

Application of Neural Network Overview In Data Mining

Hendrayana Putu Bagus Surya^[1], Rifky Lana Rahardian^[2], and Made Sudarma^[3]

[1][2] Department of Electrical and Computer Engineering, Post Graduate Program, Udayana University

[3] Department of Electrical and Computer Engineering, Udayana University

Email: ags.hendrayana@gmail.com

Abstract - Data Mining is the term used to describe the process extract value / information from the database. Four things are needed in order to effectively data mining: data that has a high quality, right of data, examples of which are adequate, and the correct device. To obtain valuable information in the required data mining algorithms applied in data mining in large databases. There are a lot of complex algorithms in data mining. One is the so-called Neural networks have an important role in data mining.

I. PRELIMINARY

Differences data and information could be analogous to the comparison between a rock and a 99.99% *ingots*. pure gold We know that there is gold in the rock. But we do not know where the exact location. As with the data. *Data mining* is the term used to describe the process *extract* value / information from the database. Four things are needed in order to effectively data mining: data that has a high quality, *right of* data, examples are *adequate*, and the correct device. There are many algorithms in *the data mining*. One is a *neural network* as a *benchmark* for build a model in data mining.

With so many existing methods, the author intends to review the application of neural network in data mining. So the next researcher can consider the selection of the right method when you want to do something with data mining techniques. The result of this paper is a review of previous researchers who have done the application of neural network in data mining. The author tries to re-explain the importance of neural networks in data mining, the advantages and disadvantages of neural networks in data mining.

II. LITERATURE

A. Neural Network

Known Neural network because it is designed to follow the way the brain processes and stores information. The human brain consists of hundreds to jutaan neurons. Each neuron has a simple design. Neural network composed of nodes that combine the inputs (variables from the database or output from the other node). These

nodes can be classified in three simple layer. Input layer, middle layer and output layer.

First we create a simple logistic model that shows the neural network. Here, it takes quite a long time to build a better model using logistic refression compared neural network. In practice, the neural network is a device non-linear statistical data modeling. Neural networks can be used to model complex yan relationship between inputs and outputs to look for patterns in the data. Neural network is essentially composed of three parts: architecture or model, learning algorithm, and activation functions. In the picture below are shown artificial neural network. [1]

Neural Network is a parallel computing system of multiple processors connected. *Neural Network* offers a mathematical model that attempts to replicate the human brain. *Backpropagation* is the generalization of Widrow-Hoff learning rule on multiple-layer networks and nonlinear transfer function is different. *Input vector* and *vector-related* targets are used to train the network to be approached a function, which connects the *vectors* input with specific output vector or classify *an input vector* into an appropriate manner that has been defined. training *Network* with *backpropagation* involves three stages, namely *feedforward* to determine input training patterns, *propagation-related error*, and adjustments on *weight*. [2]

B. The advantages Neural

1. of high accuracy Networks: *Neural network* is used for *mapping* complex non-linear approximation.
2. Tolerance to *noise*: The network was very flexible with *noisy* data.
3. Independence of the *prior* assumption:

The network did not make *priori* assumptions about the distribution of data or a form of interaction of factors.

4. Easy to manage: *Neural networks* can be updated with the new data, make useful for dynamic environments.
5. Neural networks can be implemented in parallel

hardware.

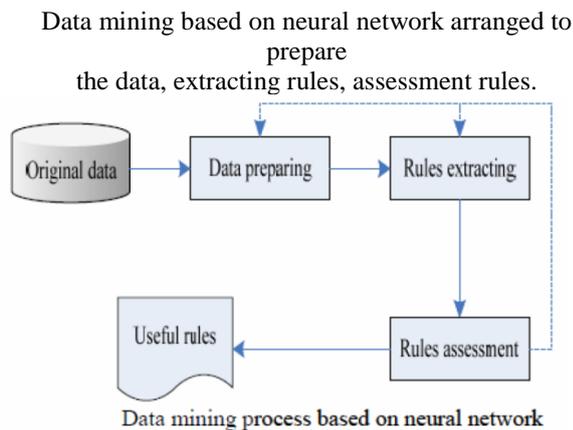
6. When the element *neural network* fails, he can proceed without problems because a parallel pattern.

7. *Neural networks* can be trained in a very large dataset iteratively. [1]

III. DISCUSSION

There are important factors in the implementation of a neural network:

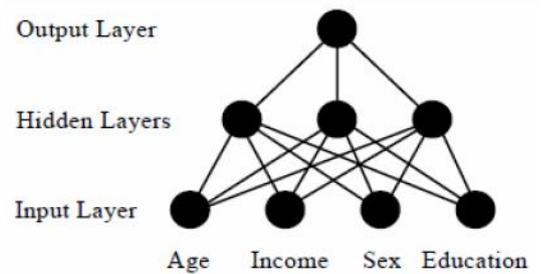
1. An effective combination of neural network and data mining technology.
2. An effective combination of process knowledge and neural computation.
3. The method using a neural network is very complex, required devices with relational databases, multidimensional data bases and data warehouses that can work together for the needs of data mining.



Example:

We have a database that consists of records of thousands of people who order a brochure about our latest products. Database consisting of four variables: **age, gender, income, education** level, and whether she booked our products after reviewing the information that we provide.

We can build the neural network with four input nodes (**age, gender, income, education level**), an output node (**behavior reservations**), and one hidden layer node. We can train the network using a database. After that, we can use it to send out brochures to people who are classified as a client by a neural network. The strength of neural network with one hidden layer which can mengaproksimasi types of non-linear function and interaction terms.



It can be concluded under the input layer we can enter the variables that needed a data mining to generate output information. Then in the hidden layer inserted the appropriate algorithms to search for that output.

In the study conducted by Meyliana titled Prediction Rental Income With Data Mining at XYZ Company is also applying neural network data mining. This study has the result that is how *data mining* will analyze several parameters of the factors that affect the rental income to generate rental income predictions are more accurate so as to support the management of PT. XYZ in strategic decision making. Research will be conducted to collect the parameters that determine the predictions of rental income, build data mining models for the prediction of rental income, tested a model of data mining on data from PT. XYZ, evaluate the test results data mining models on data from PT. XYZ. Data mining technique used is *Neural Network algorithm Backpropagation* [2].

The results obtained by these studies is the prediction of rental income with data mining produces more accurate prediction than the previous prediction of rental income which is constant based on last month's revenue in the previous year. Data mining technique used is *Neural Network* with algorithm *backpropagation Levenberg Marquardt* for this algorithm is the best algorithm in function approach.

To get the right model, training is done many times with variations in the value of μ and hidden neurons. Neural Network Model achieve a minimum MSE value ie when $\mu = 1$ and the number of hidden neurons = 125. Then the model used in predicting the rental income in the next year. From experiments showed that this model has produced a total value of error is much smaller than the total value of the previous error to find patterns in the data in previous years. Predictions with data mining lease income is 582 times more precise than the previous prediction. Thus, the benefits of this research has been reached that could help XYZ in making strategic decisions based on secret information from an existing database.

Prediction rental income that has been made can be developed using the Bayesian regularization training techniques because these techniques can result in a smaller MSE compared with *early stopping* in some cases, especially on the small amount of data when the data contains little variation (Demuth, 2009). Bayesian regularization technique will be much longer process than *early stopping*. In addition, this

rental income predictions can also be expanded in scope covering all subsidiaries.

IV. CONCLUSION

Neural networks can be trained on a very large dataset iteratively. Data mining is becoming more useful to combine the power of the neural network with statistical tools. Trust with this combination, can generate synergies and good sense to the data mining.

Prediction rental income with data mining produces more accurate prediction than the previous prediction of rental income which is constant based on last month's revenue in the previous year. Data mining technique used is *Neural Network* with algorithm *backpropagation* Levenberg Marquardt for this algorithm is the best algorithm in function approach.

References

- [1]. Rudolf Rudi Hermanto, "Neural Network and Implementation In Data Mining", 2010.
- [2]. Meyliana, "prediction sets include Data Mining Lease Revenues At XYZ Company", 2010.
- [3] Demuth, H., et al. (2009). *Neural Network Toolbox™ 6 User's Guide*. Natick: The MathWorks.
- [4] Yashpal Singh. *Neural Networks In Data Mining*. United Institute of Engineering & Technology India.
- [5] A. Ainslie, X. Dreze. "Data Mining: Using Neural Network as a Benchmark for Model Building"
- [6] Jianlin Cheng. "A Neural Network Approach to Ordinal Regression".
- [7] Yang gao. "Learning Classifier System Ensemble for Data Mining". Nanjing University.
- [8] Yashpal Singh. "Neural Network In Data Mining" United Institute of Engineering & Technology India.
- [9] Xianjun Ni. "Research of Data Mining Based on Neural Networks".
- [10] Mohd, Z., Azlan, et al. "Application of Regression and ANN Techniques for Modeling of the Surface Roughness in End Milling Machining Process". IEEE Computer Society, DOI 10.1109/AMS.2009.76.
- [11] Zhang, H., et al. "Multi-Objective simultaneous prediction of waterborne coating properties. Journal of Mathematical Chemistry 46,1050-1059.
- [12] Yulianton, Heribertus. "Data Mining For Business", 2008.
- [13] Wahyudi, Eko Nur. "Analisa Profil Data Mahasiswa Baru terhadap Program Studi yang dipilih di Perguruan Tinggi Swasta Jawa Tengah dengan Menggunakan Teknik Data Mining", 2011.
- [14] Wahyudi, Eko Nur. "Analisa Profil Data Mahasiswa Baru UNISBACK Semarang Tahun 2005-2010 Dengan Teknik Data Mining", 2010.
- [15] Mujiono. "Studi Awal Perkiraan Perilaku Konsumsi BBM Bersubsidi Di Kabupaten Kota Menggunakan Data Mining". 2013.