

# Scholarship Acceptance Selection Using Neural Network Method

Rammadhany, Slamet Riyadi, and Asroni\*

*Universitas Muhammadiyah Yogyakarta, Jln.Brawijaya, Tamantirto, Kasihan,  
Bantul, Yogyakarta 55183, Indonesia*

*\*Corresponding author: asroni@umy.ac.id*

## **Abstract**

*Muhammadiyah University of Yogyakarta is one of the private tertiary institutions that provides scholarship programs. UMY provides scholarship programs for outstanding students and middle-lower class students. In the process of receiving a scholarship, UMY has two stages, namely registration and selection. In order to obtain a scholarship, students who register must be declared to pass administrative selection and meet the assessment component as a condition for students to be eligible for scholarships. In order to know the scholarship information received by students, data processing is needed, data processing is often referred to as data mining. This writing is done to predict prospective scholarship recipients using the Neural network algorithm. The methodology at this writing begins with searching for literature studies, choosing data mining methods, data collection, data processing, application and testing of models, results and conclusions. The data used at this writing are 2019 general scholarship data, GPA data, and organizational active information data. The attributes used are parental income, GPA, scholarships, qualification and non-qualification, and active organizational information. The caption attribute qualification and non-qualification is used as a label. At this writing the authors get an accuracy of 72.62%..*

**Keywords:** *Algorithm Neural network, Data Mining, Scholarship*

## **1. Introduction**

In this day and age Technology has entered the era where the fourth industrial revolution or industry 4.0 The development of the industrial revolution 4.0 caused an increasing number of information exchanges. Due to the unstoppable exchange of information, technology is needed that is able to mine data to be analyzed into information that is useful for educational institutions or institutions such as universities.

In every college, especially in universities, there are many work programs offered to students, one of which is a scholarship program. Scholarships a gift in the form of financial assistance given to individuals that aims to be used for the sustainability of the education being pursued. This program is held to help ease the burden on students who carry out their studies. Scholarships offered are for highachieving and underprivileged students so that students do not need to think about the burden of tuition fees and focus on academic achievement.

As one of the institutions engaged in the academic field selecting scholarships at Yogyakarta Muhammadiyah University, it still faces obstacles in making decisions. The process sometimes requires quite a long time, this is because the selection process is still done manually and the existence of data processing that has not been transparent, so it needs to be made a system that selects automatically that can help

the institution in making decisions quickly so that scholarships can be distributed on time and right on target.

The existence of a lot of data causes decision makers needs tools that help in selecting prospective scholarship recipients quickly in decision making. To carry out an assessment process with a lot of data, we need a system that is able to select to improve the effectiveness of decision making and reduce subjectivity in the decision-making process (Frieyadi, 2017).

In the field of data mining there are several algorithms that can be used to conduct data selection, one of which is the neural network algorithm. Neural network algorithm is a method inspired by the neural network of the human brain because it is designed to follow the way the human brain processes and stores information (Hermanto & Rudi, 2010).

Therefore, based on the problems that arise in this study, the writer thinks to raise it to the title "Scholarship Acceptance Selection Using Neural Network Method" which can help students in finding solutions to determine the scholarship recipients at Muhammadiyah University, Yogyakarta.

## 2. Mining Data

Data Mining or also often called Knowledge Discovery in Database (KDD). Is a process that aims to get information by extracting data in the database into information, then information becomes knowledge, and knowledge becomes an experience. The Knowledge Discovery in Database process consists of a combination of several techniques such as Machine Learning, Statistics, Pattern Recognition (Fayyad, Usama, Piatetsky-Shapiro, & Padh, 1996). Data Mining in general can be applied to all types of data as long as the data can be used. In general, data used for data mining is large-sized data such as databases, data warehouses, and transaction data (Han, Kamber, & Pei, 2011). Data mining can also be applied to analyze existing data at a university to see an increase in student ability based on the history of existing grades (Ridwan, Mujib, Suyono, & M. Sarosa, 2013).

In processing data mining there are 5 commonly used processing techniques, i.e (Larose, Daniel T, & D. Larose, 2014):

1. Description: Description techniques are usually used by writers and analysts to look for patterns in the data.
2. Estimation: Estimation techniques are used by the authors to estimate the value of the target variable with numerical predictor variables.
3. Prediction: Prediction techniques are similar to estimation and classification techniques. It's just a little difference lies in the final result, the results of prediction techniques are in the future.
4. Classification: Classification techniques are similar to estimates. The difference is invariable the categorized target is not numeric.
5. Clustering: Technique clustering is a data set similar to another, and usually the authors group similar data into the same object class. Technique clustering not the same as classification because of absence variable target for clustering.

## 3. Neural Network

Neural network algorithm is a method inspired by the neural network of the human brain because it is designed to follow the way the human brain processes and stores information (Hermanto & Rudi, 2010). Neural network algorithm is used as a tool that describes non-linear statistical data, by describing a complex relationship between input and output (Singh, Yashpal, & Chauhan, 2009).

Neural network consists of several layers, each layer usually has at least 1 or more Processing Elements (PE). Processing Elements are used to simulate the workings of neurons in the human brain. Therefore, PE is also often called a neuron or node, each PE receives input from the previous layer (Singh, Yashpal, & Chauhan, 2009).

Neural network consists of neurons that are related between one neuron with another neuron. In general, there are 3 types of neural networks that are often used (Aprilla, Dennis, Baskoro, & Ambarwati, 2013):

1. Single-Layer Neural Network
2. Multilayer Neural network
3. Recurrent Neural networks

In this study, the authors use a type of Multilayer Perceptron Neural Network has one additional layer, the hidden layer between the input layer and output. This type of neural network has at least 3 layers, namely the input layer, hidden layer, and output layer. This type of neural network is one type of network that is often used (Aprilla, Dennis, Baskoro, & Ambarwati, 2013).

Multilayer Perceptron is a feedforward neural network in which the information received moves only one direction forward from the input layer through the hidden layer and the output layer (Meinanda, Hanief, Annisa, & Mu, 2009).

In the third edition of the Data Mining Concepts and Techniques book by Jiawei Han, Micheline Kamber, and Jian Pei, the mathematical formula of Neural Network has been calculated to calculate net input to units, each input connected to a unit multiplied by weight and then summed with bias. The Neural Network calculation formula is as follows (Han, Kamber, & Pei, 2011):

$$I_j = \sum_i W_{ij} O_i + \theta_j$$

- $J$  = hidden or output layer  
 $I_j$  = Net Input  
 $W_{ij}$  = connection weight from unit  $i$  layered previously to unit  $j$   
 $O_i$  = Output from unit  $i$   
 $\theta_j$  = bias

#### 4. Research Stage

Before conducting research, it is necessary to determine what stages will be passed, the stages in this study can be seen in Figure 1.

1. Study of literature

The purpose of conducting a literature study at this writing is to look for research related to data mining, and the use of Neural Network algorithms. Conducting a literature study also aims to strengthen the theories contained in this paper.

2. Data collection

The data used in this paper are secondary data sources, the data used in this paper the author gets from one source. Sources obtained at LPKA (Student and Alumni Development Institute) Muhammadiyah University Yogyakarta.

3. Data processing

Data selection (Data Selection) in a database of course there is so much data, of course, not all data that will be used, therefore data selection is needed to determine the data that suits your needs.

Data cleaning. After going through the data processing at this stage, cleaning of the values of the attributes is performed. Because there are still attribute values that cannot be used because there is a lot of null, noise, and irrelevant data, it is necessary

to clean up the data to ensure the values of each attribute are in accordance with what is needed.

This stage requires a long process because the writer wants to get the best accuracy results so that in the process of cleaning up the data, it becomes a concern for the writer to ensure that the data that will be used is completely clean from null data, inconsistent data, and from irrelevant data.

Data Transformation Data that has been combined is then transformed into an appropriate form for later processing with data mining using the RapidMiner software. At this stage of data transformation, the author makes several initialization values for each attribute that will be used. As for some of the attribute values that are initialized:

1. Parents' income / salary: Family income is the sum of all income and wealth used to divide the family.
2. GPA: is an abbreviation of the Grade Point Average is a measure of a student's ability up to a certain period calculated based on the number of SKS that have been taken.

Because the existing attributes are considered to be insufficient, it is necessary to add two attributes, namely the Organizational Active attribute and the passing and non-qualifying information, while the objective in adding the Organization Activation attribute is to strengthen the level of accuracy so that it is more accurate and targeted, and the information passing and not passing is appropriate with the rector's decision Yogyakarta Muhammadiyah University Number: 337 / KEP-UMY / VII / 2019 Regarding the awarding of scholarship recipients for Yogyakarta Muhammadiyah University students.

Furthermore, the authors classify the attributes of parental income and GPA. The purpose of doing classification is to make it easier to predict the level of accuracy in rapid miners.

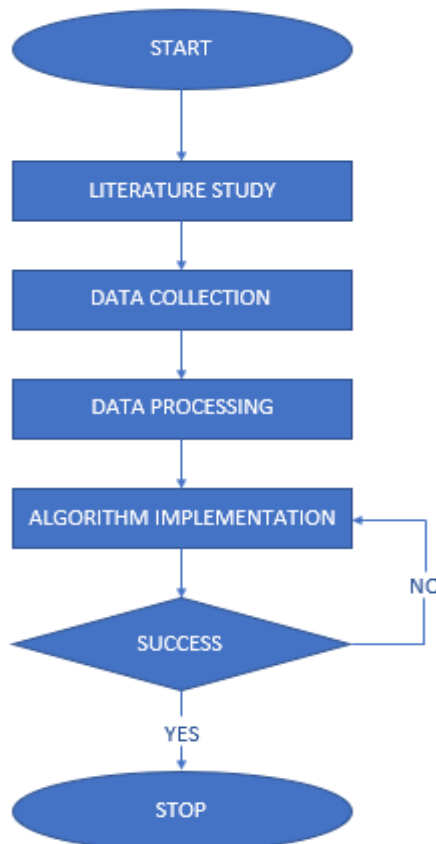


Figure 1. Research Stage

## 1. Algorithm Implementation

This stage is the most important process in this writing because it is at this stage that data mining to find information is carried out. Previously processed data is referred to as training data, which data will be trained using the Neural Network method using RapidMiner software. The attributes that will be tested at this stage are the pass and not pass tables, which are calculated from the number of applicants for general scholarships.

The implementation phase of this algorithm is the writer doing the test. In testing out of a total of 556 preprocessing data the authors divide the data into training data (training) and test data (testing).

In this test the writer uses cross validation operator. This cross-validation operator will later separate the data by dividing the total data from 90% to training data and the remaining 10% to test data.

In this study the authors use a Multilayer Neural network consisting of 3 layers, namely: Input layer, Hidden layer, and Output Layer. The number of nodes in the Input layer is based on the number of values that exist in each attribute. The number of nodes in the hidden layer is based on the number of hidden nodes that have been configured. The number of nodes in the Output layer is based on the number of values in the attribute that is a class or label. The type of topology used in this study is the feedforward topology, because at the testing stage there is no feedback to the previous node. Then the operator is added as shown in Figure 2.

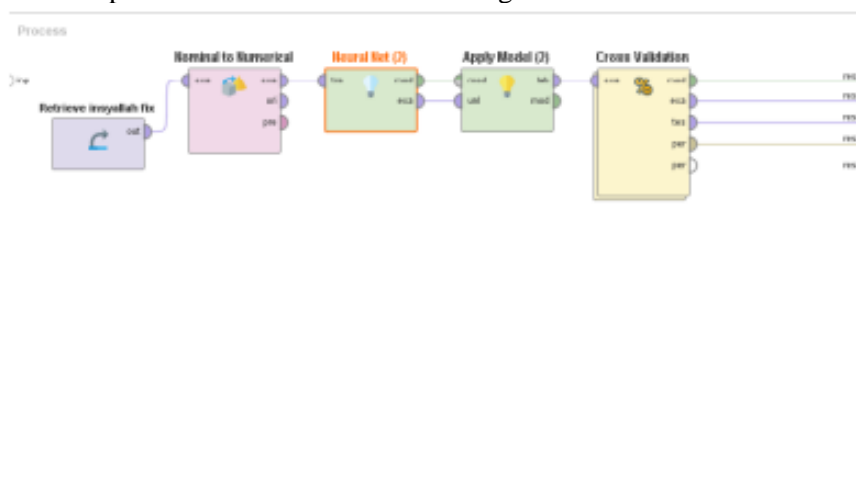


Figure 2. Rapid miner configuration process

After all the operators and files used are connected, to find out the results of the test click the run button on the top toolbar like Figure 4.8. After clicking on the run button, wait a few moments for the computer to process.

After waiting a while, the RapidMiner display will move to the Result sheet which displays the results of the process using the Neural network. The results of Neural network modeling obtained as shown in Figure 3. Neural network modeling obtained consists of 3 layers, namely: input layer, hidden layer and output layer. At the input layer there are 25 nodes, each node derived from the values that exist in each attribute used. In the hidden layer there are 2 nodes, each node in the hidden layer receives input from every node in the input layer. For the output layer there are 2 nodes which correspond to the amount of data in the attribute label.

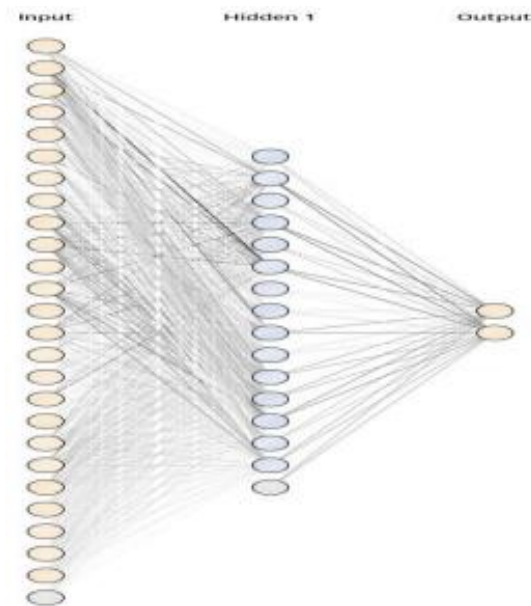


Figure 3. Modeling Neural Network

Performance Vector Tab will display the level of accuracy generated by using a Neural network. In table 4.3 it can be seen that the level of accuracy obtained using Neural networks is equal to 72.62%. For the class precision results passed that is equal to 57.03%. and class precision does not pass that is 78.82%. For class recall, it passed 51.77%. and class recall did not pass 82.14%.

Table 1. Confusion Matrix Result

Accuracy: 72.62%			
	True does not pass	True passed	Class precision
Pred. Did not pass	253	68	78.82%
Pred. Get away	55	73	57.03%
Class recall	82.14%	51.77%	

In testing using cross validation there are several types of data retrieval, for this test the author not only uses 1 type of data retrieval but uses 3 types of data retrieval as in table 1. the goal is to compare and to find out which type of data retrieval is the best. Even the results can be seen in table 2. It can be seen that the type of Stratified sampling taking results in an accuracy rate of 72.62%. Stratified sampling is a type of data collection by constructing a random set, which distinguishes from shuffle sampling, that is, stratified sampling ensures that the distribution of the subset is equal to the entire example set.

Table 2. Sampling Type Result

<i>Sampling Type</i>	<i>Result</i>
Stratified Sampling	72.62%
Shuffled Sampling	68.84%
Linear Sampling	69.72%

Based on the results of table 4.4, the best results are obtained with stratified sampling type sampling with an accuracy of 72.62%. The process of collecting stratified data is randomized, but the difference is that stratified sampling ensures the

distribution of attributes in the subset is the same as in all data sets. (Rapid miner documentation, 2020).

## 5. Conclusion

From the results of the discussions carried out in this study, the following conclusions are obtained:

1. The neural network method can be applied for scholarship recipient selection, with an accuracy of 72.62%.
2. Neural network algorithm is able to quickly process data and also produces an accurate level of accuracy that is equal to 72.62%.

## References

- [1] LPKA UMY. (2020, 4 26). LPKA UMY. Retrieved from <https://lpka.umy.ac.id/beasiswa/>.
- [2] Aprilla, Dennis, Baskoro, D., & Ambarwati, L. (2013). "Learn data mining with RapidMiner." Jakarta: Gramedia Main Library.
- [3] Dewi H, R., Yunita, & Indrawati, N. (2015). Design and Build a Scholarship Decision Support System Using the Naive Bayes Classifier Method.
- [4] Fayyad, Usama, Piatetsky-Shapiro, G., & Padh. (1996). From data mining to knowledge discovery in databases. " AI magazine 17 (3): 37–37.
- [5] Frieyadie. (2017). USE OF SIMPLE ADDITIVE WEIGHTING METHOD FOR DETERMINING FEASIBILITY OF GIVING SCHOLARSHIP FOR ACHIEVING STUDENTS.
- [6] Han, J., Kamber, M., & Pei, J. (2011). Data Mining Concepts and Techniques.
- [7] Haryanto, D., Ramdani, C., Wahidah, W., Dinia, A., & Oktaviani, S. (2019). Design of Decision Support System for Determination of Scholarship Acceptance Using Artificial Neural Network (ANN) Method at Telkom Purwokerto Technology Institute.
- [8] Hermanto, & Rudi, R. (2010). Neural Network and Its Implementation in Data Mining.
- [9] Hofmann, Markus, & Klinkenberg, R. (2013). RapidMiner: Data mining use cases and business analytics applications. CRC Press.
- [10] Larose, Daniel T, & D. Larose, C. (2014). Discovering knowledge in data: an introduction to data mining. John Wiley & Sons.
- [11] Meinanda, Hanief, M., Annisa, M., & Mu, N. (2009). "Predicting undergraduate studies with an artificial Neural Network." Internetworking Indones. J 1 (2): 31–35.
- [12] Merdekawati, A. (2013). DECISION SUPPORT SYSTEM FOR RECEIVING SCHOLARSHIP USING ALGORITHM C4.5.
- [13] Nabahan, D., Wijaya, A., & William. (2019). Analysis of Factors that influence students in taking scholarships using the C4.5 algorithm.
- [14] Pauziah, U. (2016). STUDY COMPARISON OF C4.5 ALGORITHM, NA BVE BAYES AND NEURAL NETWORK IN SELECTING SCHOLARSHIP RECIPIENTS.
- [15] Permana, R., & Sahara, S. (2019). Support Vector Machine Method as a Determination of Student Graduation in Electronic Learning. Journal of the Equator Informatics 7.

- [16] Pujianto, A., Kusrini, & Sunyoto, A. (2018). DECISION SUPPORT SYSTEM DESIGN FOR SCHOLARSHIP RECIPIENT PREDICTION USING THE NEURAL NETWORK BACKPROPAGATION METHOD.
- [17] Rapidminer documentation. (2020, 4 26). Retrieved from [https://docs.rapidminer.com/latest/studio/operators/blending/examples/sampling/sample\\_stratified.html](https://docs.rapidminer.com/latest/studio/operators/blending/examples/sampling/sample_stratified.html).
- [18] Ridwan, Mujib, Suyono, H., & M. Sarosa. (2013). Application of Data Mining for Student Academic Performance Evaluation Using the Naive Bayes Classifier Algorithm. " EECCIS Journal 7 (1): 59–64.
- [19] Romero, C., Ventura, S., & García, E. (2008). "Data mining in course management systems: Moodle case study and tutorial.
- [20] Siddiq, M., Tamin, R., & Azis, S. (2019). Data Mining Technique as Majors Support System Management with Classification Approach.
- [21] Singh, Yashpal, & Chauhan, A. (2009). "NEURAL NETWORKS IN DATA MINING." Journal of Theoretical & Applied Information Technology 5 (1).
- [22] Sumiah, A., Maulana, W., & Rinaldi, D. (2018). "Implementation of data mining for the recommendation of determining scholarship recipients using the Naive Bayes method".
- [23] Suryanto, Anwari Rahman, A., & Agus. (2017). "IMPLEMENTATION OF SCHOLARSHIP RECIPIENT SELECTION INFORMATION SYSTEM USING NAIVE BAYES CLASSIFIER METHOD".
- [24] Wibowo, a. (2017, 11 24). BINUS MTI. Retrieved from <https://mti.binus.ac.id/2017/11/24/10-fold-cross-validation/>