Comparison of Fuzzy Logic and Multiple Linear Regression in Forecasting Rice Production in Toba District

Agustaeys Pasaribu¹, M.R Syahputra²

¹Student of Mathematics Study Program, Faculty of Mathematics and Natural Sciences, Universitas Sumatera Utara Medan-Indonesia 20155
²Mathematics Undergraduate Study Program, Faculty of Mathematics and Natural Sciences, Universitas Sumatera Utara Medan-Indonesia 20155

Email: ¹agustaeysp1000@gmail.com

ABSTRACT

The purpose of this study is to forecast rice production in Toba Regency in the future. Factors affecting rice production studied are land area, subsidized fertilizers, rice pest populations, average rainfall on rainy days. The methods used in this study are fuzzy logic methods and linear regression. This study compared more accurate forecasting results based on errors from both methods. The data used is secondary data from the Toba Regency Agriculture Office. The data studied were in the form of data on rice production, land area, subsidized fertilizers, pest populations, average rainfall and rainy days in 2016-2021. This research was conducted with the help of matlab toolbox fuzzy and SPSS software. From the forecasting results obtained, the multiple linear regression method is closer to the actual result of 878,428.07 Tons compared to using fuzzy logic of 1,032,300 Tons. Based on the error standards of the two methods, it can be concluded that the multiple linear regression method has a more accurate result with an error of 3.11% compared to the fuzzy logic method with an error of 22.05%.

Keywords: Rice, Production, Fuzzy Logic, Linear Regression, Standard Error.
A. Pendahuluan
Forecasting is the activity of predicting the future value of the value obtained in the previous period. Forecasting is the ability to minimize errors or errors when predicting a problem optimally. The inability to accurately predict future events can be a big risk, for which strong forecasting is needed. One of them is the selection of the right forecasting method with high accuracy and minimal error (Arfina & Tomi, 2018). A commonly used method in forecasting is linear regression. The linear regression method is used to form the equations of several free variables that are considered to relate to bound variables. In addition to multiple linear regression the method used to foresee a future event is fuzzy logic. Fuzzy logic has the advantage of foreseeing future production with reasonings in the data of previous periods (Yulia R.S., dkk, 2021).

One of the rice producing areas is Toba Regency. Toba Regency is one of the regions in North Sumatra where the majority of people's income is from agricultural products, one of which is as a rice farmer. The Toba Regency area is suitable as a rice field because the area is in the mountains and there are many small rivers. In 2016, Toba Regency had a rice field area of 22,895 ha with a production of 146,701 Tons. Meanwhile, in 2021, Toba Regency is only able to produce rice production of 147,001.5 Tons with an area of 22,450.6 ha (tobakab.go.id, 2022).

The area of rice sowing land is one of the main factors of rice production, the wider the rice planting land, the greater the probability of the yield produced. In addition to land area, the triggering factor that causes rice production is the use of fertilizers. In this study, the fertilizer in question is the availability of government subsidized fertilizers. Many of the farmers do not get subsidized fertilizer from the government. Meanwhile, the price of non-subsidized fertilizers continues to increase in prices so that farmers are overwhelmed to fertilize rice crops. This makes the harvest obtained unsatisfactory (Yosefina, Made & Dewa, 2018). In addition to fertilizers, the disruptive factor of rice plants is plant pests. Pests of rice plants often become destroyers or make rice farmers fail to harvest, such as rats, sparrows, caterpillars, grasshopper and so on (Tri, Made & Lien, 2019). Another factor to be studied in this study is the average rainfall and the number of rainy days. Rainfall and rainy days that are too high or too low can cause less than perfect rice growth (Isnain, dkk 2021).

This study predicted rice production in Toba Regency using multiple linear regression methods and fuzzy logic. Factors affecting rice production studied were land area ($X_1$), the use of subsidized fertilizers ($X_2$), rice pests ($X_3$), average rainfall ($X_4$) and rainy days ($X_5$). The fuzzy logic method used is the Mamdani method, and the linear regression method used is the least squared. The composition of the rules that is often used is the Max method. The solution of the Max method is the maximum value of the fuzzy inference of the membership function. Fuzzy inference is the merging of several rules from a fuzzy set. The Max method can be formulated in the following equation 2. The composition of the rules that is often used is the Max method. The solution of the Max method is the maximum value of the fuzzy inference of the membership function. Fuzzy inference is the merging of several rules from a fuzzy set. The Max method can be formulated in the following equation 2. multiple regression. From the results obtained, it can be known which method is better to predict the outcome.

B. Metode Penelitian
1. Fuzzy Mamdani
The fuzzy mammadi method is a forecasting method used to estimate the future based on previous data. The Mamdani method is often also referred to as the Max-Min method. In 1975 Ebrahim Mamdani began to introduce the mammadi method (Aulia Ishak, 2010). To get a good result or output, there are 4 basic stages that must be considered:

1. Formation of fuzzy sets
The formation of a fuzzy set begins with determining the input and output variables of the set, determining the universe of speech and dividing the domain of each variable.

2. Application of the fuzzy implication function of mammadi
An implication function that is often used in research is the Min function. The general form of the implication function is as follows:

\[
IF \ x \ is \ A \ THEN \ y \ is \ B
\]
3. Create a combination or rule composition

The composition of the rules that is often used is the Max method. The solution of the Max method is the maximum value of the fuzzy inference of the membership function. Fuzzy inference is the merging of several rules from a fuzzy set. The Max method can be formulated in the following equation 2.

\[ \mu_{sf}(x_i) = (\text{max } \mu_{sf}(x_i), \mu_{kf}(x_i)) \]

4. Affirmation function (defuzzyfication)

The affirmation function is the output of an assertive number in the fuzzy set domain. The affirmation function is obtained from the regional solution of the composition of the rules. The affirmation function used is the centroid method. The general form of the centroid method is as follows:

\[ y^* = \frac{\int_a^b y \mu(y) dy}{\int_a^b \mu(y) dy} \]

2. Regresi Linier Berganda

Multiple linear regression is a test method used to find out how the influence or relationship of free variables affects the production of bound variables. The results of forecasting using the method of multiple linear regression are obtained by forming an equation that connects free variables and bound variables (Yulia, dkk., 2019). The equation of multiple linear regression can be formulated in the following equation.

\[ \hat{Y}_i = b_0 + b_1X_1 + b_2X_2 + \cdots + b_nX_n + \varepsilon_i \]

The stages to solve the multiple regression equation and the testing of the function of the multiple linear regression coefficient consisting of 5 independent variables and 1 bound variable include:

1. It starts by searching and determining the price of each variable to get the price of \( b_0, b_1, b_2, b_3, b_4 \) and \( b_5 \).
2. Create a multiple linear regression equation of 5 independent variables as in the general formula above:

\[ \hat{Y}_i = b_0 + b_1X_1 + b_2X_2 + b_3X_3 + b_4X_4 + b_5X_5 \]

3. Standard error

Standard error is a relative error in forecasting calculations. Here is a standard description of the error of each method:

1. Standar error fuzzy mamdani

\[ S_e = \sqrt{\frac{s^2}{n}} \]

2. Standar error regresi linier

\[ S_e = \sqrt{\frac{\sum(Y - \hat{Y}_i)^2}{n-k-1}} \]

3. Then the standard percentage of error can be calculated with the formula:

\[ \text{Percentage} = \frac{S_e}{\overline{X}} \times 100\% \]

C. Hasil dan Pembahasan

The data used for this study were data on rice production, land area, availability of subsidized fertilizers, pest population, average rainfall and rainy days within 6 years (tobasamosir kab.go.id, 2022).

Table 1 Production Data, Land Area, Subsidized Fertilizers and Rice Pests 2016-2021.

<table>
<thead>
<tr>
<th>Year</th>
<th>Rice Product (Ton)</th>
<th>Land Area (Ha)</th>
<th>Fertilizer Subsidy (Ton)</th>
<th>Pests (Ha)</th>
<th>Rain fall (mm)</th>
<th>Rainy Day</th>
</tr>
</thead>
<tbody>
<tr>
<td>January-April 2016</td>
<td>14441</td>
<td>2208</td>
<td>3348</td>
<td>116.5</td>
<td>79</td>
<td>48</td>
</tr>
<tr>
<td>May-August 2016</td>
<td>12531</td>
<td>1944</td>
<td>2853.1</td>
<td>109.5</td>
<td>90.5</td>
<td>55</td>
</tr>
<tr>
<td>September-December 2016</td>
<td>13472</td>
<td>2315</td>
<td>3854.9</td>
<td>102.5</td>
<td>138</td>
<td>61</td>
</tr>
<tr>
<td>January-April 2017</td>
<td>22254</td>
<td>3854</td>
<td>2920</td>
<td>76</td>
<td>132.75</td>
<td>73</td>
</tr>
<tr>
<td>May-August 2017</td>
<td>113518</td>
<td>1719</td>
<td>1697</td>
<td>73.5</td>
<td>99.5</td>
<td>52</td>
</tr>
<tr>
<td>September-December 2017</td>
<td>24653</td>
<td>3780</td>
<td>3609</td>
<td>77</td>
<td>178.75</td>
<td>84</td>
</tr>
<tr>
<td>January-April 2018</td>
<td>14366.74</td>
<td>2867.70</td>
<td>3369</td>
<td>39</td>
<td>124.75</td>
<td>30</td>
</tr>
<tr>
<td>May-August 2018</td>
<td>105523.95</td>
<td>16574.70</td>
<td>1938</td>
<td>47</td>
<td>70.5</td>
<td>39</td>
</tr>
<tr>
<td>September-December 2018</td>
<td>31392.84</td>
<td>4819.30</td>
<td>2884</td>
<td>46</td>
<td>210</td>
<td>75</td>
</tr>
<tr>
<td>January-April 2019</td>
<td>27272.93</td>
<td>3832.20</td>
<td>3759</td>
<td>80.5</td>
<td>140</td>
<td>52</td>
</tr>
<tr>
<td>May-August 2019</td>
<td>99419.11</td>
<td>15906.25</td>
<td>2128</td>
<td>60.5</td>
<td>195.75</td>
<td>62</td>
</tr>
<tr>
<td>September-December 2019</td>
<td>23491.84</td>
<td>3725.10</td>
<td>2716</td>
<td>95.5</td>
<td>120.15</td>
<td>54</td>
</tr>
<tr>
<td>January-April 2020</td>
<td>13923.34</td>
<td>3400.40</td>
<td>2908</td>
<td>43.5</td>
<td>83.75</td>
<td>55</td>
</tr>
<tr>
<td>May-August 2020</td>
<td>97651.11</td>
<td>15559.10</td>
<td>2510.5</td>
<td>89</td>
<td>121.75</td>
<td>62</td>
</tr>
<tr>
<td>September-December 2020</td>
<td>15973.77</td>
<td>2819.40</td>
<td>4158.5</td>
<td>38.5</td>
<td>171</td>
<td>71</td>
</tr>
<tr>
<td>January-April 2021</td>
<td>23106.56</td>
<td>3908.10</td>
<td>3437</td>
<td>107.5</td>
<td>119</td>
<td>42</td>
</tr>
<tr>
<td>May-August 2021</td>
<td>92019.54</td>
<td>3046.10</td>
<td>2220</td>
<td>115.5</td>
<td>158.5</td>
<td>53</td>
</tr>
<tr>
<td>September-December 2021</td>
<td>25721.95</td>
<td>3589.00</td>
<td>3389</td>
<td>110.5</td>
<td>210</td>
<td>68</td>
</tr>
<tr>
<td>Total</td>
<td>879,552</td>
<td>141,555</td>
<td>54,117</td>
<td>1,333</td>
<td>2,508</td>
<td>1,073</td>
</tr>
</tbody>
</table>

1. Fuzzy Set Formation

Based on Table 1, the sum of the data consists of 18 periods for which the universe value will be determined between the lowest value and the highest value. It can then be divided the domains of each fuzzy set.
The implication function used is the min (minimum) function. In this study, 74 fuzzy rules were used based on the data in table 1. As an example, you can see the fuzzy rules in the data for the January-April 2016 period as follows:

**R1** IF Land Area IS SMALLEST AND Fertilizer Subsidied IS MEDIUM AND Pest Population IS HIGHEST AND Rainfall IS LOWEST AND Rainy Days IS SMALLEST THEN Production IS DECREASED

**R2** IF Land Area IS SMALLEST AND Fertilizer Subsidied IS MOST AND Pest Population IS HIGHEST AND Rainfall IS LOWEST AND Rainy Days IS SMALLEST THEN Production IS DECREASED

In table 1, the land area for the January-April 2016 period is 2,808 Ha, subsidized fertilizers are 3,438 tons, pest populations are 116.5 Ha, average rainfall is 79 mm and rainy days are 48 days, then the membership function of each variable can be calculated, including:

1. For a land area of 2,808 Ha, it is calculated:
   \[
   \mu_x_{Small}(2,808) = \frac{2,105 - 2,808}{1,444} = 0.92
   \]
   \[
   \mu_x_{Medium}(2,808) = 0
   \]
   \[
   \mu_x_{Most}(2,808) = 0
   \]

2. For subsidized fertilizers as many as 3,438 then:
   \[
   \mu_x_{Small}(3,438) = 0
   \]
   \[
   \mu_x_{Medium}(3,438) = \frac{2,097.75 - 3,438}{2,097.75} = 0.91
   \]
   \[
   \mu_x_{Most}(3,438) = 0
   \]

3. For a pest population of 116.5 Ha then:
   \[
   \mu_x_{Low}(116.5) = 0
   \]
   \[
   \mu_x_{Medium}(116.5) = 0
   \]
   \[
   \mu_x_{High}(116.5) = 1
   \]

4. For an average rainfall of 79 mm then:
   \[
   \mu_x_{Low}(79) = \frac{144.25 - 79}{56.25} = 0.88
   \]
   \[
   \mu_x_{Medium}(79) = 0
   \]
   \[
   \mu_x_{High}(79) = 0
   \]

5. For a 48-day rainy day then:
   \[
   \mu_x_{Small}(48) = \frac{61.5 - 48}{22.5} = 1.35 = 0.6
   \]
   \[
   \mu_x_{Medium}(48) = 0
   \]
   \[
   \mu_x_{Most}(48) = 0
   \]

\[\alpha\text{-predikat}_1 = \mu_x_{Small} \cap \mu_x_{Medium} \cap \mu_x_{High} \cap \mu_x_{Low} \cap \mu_x_{Small} = min (\mu_x_{Small}(2,808), \mu_x_{Medium}(3,438), \mu_x_{High}(116.5), \mu_x_{Low}(79), \mu_x_{Small}(48)) = min (0.92;0.15;1;0.88;0.6) = 0.15\]

\[\alpha\text{-predikat}_2 = \mu_x_{Small} \cap \mu_x_{Most} \cap \mu_x_{High} \cap \mu_x_{Low} \cap \mu_x_{Small} = min (\mu_x_{Small}(2,808), \mu_x_{Most}(3,438), \mu_x_{High}(116.5), \mu_x_{Low}(79), \mu_x_{Small}(48)) = min (0.92;0.43;1;0.88;0.6) = 0.43\]

3. **Rule Composition**

The composition of the rules used is a non-zero min function. From the \(\alpha\)-predicate values obtained in the implication function above, \(\alpha\)-predicate2 is greater than \(\alpha\)-predicate1, then the composition of the rules can be calculated as follows:

\[\alpha\text{-predikat}_2 = min (0.92;0.43;1;0.88;0.6) = 0.43\]

\[\mu_y\text{Reduced} = \frac{68.99 - y}{54.54} = 0.43\]
\[ y = 68,990 - 54,546x + 0.43 = 45,535.22 \]

So that the fuzzy area solutions for the period January-April 2016 include:
\[
\mu_{\text{Reduced}} = \begin{cases} 
\frac{1}{(68,990 - y)} & ; \quad y \leq 14,444 \\
\frac{54,546}{68,990 - y} & ; \quad 14,444 < y < 68,990 \\
0 & ; \quad y \geq 68,990 
\end{cases}
\]

\[ \text{Figure 2} \quad \text{Fuzzy area solution for the period January-April 2016} \]

The affirmation function used in fuzzy mamdani is the centroid method. To calculate the affirmation function in the period January-April 2016, it is calculated the moment (M) and area of the area (A) to calculate the value of the center point (y*) from figure 2. The first step is to calculate the \( M_1 \) moment and \( M_2 \):

1. The first inference is a linear function, then it can be calculated:
   \[ M_1 = \int_{14,444}^{45,535.22} 0.43y \, dy \]
   \[ M_2 = \int_{14,444}^{45,535.22} \frac{0.43y^2}{2} \, dy \]
   \[ M_1 = 0.215(45,535.22) - 0.215(14,444)^2 
   \]
   \[ M_2 = 400,937,832 \]

2. The second inference is the descending function, then it can be calculated:
   \[ M_2 = \int_{14,444}^{45,535.22} \frac{68,990 - y}{2 \times 54,546} \, dy \]
   \[ M_2 = \frac{68,990}{2 \times 54,546} - \frac{68,990}{3 \times 54,546} \]
   \[ M_2 = (68,990 - 109,092)^2 \]
   \[ M_2 = 11,727.39 \]

Furthermore, the central point is calculated by summing the \( M_1 + M_2 \) divided by \( A_1 + A_2 \), then obtained:
\[ y^* = \frac{400,937,832 + 269,049,739}{13,369,22 + 11,727.39} \]
\[ = 36,388,63172 \]

The value of the central point (y*) obtained was 36,388,63172, meaning that the optimum rice production in the January-April 2016 period was 36,388,63172 Tons. Figure 3 is the result of testing rice production in the period January-April 2016 using the matlab toolbox (Sri Kusumadewi, 2002).

\[ \text{Figure 3} \quad \text{Production output for the period January-April 2016 by centroid method} \]

4. Multiple Linear Regression Equation

The normal equation of 5 free variables is \( \hat{Y}_i = b_0 + b_1X_1 + b_2X_2 + b_3X_3 + b_4X_4 + b_5X_5 \). With the help of the SPSS program, the value of the regression coefficient can be calculated from the table data in table 3, including:

\[ \text{Table 3} \quad \text{Regression coefficient of 5 independent variables} \]

<table>
<thead>
<tr>
<th>Model</th>
<th>B</th>
<th>Std. Error</th>
<th>Standardized Coefficients</th>
<th>Beta</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (Constant)</td>
<td>-3711.185</td>
<td>4800.976</td>
<td>0.001</td>
<td>0.001</td>
</tr>
<tr>
<td>Land Area</td>
<td>6.449</td>
<td>.097</td>
<td>1.008</td>
<td>0.008</td>
</tr>
<tr>
<td>Subsidy</td>
<td>.372</td>
<td>.940</td>
<td>.006</td>
<td>.006</td>
</tr>
<tr>
<td>Fertilizer</td>
<td>-1.045</td>
<td>14.465</td>
<td>-.001</td>
<td>-.001</td>
</tr>
</tbody>
</table>

Copyright (c) 2022 FARABI: Jurnal Matematika dan Pendidikan Matematika
Furthermore, the value of the coefficient obtained is substitution into equation 11 so as to form a linear regression equation, including:

\[ \hat{Y}_i = b_0 + b_1 X_1 + b_2 X_2 + b_3 X_3 + b_4 X_4 + b_5 X_5 \]

\[ \hat{Y}_i = -4,800.976 + 6.449 X_1 + 0.372 X_2 - 1.045 X_3 - 4.83 X_4 + 42.243 X_5 \]

Than can be calculated multiple regression forecasting values \( \hat{Y}_i \) by substituting the values of variables in table 4.

**Table 4** Production results of fuzzy mamdani and regression

<table>
<thead>
<tr>
<th>Y</th>
<th>( Y_{fuzzy} )</th>
<th>( Y_{regresi} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>14,444</td>
<td>36,100</td>
<td>16,111.10</td>
</tr>
<tr>
<td>123,536</td>
<td>102,000</td>
<td>123,325.62</td>
</tr>
<tr>
<td>14,572</td>
<td>36,300</td>
<td>12,011.36</td>
</tr>
<tr>
<td>23,354</td>
<td>36,600</td>
<td>23,309.65</td>
</tr>
<tr>
<td>113,548</td>
<td>97,900</td>
<td>111,739.38</td>
</tr>
<tr>
<td>24,653</td>
<td>35,700</td>
<td>23,620.10</td>
</tr>
<tr>
<td>14,766.74</td>
<td>36,800</td>
<td>16,036.05</td>
</tr>
<tr>
<td>105,523.95</td>
<td>103,000</td>
<td>104,068.05</td>
</tr>
<tr>
<td>31,292.84</td>
<td>35,700</td>
<td>30,063.65</td>
</tr>
<tr>
<td>22,722.95</td>
<td>36,200</td>
<td>22,742.65</td>
</tr>
<tr>
<td>99,449.11</td>
<td>102,000</td>
<td>100,051.11</td>
</tr>
<tr>
<td>21,491.84</td>
<td>37,300</td>
<td>21,872.61</td>
</tr>
<tr>
<td>19,823.36</td>
<td>37,700</td>
<td>20,114.42</td>
</tr>
<tr>
<td>97,631.18</td>
<td>100,000</td>
<td>98,466.80</td>
</tr>
<tr>
<td>15,973.77</td>
<td>36,900</td>
<td>16,995.91</td>
</tr>
<tr>
<td>21,016.56</td>
<td>37,100</td>
<td>20,191.00</td>
</tr>
<tr>
<td>92,019.54</td>
<td>87,300</td>
<td>94,400.10</td>
</tr>
<tr>
<td>23,732.95</td>
<td>37,700</td>
<td>23,308.49</td>
</tr>
<tr>
<td><strong>879,552</strong></td>
<td><strong>1,032,300</strong></td>
<td><strong>878,428.07</strong></td>
</tr>
</tbody>
</table>

**5. Standar Error**

From the results of the production obtained, the error standards of the two methods can be calculated by the following equation:

**Table 5** Standard Error

<table>
<thead>
<tr>
<th>Standard Error</th>
<th>Fuzzy Mamdani</th>
<th>Linear Regression</th>
</tr>
</thead>
<tbody>
<tr>
<td>( S_e ) = \sqrt{ \frac{s^2}{n} }</td>
<td>( S_e ) = \sqrt{ \frac{\sum(y - \bar{y})^2}{n-k-1} }</td>
<td></td>
</tr>
<tr>
<td>( = \sqrt{ \frac{(41,729.89)^2}{18} } )</td>
<td>( = \sqrt{ \frac{27,826,328.97}{18-5-1} } )</td>
<td></td>
</tr>
<tr>
<td>( = 10,774.6112 )</td>
<td>( = 1,522.78 )</td>
<td></td>
</tr>
<tr>
<td>( \text{Percentage} = \frac{S_e}{X} )</td>
<td>( \text{Percentage} = \frac{S_e}{X} )</td>
<td></td>
</tr>
</tbody>
</table>

**D. Kesimpulan dan Saran**

1. Kesimpulan:

Using 74 propositions of rules, the result of forecasting using fuzzy mamdani logic reached 1,032,300 Tons. While the forecasting results using linear regression with the equation \( \hat{Y}_i = -4,800.976 + 6.449 X_1 + 0.372 X_2 + 1.045 X_3 - 0.83 X_4 + 42.243 X_5 \) is 878,428.07 Tons. Based on the standard error average, multiple linear regression forecasting has a smaller error of 1,522.78 with a percentage of 3.11%. Meanwhile, forecasting using the fuzzy logic method has an error of 10,774.61 with a percentage of 22.05%. It can be concluded that multiple linear regression forecasting has more accurate results than mamdani’s fuzzy logic.

2. Saran

For readers of this study, it is hoped that it is better to use data that tends to be more constant so that the forecasting results obtained are more accurate. If the data used is not constant, then the standard deviation or standard error value of the study is greater, resulting in the forecasting result far from the actual result.

For researchers who want to conduct further research, it is expected to use data on time intervals longer or a period of more than 6 years, so that the forecast obtained is more accurate.

**E. Daftar Pustaka**


tobakab.go.id. (2022). Retrieved Februari 17, 2022, from https://tobakab.go.id/

