International Journal of Electrical Engineering, Mathematics and Computer Science

E-ISSN: 3048-1945
P-ISSN: 3048-1910

Research Article

Comparison of Multiple Linear Regression, Backpropagation and Fuzzy Mamdani Methods in Predicting the Revenue of PLN Takengon Unit

Richasanty Septima[1*], Hendri Syahputra[2], Husna Gemasih[3]

[1-3] Faculty of Engineering, Gajah Putih University, Indonesia, richaseptima@gmail.com

*Corresponding Author: Richasanty Septima

Abstract: The performance of data mining techniques has been proven accurate in many studies, but each method in data mining techniques has different accuracy depending on the type of data that is the object of research. Methods in data mining techniques are divided into several functions, namely: clustering, association, classification, and prediction, where each data mining technique objective has a superior method. Therefore, in this case the author will compare the performance of the multiple linear regression method, and neural networks with fuzzy mamdani in predicting the income of PLN Unit Takengon. In several studies, the Backpropagation method shows the highest accuracy compared to other methods. Then the prediction model with multiple linear regression also has the highest accuracy as well as the Fuzzy Mamdani method has high accuracy too. Therefore, the purpose of this study is to compare the three methods, so that it can be determined which method has a higher accuracy value. The results of this study indicate that the Back propagation method has the highest accuracy and the lowest average error, namely a MAPE value of 5.9% with an accuracy of 94.1% and an RMSE of 14398.14, followed by the multiple linear regression method obtaining a MAPE value of 6.9% with an accuracy of 93.1% and an RMSE of 15527.41, then for Fuzzy Mamdani obtaining a MAPE value of 7% with an accuracy of 93% and an RMSE of 16077.76.

Keywords: Back propagation, Data Mining, Fuzzy Mamdani, Multiple Linear Regression, Prediction.

Received: 13 April, 2025
Revised: 15 May, 2025
Accepted: 09 June, 2025
Published: 11 June, 2025
Curr. Ver.: 11 June, 2025

1. Introduction

Prediction can be done on income data to estimate and predict the possible amount of income in the future. One technique that can be used in predicting is data mining techniques. Data mining or also called Knowledge Discovery in Database (KDD) is the process of extracting information contained in a large database through the process of collecting, cleaning, processing, and analysis, so that something useful is found from a set of data. Data mining itself can solve problems from patterns of computer learning identification results or also called machine learning.

The performance of data mining techniques has been proven accurate in many studies, but each method in data mining techniques has different accuracy depending on the type of data that is the object of research. Methods in data mining techniques are divided into several functions, namely: clustering, association, classification, and prediction, where each data mining technique objective has a superior method. For example, in the case of numerical

prediction of time-series data, the neural network and multiple linear regression methods are superior (Maulida, 2022) (Winnos, 2022). But what if the two methods are compared with a logic-based method, namely the fuzzy mamdani method.

Fuzzy mamdani is a branch of artificial intelligence that emulates human thinking ability into an algorithm that is then run by a machine. Where the algorithm in the form of a rule in the fuzzy inference system is very dependent on the ability and expertise of the human being to be able to produce good logical accuracy. The best way that is commonly used is to increase the membership characteristics of the fuzzy set, so that the more membership characteristics, the better the chance of getting the best results can be achieved. This fuzzy method is very different from data mining techniques, where in data mining techniques the machine is the key to learning

2. Literature Review

Prediction

Prediction is one of the most common applications of time series analysis. Prediction of future trends has applications in retail sales, economic indicators, weather forecasting, stock markets, and many other application scenarios. In this case, we have one or more series of data values, and we want to predict the future values of the series using the history of previous values (Aggarwal, 2015).

In summary, there are three stages of prediction that must be passed in designing a prediction method, namely (Wardhono et al., 2019):

a. Conmducting analysis on past data. This step aims to obtain a pattern of the data concerned.
b. Choosing the method to be used. There are various methods available with their needs. The choice of method can affect the prediction results. Prediction results are measured by calculating the smallest error or mistakes.
c. The transformation process of past data using the selected method. Do you agree with other researchers' arguments and conclusion? If not, why?

3. Proposed Method

Data Mining

According to Han et al., (2012), data mining is knowledge mining from data, although in the short term, knowledge mining may not reflect the emphasis on mining from large amounts of data. According to Aggarwal, (2015), data mining is the science that studies the collection, cleaning, processing, analyzing, and gaining insight from data. Wide variations exist in terms of problem domains, applications, formulations, and data representations encountered in real applications. It is also called "data mining" because the term is commonly used to describe different aspects of data processing.

The knowledge discovery process is shown as an iterative sequence of the following steps (Han et al., 2012):

a. Data cleaning and integration is to remove noise and inconsistent data and also merge data from multiple sources.

b. Data selection and transformation is where data relevant to the analysis task is taken from the database and transformed and consolidated into a form suitable for mining by performing summary or aggregation operations.
c. Data mining is an important process where intelligent methods are applied to extract data patterns.
d. Pattern evaluation and presentation is to identify truly interesting patterns that represent knowledge based on the measure of interest and the knowledge gained is visualized to the user.

1. Results and Discussion

Multiple Linear Regression Method

Linear Regression involves finding the "best" line to fit two attributes (or variables) so that one attribute can be used to predict the other. Multiple linear regression is an extension of linear regression, where more than two attributes are involved and the data is fit to a multidimensional surface (Han, et al., 2012).Finding appropriate literature on a specific topic.

In this study, the regression used is multiple linear regression. The multiple linear regression formula is shown in the following equation (1) (Prasetyo, et al., 2021):

$$\boldsymbol{y = a + b_1x_1 + b_2x_2 + \cdots + b_nx_n} \quad \text{.....(1)}$$

Where y is the dependent variable, which depends on the value of X (independent variable). The value of a is a constant and b is a regression coefficient of the variable X. To obtain the values of a and b against the value of the variable X, it can be written in equations (2) and (3).

$$\boldsymbol{a} = \frac{(\sum \boldsymbol{y})(\sum \boldsymbol{x^2}) - (\sum \boldsymbol{x})(\sum \boldsymbol{xy})}{\boldsymbol{n}(\sum \boldsymbol{x^2}) - (\sum \boldsymbol{x})^2} \quad \textbf{.....(2)}$$

$$b = \frac{n(\sum xy) - (\sum x)(\sum y)}{n(\sum x^2) - (\sum x)^2} \quad \text{.....(3)}$$

Backpropagation Neural Network (NN)

A neural network is a set of connected input/output units where each connection has a weight associated with it. During the learning phase, the neural network adjusts the weights so that it can predict the correct class label from the input tuples. Neural network learning is also called connectionist learning because of the connections between the units (Han, et al. 2012).

Neural networks are also a simulation model of the human nervous system. The human nervous system is made up of cells called neurons. Biological neurons are connected to each other at points of contact, called synapses. Learning occurs in living organisms by changing the strength of synaptic connections between neurons. Typically, the strength of these connections changes in response to external stimuli. Neural networks can be thought of as a simulation of this biological process (Aggarwal, 2015).

Fuzzy Logic

Fuzzy logic is a problem-solving control system methodology, which is suitable for implementation in systems ranging from simple systems, small systems, embedded systems, PC networks, multichannel or workstation-based data acquisition and control systems (Wibowo, 2015).

There are several things that are the basis for understanding Fuzzy Logic, including:

1. Fuzzy variables, which are variables that will be discussed in a fuzzy system.
2. Fuzzy sets, which are groups that represent a certain condition in a fuzzy variable. Fuzzy sets have 2 attributes, namely linguistic and numeric, while the characteristics of the set itself are largely determined by its membership function.
3. Universe of discourse, namely all values that are permitted to be operated on in a fuzzy variable.

To solve the problems in this study, the author created a research methodology scheme using the data mining / knowledge discovery in database (KDD) method, where the research process flow begins with conducting a literature study to add and understand the theory related to data mining techniques, then carrying out the data cleaning and integration stages, data selection and transformation. In the data mining process, three methods are used, namely multiple linear regression, backpropagation neural network, and fuzzy mamdani.

After the method is applied, the next process is evaluation using the root mean square method and mean absolute percentage error. Then the last is the conclusion of the best accuracy of the method used. The planned stages that will be applied are as shown in Figure 1 below:

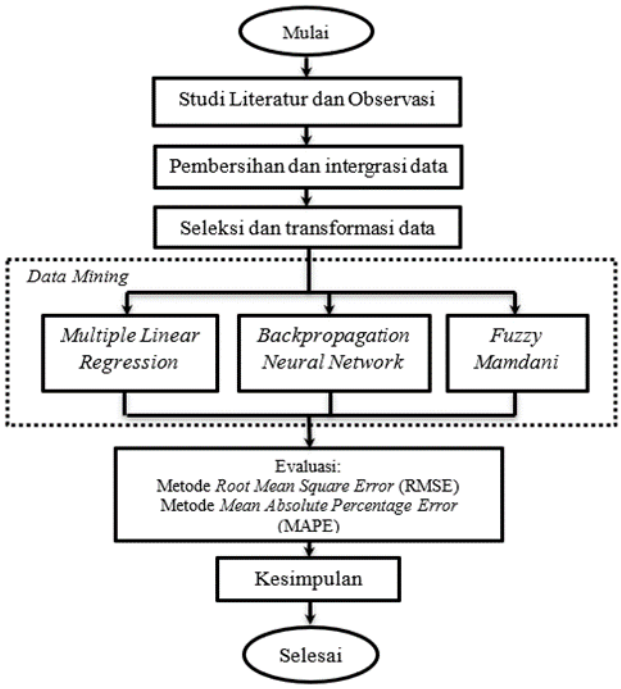


Figure 1. Research methodology flow

There are 4 (four) stages in the multiple linear regression method as can be seen in Figure 2 below:

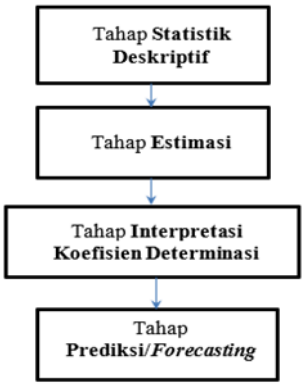


Figure 2. Stages of the Multiple Linear Regression Method

The training process in backpropagation neural networks aims to determine the best weights in each layer with several process stages, namely divided into 3 (three) stages, which can be seen in Figure 3. below:

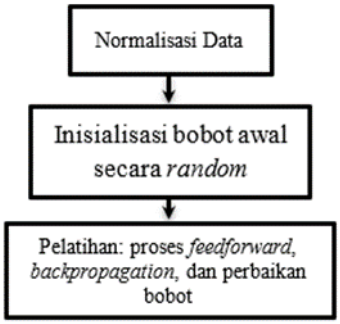


Figure 3. Stages of the NN method

The fuzzy mamdani method has process stages according to figure 4 below. The stages are as follows:

Figure 4. Stages of the Mamdani fuzzy method

The analysis process for the PLN Takengon Unit revenue dataset uses the RStudio application as follows:

1. Multiple linear regression

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 3.821e+09  4.779e+08   7.996 3.28e-11 ***
BULAN       7.123e+06  1.426e+07   0.500    0.619
KWH         1.027e+02  8.658e+01   1.186    0.240
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 397700000 on 64 degrees of freedom
Multiple R-squared:  0.0266,    Adjusted R-squared:  -0.003815
F-statistic: 0.8746 on 2 and 64 DF,  p-value: 0.422
```

Figure 5. Results of regression modeling

2. *Back propagation*

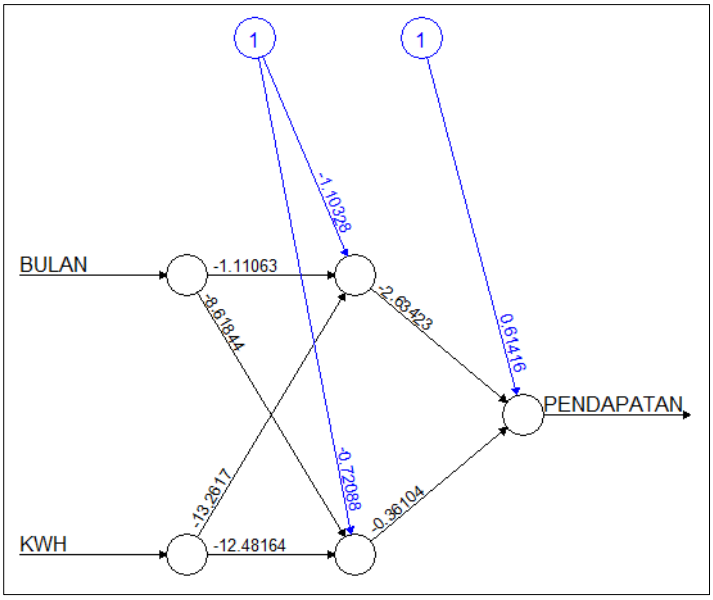
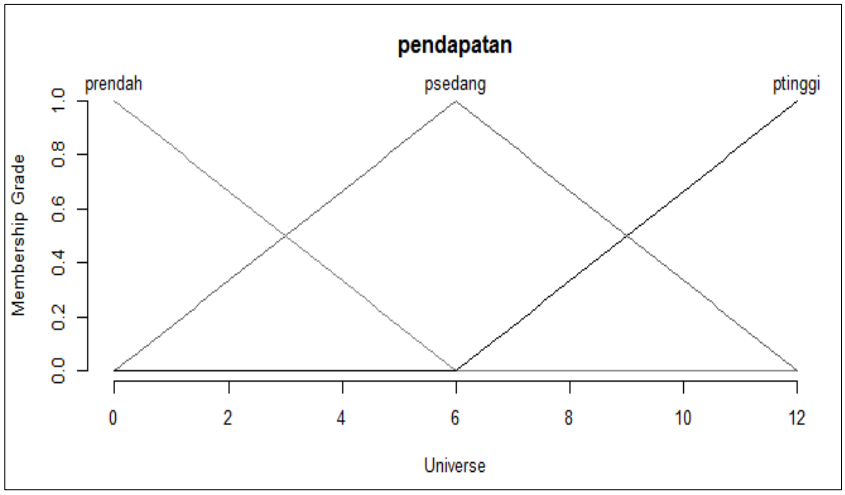


Figure 6. Results of backpropagation modeling

3. Fuzzy Mamdani



Gambar 7. Hasil fungsi keanggotaan variabel pendapatan

In this discussion, the final stage will be carried out, namely forecasting or prediction of the dataset used. The prediction is divided into three parts, namely prediction with multiple linear regression, backpropagation, and fuzzy mamadani.

1. Testing with the multiple linear regression method

   Based on the results obtained, it can be explained that the MAPE value obtained is 6.90281% or rounded to 6.9% so that based on the MAPE rule, the prediction has high accuracy (if the absolute value of the percentage <10%), while the RMSE value obtained is 15527.41.
2. Testing with the backpropagation method

   Based on the results obtained, it can be explained that the MAPE value obtained is 5.881393% or rounded to 5.9% so that based on the MAPE rule, the prediction has high accuracy (if the absolute value of the percentage <10%), while the RMSE value obtained is 14398.13763.
3. Testing with fuzzy mamadani

   Based on the results obtained, it can be explained that the MAPE value obtained is 7.035069499% or rounded to 7% so that based on the MAPE rule, the prediction has high accuracy (if the absolute value of the percentage <10%), while the RMSE value obtained is 16077.76.

Thus, based on the test results in tables 4.6, 4.8 and 4.9, a comparison of the accuracy of each method can be produced based on the MAPE and RMSE indicators as can be seen in the following table:

Table 1 Comparison of test results

Metode	MAPE	RMSE
Multiple Linear Regression	6,9%	15527.41
Back propagation	5,9%	14398.14
Fuzzy Mamdani	7,0%	16077.76

Based on table 1, it can be seen that the comparison shows that the smallest MAPE value was obtained by the Backpropagation method 5.9%, followed by the Multiple Linear Regression method 6.9% and the Fuzzy Mamdani method 7.0%, so it can also be explained in the form of a graph as below.

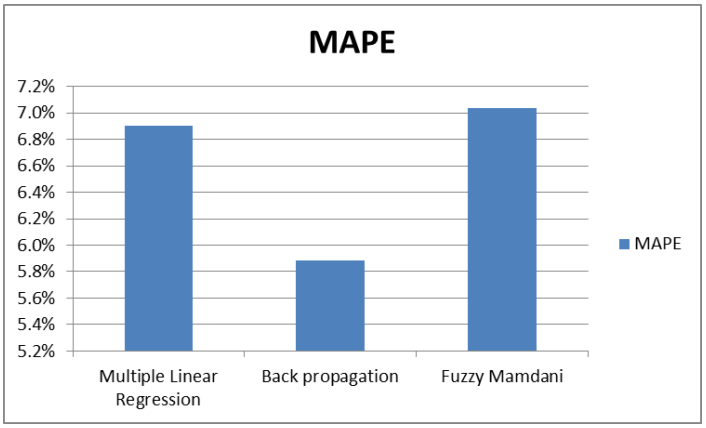


Figure 8. Comparison graph of MAPE values

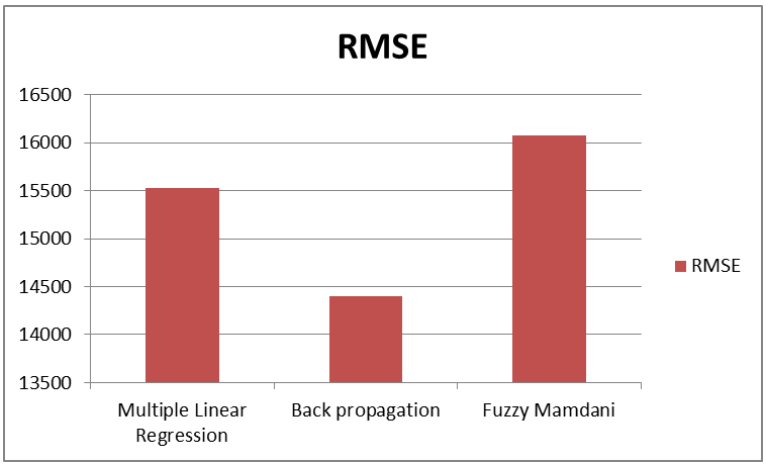


Figure 9. Comparison graph of RMSE values

Based on the graphs in Figures 4.15 and 4.16 above, it can be concluded that the accuracy value of the Backpropagation method if 100% -5.9%, then it is equal to 94.1% then followed by the Multiple Linear Regression method if 100% -6.9%, then it is equal to 93.1% and the Fuzzy Mamdani method if 100% -7.0%, then it is equal to 93.0%.

5.Conclusion

Based on the test results of the PLN Takengon Unit's revenue prediction, the conclusions drawn are: The best equation pattern in predicting PLN Takengon Unit revenue using the multiple linear regression method is the intercept value or b_0=3821287632, the coefficient value for month or b_1=7123213, and the coefficient value for kWh or b_2=102.672. Meanwhile, the best model with the back propagation method is 2 input layers (weight values v_1 = -1.11063, v_2 = -12.48164, v_3 = -8.61844, v_4 = -13.2617), plus 1 bias layer (weight values v_j1 = -1.10328, v_j2 = -0.72088), then 2 hidden layers (weight values z_1 = -2.63423, z_2 = -0.36104), plus 1 bias layer (weight value w_j = 0.61416), and 1 output layer.

The best algorithm rule results in predicting PLN Takengon Unit revenue using the fuzzy mamdani method obtained 8 rulesI, namely:

[R1] If Kwh = Low, then Income = Low
[R2] If Month = Early and Kwh = Medium, then Income = Medium
[R3] If Month = Early and Kwh = Medium, then Income = High
[R4] If Kwh = High, then Income = High
[R5] If Month = Middle and Kwh = Medium, then Income = Medium
[R6] If Month = Mid and Kwh = Medium, Then Income = High
[R7] If Month = End and Kwh = Medium, Then Income = Medium
[R8] If Month = End and Kwh = Medium, Then Income = High

The multiple linear regression method obtained a MAPE value of 6.9% with an accuracy of 93.1% and an RMSE of 15527.41, for Backpropagation, it obtained a MAPE value of 5.9% with an accuracy of 94.1% and an RMSE of 14398.14, while for Fuzzy Mamdani it obtained a MAPE value of 7% with an accuracy of 93% and an RMSE of 16077.76, so the method with the highest accuracy and the lowest average error is Back propagation.

References

[1] Aggarwal, C. C. *Data Mining.* New York: Springer, 2015.

[2] Decky Irmawan, Khamami Herusantoso " Penerapan Logika Fuzzy sebagai Sistem Penunjang Keputusan Prakiraan Cuaca", e-Indonesia Initiative, Bandung, 2011.

[3] Faisal, M Reza dan Dodon T. Nugrahadi. *Belajar Data Science: Klasifikasi dengan Bahasa Pemrograman R*. Banjar Baru: Scripta Cendekia, 2019.

[4] Han, Jiawei dkk. *Data Mining: Concepts And Techniques 3rd Edition*. United States: Morgan Kaufmann. 2012.

[5] Kharomah, Siti Ismiatul, dkk., "Implementasi Model Fuzzy-Wavelet Dan Fis Metode Mamdani Dalam Prediksi Nilai Tukar Eur/Idr". Journal of Mathematics, Vol. 8 (2) (2019). Hal: 60-68

[6] Kusumadewi, Sri, dan Hari Purnomo. *Aplikasi Logika Fuzzy Untuk Pendukung Keputusan*. Yogyakarta: Graha Ilmu, 2010

[7] Laila dan Cahyono,"Estimasi Data Insight Social Media Ads Menggunakan Neural Network, Linear Regression dan Deep Learning" Jurnal Sistem Komputer dan Informatika Vol 4 (3) (2023) Hal:562-569

[8] Larose, T. Daniel. *Discovering Knowledge In Data: An Introduction to Data Mining*. Canada: Jhon Wiley & Sons, Inc. 2005.

[9] Maulida, Rizki, *Perbandingan metode K-Nearest Neighbor (K-NN) dan Neural Network (NN) dalam prediksi harga saham PT. Telkom Indonesia Tbk (TLKM)*. Teknik Informatika, Fakultas Teknik Universitas Gajah Putih. Takengon, 2022.

[10] Maulidiyah, Salasabila Talitha. *Perbandingan Metode Ann-Backpropagation Dan Linear Regression Yang Telah Dioptimasi Untuk Prediksi Angka Kemiskinan Di Jawa Barat*. Teknik Informatika, Fakultas Sains dan Teknologi Universitas Negeri Malang 2023.

[11] Munir, Rinaldi. *Matematika Diskrit*. Bandung: INFORMATIKA, 2012

[12] Rahmaddeni, Rahmaddeni. "Penerapan Fuzzy Logic dalam Menganalisis Tingkat Pendapatan Akhir Konsultan Produk Multi Level Marketing (Studi Kasus: PT. Orindo Alam Ayu Cabang Pekanbaru)." *Jurnal Sains dan Teknologi Industri* 11, no. 2 (2015) Hal: 192–199.

[13] Wardhono, Adhitya dkk. *Analisis Data Time Series Dalam Model Makroekonomi*. Jawa Timur: CV. Pustaka Abadi. 2019.

[14] Wibowo, S. Penerapan Logika *Fuzzy* Dalam Penjadwalan Waktu Kuliah. *Informatika UPGRIS*, *1*(Juni) (2015), 59–77

[15] Winnos, Hilman. *Perbandingan Metode Regresi Linier dan ARIMA Untuk Prediksi Saham PT. BSI, Tbk*. Teknik Informatika, Fakultas Teknik Universitas Gajah Putih. Takengon, 2022.