



COMPARISON OF AIR QUALITY DATA ACCURATION USING DECISION TREE AND NEURAL NETWORK METHOD

¹Fahmi Izhari., ²Hanna Willa Dhany

Sains dan Teknologi Universitas Pembangunan Pancabudi Medan

E-mail: ¹fahmi_izhari@dosen.pancabudi.ac.id., ²hdhany@dosen.pancabudi.ac.id

Abstract

In research conducted on the Neural Network classification model that has been tested has an accuracy of 82.04% with a classification error rate of 17.96%. Meanwhile, the Decision Tree classification model has an accuracy rate of 99.38 % with a classification error rate of 0.62%. Based on the test results from the two classification models, it can be concluded that the success of the Decision Tree can be used as a reference to improve the performance of the classification model's accuracy compared to the Neural Network Backpropagation model.

Keywords: Decision Tree, Neural Network, Classification, Accuracy.

INTRODUCTION

Information Technology is crucial in every life management, from its use in education to industry for management decision making. With the information technology that has been developed this has a positive impact on human life. Information technology is very valuable, because it provides benefits both directly and indirectly. Making people more productive in making changes and enhancing knowledge.

Data classification is data categorization into different categories according to the rules. In this grouping to change the structure of the object instance. The classification algorithm is made from training sets and models and models used to classify new objects. Decision trees improve composition by analyzing performance and the results of their analysis [1]. Classification of data objects based on objects that have been determined in a data. There are many classification algorithms, but decision trees are most often used[2]. Decision Tree Algorithm is one of the most

important collection sizes in data mining. Classification is one type of grouping diagram diagram like a tree structure, where each internal node determines the test on each attribute, each branch represents the results of the test, and each leaf binding represents the class.

The model for classifying records to find leaf paths to measure leaf attributes and attributes is the result of the classification used by Decision Tree[3].

Research conducted on trials that used data derived from results in the psychomotor aspects of children. By comparing the results of counting manually with the program obtained 85%[4].

Backpropagation is one of the methods in the artificial neural network that is popular and has the advantage in its ability to learn, adapt to an object, and compatibility with errors [5]. But backpropagation has weaknesses in the learning process that requires a long time so this method requires a long time in reaching convergent levels [6]. The degree of convergence is the value of the mean squared error (the difference



between the actual output and the target output [7]. Factors affecting learning in backpropagation are input, bias, the pace of learning or the level of learning and momentum. One way that is widely used to improve convergence in the backpropagation method is to provide a good input value on the backpropagation method. Meanwhile, according to Backpropagation is one of the artificial neural networks that has a process of forward learning and backward error[8]. In testing conducted on backpropagation neural networks requires a combination of data that reaches 60% of the total data. From the background of the problems that have been resolved before, the writer takes the formulation of the problem that is needed in the performance evaluation process of the algorithm applied can learn which algorithm is more effective and accurate in carrying out the classification process.

METHOD

Research conducted in trials used data derived from the results of assessments in the psychomotor aspects of children. By comparing the results of manual calculations with the program, the success rate is 85%[9]. Backpropagation is one of the methods in the artificial neural network which is popular and has advantages in its ability to learn, adapt to an object, and tolerate mistakes [10]. But backpropagation has weaknesses in the learning process that requires a long time so this method requires a long time in reaching convergent levels[11]. The degree of convergence is the mean square error value (the difference between actual output and target output) approaches zero when the algorithm studies each pattern (Hagan et al, 2002). One way that is widely used to improve convergence

in the backpropagation method is to provide a good input value on the backpropagation method. Meanwhile according to[12].

Whereas the Decision Tree is a flow chart that is shaped like a tree structure with each internal node testing the attributes, the branches do the output of the test and the leaf nodes do class grouping or class distribution. The most important node is called the root node or root. A root node has several exit edges but no entry edge. Internal nodes will have one entry edge and several exit edges, while the leaf node will only have one edge without having an exit edge. Decision tree is used to classify a data that still does not know its class to existing classes. The data testing path is the first step that the root node goes through and finally the leaf node will predict the class for the data that has been deduced.

RESULT

Dataset Pekanbaru City Air Quality has 5 attributes, 6 classes and 1,096instances, Class distribution in the form of good (746instances), dangerous (33instances), very unhealthy (8instances), moderate (237instances), no data (11instances), unhealthy (61instances)). The data is shared as much as 70% of the data will be used as training data and as much as 30% of the data will be used as random test data.

Table1. Dataset Attribute Information

No.	Atribut	Nilai
1	<i>PM10</i>	[3-870]
2	<i>SO₂</i>	[1-57]
3	<i>CO</i>	[1-150]
4	<i>O₃</i>	[2-211]
5	<i>NO₂</i>	[1-105]
9	<i>Kategori</i>	{baik, berbahaya, sangat tidak sehat, sedang, tidak ada data, tidak sehat}



The details of the Pekanbaru City Air Quality dataset can be seen in table 2

Table 2 Data Details

No	PM10	SO ₂	CO	O ₃	NO ₂	Kategori
1	47	51	8	67	2	Sedang
2	48	51	9	37	2	Sedang
3	37	51	9	26	2	Sedang
4	24	50	2	51	1	Sedang
1096	37	11	7	23	7	Baik

Based on the above table, then proceed with calculating the Accuracy value and the classification error rate (Classification_error) of the Decision Tree model. Here are the results of the calculation:

a. Accuracy

$$= \frac{67+227+19+1+8+0}{67+227+19+1+8+0+1+0+1+0+0+0} = \frac{322}{324} = 0.9938 * 100\% = \mathbf{99.38\%}$$

Table 3. Hidden Layer Value

Atribut	Nod e 1	Nod e 2	Nod e 3	Nod e 4	Nod e 5	Nod e 6	Nod e 7
PM10	-	-	-	-	-	-	-
	1.95	4.88	0.11	5.96	1.67	2.45	5.20
SO2	-	-	-	-	-	-	-
	0.22	0.19	0.46	0.11	0.17	0.31	0.16
CO	-	-	-	-	-	-	-
	0.48	0.91	0.62	1.10	0.57	0.52	0.99
O3	-	-	-	-	-	-	-
	0.53	0.98	0.86	1.16	0.55	0.60	1.00
NO2	-	0.09	-	0.23	-	-	0.11
	0.25	-	0.50	-	0.45	0.18	-
Bias	-	-	1.04	-	-	-	-
	2.56	5.86	-	6.99	1.89	3.23	6.22

b. Classification_error=

$$\frac{1+1+0+0+0+0}{67+227+19+1+8+0+1+0+1+0+0+0} = \frac{2}{324} =$$

$$0.00617 * 100\% = \mathbf{0.62\%}$$

Then testing the Neural Network method. The results of the Hidden Layer 1 weight propagation on each attribute of Pekanbaru City Air Quality Data use the Backpropagation classification model:

Table 4 Results of Propagation of Hidden Layer Weights

Kinerja Klasifikasi	Predicted Class						
	Actual Class	Predicted Sedang	Predicted Baik	Predicted Tidak Sehat	Predicted Sangat tidak Sehat	Predicted Berbahaya	Predicted Tidak Ada Data
Sedang		67	1	0	0	0	0
Baik		1	227	0	0	0	0
Tidak Sehat		0	0	19	0	0	0
Sangat Tidak Sehat		0	0	0	1	0	0
Berbahaya		0	0	0	0	8	0
Tidak Ada Data		0	0	0	0	0	0



Table 5. Propagation Results Layer Output Weight

Node	Bobot Output Layer (Class)					
	Sedang	Baik	Tidak Sehat	Sangat Tidak Sehat	Berbahaya	Tidak Ada Data
Node 1	-4.649	4.031	-2.664	0.819	-1.604	0.798
Node 2	4.238	0.050	0.576	1.026	-2.856	-1.174
Node 3	-1.044	2.510	-1.985	0.908	-1.776	-0.973
Node 4	-3.007	3.390	-2.501	0.883	-1.621	-0.896
Node 5	3.976	-0.122	0.905	0.937	-2.918	-1.128
Node 6	-0.132	1.831	-1.498	0.903	-1.838	-1.004
Node 7	-1.758	2.944	-2.271	0.934	-1.731	-0.959
Threshold	-3.910	-6.199	-0.151	1.595	1.005	-2.566

So we can get the results of the conventional classification of Neural Network Backpropagation (Pekanbaru Air Quality) accuracy:

Table 6 Confusion Matrix classification Neural Network Backpropagation

Kinerja Klasifikasi	Predicted Class					
	Predicted Sedang	Predicted Baik	Predicted Tidak Sehat	Predicted Sangat Tidak Sehat	Predicted Berbahaya	Predicted Tidak Ada Data
Actual Class Sedang	57	4	10	0	0	0
Actual Class Baik	41	181	0	0	0	0
Actual Class Tidak Sehat	0	0	18	0	0	0
Actual Class Sangat Tidak Sehat	0	0	1	0	1	0
Actual Class Berbahaya	0	0	0	1	9	0
Actual Class Tidak Ada Data	0	0	0	0	0	0

Based on the above table, then proceed with calculating the Accuracy value and the classification error rate (Classification_error) of the conventional Neural Network Backpropagation classification model (Pekanbaru Air Quality). The following results are calculated:

a. Accuracy

$$\frac{57+181+18+0+9+0}{57+181+18+0+9+0+41+4+10+1+1+1} = \frac{265}{323} = 0.8204 * 100\% = \mathbf{82.04\%}$$

b. Classification_error=

$$\frac{41+4+10+1+1+1}{57+181+18+0+9+0+41+4+10+1+1+1} = \frac{58}{323} = 0.1795 * 100\% = \mathbf{17.96\%}$$

CONCLUSION

In research conducted on the Neural Network classification model that has been tested has an accuracy of 82.04% with a classification error rate of 17.96%. Meanwhile, the Decision Tree classification model has an accuracy rate of 99.38 % with a classification error rate of 0.62%. Based on the test results from the two classification models, it can be concluded that the success of the Decision Tree can be used as a reference to improve the performance of the classification model's accuracy compared to the Neural Network Backpropagation model..

REFERENCE

- [1] Arif Jurmawanto., Rudy Hartono., Dhidik Prastiyanti. 2013. *Aplikasi Jaringan Syaraf Tiruan Backpropagation untuk Memprediksi Penyakit THT di Rumah Sakit Mardi Rahayu Kudus.* Kudus..
- [2] Brian Thomas. 2016. *Analisis Learning Rates Pada Algoritma*



- Backpropagation Untuk Klasifikasi Penyakit Diabetes.* Universitas Trunojoyo Madura Bangkalan, Indonesia.
- [3] Dai Qin-yun., Zang Chun-Ping., Wu Hao. 2016. *Research of Decision tree Classification Algorithm in Data Mining.* Dept. of Electric and Electronic Engineering, Shijiazhuang Vocational and Technology Institute. China
- [4] Dimililer Kamil. 2013. *Backpropagation Neural Network Implementasi For Medical Image Compression.* Electrical and electronic engineering department. Turkey.
- [5] Han. P. H, Chao C. L, 1998. *A Learning Fuzzy Decision Tree and its Application to Tactile Image. Intelligent Robots and System,* Proceedings IEEE/RSJ International Conference on Volume : 3.
- [6] Izhari F, Dhany H W, Zarlis M, et al. 2017. *Analysis Backpropagation Methods With Neural Network For Prediction Of Children's Ability In Psychomotoric.* 2nd International Conference on Computing and Applied Informatics
- [7] Manasi M. Phadatare, Sushma S. Nandgaonkar. 2014. *Uncertain Data Mining usig Decision Tree and Bagging Technique.* Department of Computer Engineering, India.
- [8] Omaima N.A., AL-Allaf. 2010. *Improving the Performance of Backpropagation Neural Network Algorithm for Image Compression/Decompression System.* AL-Zaytoonah Private University of Jordan.
- [9] Pal. M. 2007. *Random Forest Classifier for Remote Sensing Classification.* National Institute of Technology, Department of Civil. Haryana
- [10] Patel, B. N., Prajapati G. Satish., Lakhtaria I. Kamaljit. 2012. *Efficient Classification of Data Using Decision Tree.* Bonfring International Journal of Data Mining, Vol. 2, No.1.
- [11] Pratap Singh Kirar Vishnu. 2015. *Improving the Performance of Back-Propagation Training Algorithm by Using ANN.* International Journal of Computer and Information Engineering.
- [12] Seema., Rathi Monika., Mamta. 2012. *Decision Tree: Data Mining Techniques.* Department of Computer Science Engineering. India.
- [13] Turban Efraim., E Jay., Aronson., Liang Ting-Peng. 2005. *Decision Support System and Intelligent System.* Andi Offset
- [14] Vapnik, V. 1998. *Statistical Learning Theory.* Wiley New York