpublished thesis. Bogor: Institut Pertanian Bogor (2008)
[13] Wuryandari, MD, & Afrianto, I. Comparison of Methods of Artificial Neural Networks Backpropagation and Learning Vector Quantitation on Facial Recognition, Journal of Computer and Information Technology (KOMPUTA), Issue 1, Vol 1 (2012)
[14] Indrabayu, Nadjamuddin AM. Saleh Pallu, Andani Achmad, Febi Febriyati. Rainfall Prediction with Fuzzy Logic, Journal of Proceedings, Vol 6, ISSN 978-979-127255-0-6 (2012)
[15] Fuad MFM. Rice Public Availability Prediction Using Fuzzy Logic and Neural Network In Effort meningkan Food Security. ARGOINTEK, Vol 5, No. 1 (2011)
[16] Rohaeti, E., Heryanto, R., Rafi, M., Wayuningrum, A., and Darusman, LK. Predicted levels of Flavoid Total Tempuyung (Sonchus arvensi L) Using IR Sprektroskopi Combined with Partial Least Squares Regression, Journal of Chemistry, ISSN 1907-9850, Vol 5, No. 2. 2011
[17] Fansuri, RM. Music Genre Classification Using Learning Vector Quantization (LVQ), Thesis Not Published. Bogor Agricultural University in Bogor (2011)
[18] Mudhakkir, I., Gratitude, A., Goddess, IN. Enhanced Accuracy Backpropagation Algorithm to Feature Selection Particle Swarm Optimization in Prediction customer- Telecommunications Customers Lost, Pseudocode Journal, Vol 1, No. 1, ISSN 2355-5920 (2014)
[19] Hartatik. Application of Learning Vector Algorithm for the Prediction Quantiozation Academic Values Using AMS Instrument (Academic Motivation Scale), TIE Scientific Journal, Vol.16, No. 3. S (2015)
[20] Risky, M., Susanto, O., Kartini, D. Application Method of Learning Vector Quantization (LVQ) on Prediction Department at SMA PGRI 1 Banjarbaru," Journal of Computer Science, Vol 4, No 1 (2016)