Fuzzy Inference System Tsukamoto–Decision Tree C 4.5 in Predicting the Amount of Roof Tile Production in Kebumen

Tundo¹*, Fajar Mahardika²
1Departement of Informatics, Universitas 17 Agustus 1945 Jakarta, Indonesia
2Departement of Computer Technology, ITS NU Pekalongan, Indonesia
asna8mujahid@gmail.com¹, fajarmahardika@itsnupekalongan.ac.id²

ABSTRACT

Tile is a product that is in great demand by many people. This has become a trigger for producers to improve their management. The company's tile production management is still experiencing problems, namely frequent miscalculations in determining the agreement that must be issued in making tile production from customer requests. One of the efforts made is to predict the production that can be done to get the optimal amount obtained, to get a big profit. In this study, to obtain a prediction of the amount of tile production, computerized calculations were carried out using the Tsukamoto fuzzy logic method. This method uses the concept of rules from the C 4.5 decision tree in the building to make it easier to determine the rules that are built without having to consult an expert because C 4.5 will study existing datasets to serve as a reference in forming these rules according to conditions that often occur. The modeling results produce relevant rules after being compared with the actual results. The results of the comparison of predictions with actual production have an error percentage of 29.34%, with a truth of 70.66% (based on the calculation of the Average Forecasting Error Rate (AFER)). Therefore when implemented in the Tsukamoto Fuzzy Inference System it can produce predictions of tile production that are quite optimum. It is said to be quite optimum because all customer requests are met, either generated by the production prediction itself or the prediction results are added up with inventory data, and all predictions are close to actual production.

A. INTRODUCTION

Tile is in great demand by various groups (Wati, Rohmah, & Rahmadani, 2021). This condition provides many opportunities for producers to implement strategies to optimize their production (Ezhilmaran & Joseph, 2017). In the industrial era 4.0, the application of technology in various ways, can be done to increase the optimization of tile production. One of the processes carried out is predicting the amount of production from producers.

One of the efforts made in a prediction is by using the Fuzzy Logic method (Tundo & Sela, 2018). Fuzzy logic is logic with a value of similarity (fuzziness) with a value of true or false (Sheena, Ramalingam, & Anuradha, 2017). This concept was implemented and introduced by Lotfi Asker Zadeh in 1965 in fuzzy set theory (Solesvik, Kondratenko, & Kharchenko, 2017). There are several types of fuzzy belonging to the type of Fuzzy Inference System (FIS), namely Sugeno, Mamdani, and Tsukamoto. This study uses the Tsukamoto FIS concept – Decision tree C 4.5 which is used to implement. Use of decision tree C 4.5 to create rules that
are built based on collected datasets, then processed using WEKA. These two concepts make
the flexible, simple structure, tolerant of the data used, and speed up the creation of rules
without expert intervention (Tseng, Konada, & Kwon, 2016).

Several studies support this research, including the application of Mamdani FIS in
predicting the amount of woven fabric production (Tundo & Saifullah, 2022). Rules are made
automatically using Random Tree with the criteria used are production costs, stock, and
demand. The resulting accuracy shows values with results that are close to actual production
with an accuracy of 97%. But unfortunately not in detail explained how the concept of
Random Tree is used. Furthermore, other predictions are also made with the application of
FIS in predicting palm oil production (Tundo & ’Uyun, 2020) carried out using the Tsukamoto
method. The rules used are the results of the decision tree J48 and REPTree with the criteria:
the amount of oil palm, demand, and supply of palm oil. The decision tree shows values with
results that are close to actual production. However, the resulting classification accuracy is
lower than J48. Furthermore, there is a wind power prediction using the comparison of Fuzzy
Mamdani and Sugeno (Topaloglu & Pehlivan, 2018). The criteria used are wind speed, power
density, capacity factor, and suitability factor. The experimental results give a better result
using the Sugeno method compared to Mamdani. However, no detailed accuracy results have
been submitted yet, it should be explained so that it can be confirmed and seen.

Modeling in this study was carried out using decision tree rules C 4.5 which were then
processed with WEKA to form a rule. These rules are used to predict tile production using the
FIS Tsukamoto method. The existence of C 4.5 rules makes it easier to determine the rules
that are built without having to consult with experts because C 4.5 will study existing datasets
to serve as a reference in forming these rules according to conditions that often occur. In
addition, evidenced by the accuracy presented (Mujahid & Sela, 2019). The modeling results
produce relevant rules after being compared with the actual results. In addition, this research
can also assist in estimating predictions of tile production which can estimate related losses
or profits that will occur.

B. METHODS
1. Data Collection

Data collection was obtained from a tile company, to be precise at TH ABADI from January
2018 - September 2022, through direct interviews with the parties concerned. Data can be
seen in Table 1.

<table>
<thead>
<tr>
<th>Month</th>
<th>Year</th>
<th>Land (Kg)</th>
<th>Demand</th>
<th>Supply</th>
<th>Production</th>
</tr>
</thead>
<tbody>
<tr>
<td>January</td>
<td>2018</td>
<td>50000</td>
<td>34000</td>
<td>875</td>
<td>20613</td>
</tr>
<tr>
<td>February</td>
<td>2018</td>
<td>55000</td>
<td>31600</td>
<td>500</td>
<td>23800</td>
</tr>
<tr>
<td>March</td>
<td>2018</td>
<td>53000</td>
<td>29950</td>
<td>400</td>
<td>25016</td>
</tr>
<tr>
<td>April</td>
<td>2018</td>
<td>40000</td>
<td>28720</td>
<td>450</td>
<td>24206</td>
</tr>
<tr>
<td>May</td>
<td>2018</td>
<td>45000</td>
<td>29860</td>
<td>230</td>
<td>25516</td>
</tr>
<tr>
<td>June</td>
<td>2018</td>
<td>42000</td>
<td>29050</td>
<td>300</td>
<td>25107</td>
</tr>
<tr>
<td>July</td>
<td>2018</td>
<td>32000</td>
<td>25050</td>
<td>250</td>
<td>27031</td>
</tr>
<tr>
<td>August</td>
<td>2018</td>
<td>56000</td>
<td>32550</td>
<td>200</td>
<td>30072</td>
</tr>
<tr>
<td>September</td>
<td>2018</td>
<td>45000</td>
<td>30050</td>
<td>250</td>
<td>27558</td>
</tr>
<tr>
<td>October</td>
<td>2018</td>
<td>42000</td>
<td>29550</td>
<td>300</td>
<td>25134</td>
</tr>
<tr>
<td>November</td>
<td>2018</td>
<td>40000</td>
<td>27700</td>
<td>500</td>
<td>23731</td>
</tr>
<tr>
<td>Month</td>
<td>Year</td>
<td>Land (Kg)</td>
<td>Demand</td>
<td>Supply</td>
<td>Production</td>
</tr>
<tr>
<td>----------</td>
<td>------</td>
<td>-----------</td>
<td>--------</td>
<td>--------</td>
<td>------------</td>
</tr>
<tr>
<td>December</td>
<td>2018</td>
<td>30000</td>
<td>24100</td>
<td>100</td>
<td>26010</td>
</tr>
<tr>
<td>January</td>
<td>2019</td>
<td>25000</td>
<td>21050</td>
<td>150</td>
<td>23841</td>
</tr>
<tr>
<td>February</td>
<td>2019</td>
<td>20000</td>
<td>18690</td>
<td>340</td>
<td>24317</td>
</tr>
<tr>
<td>March</td>
<td>2019</td>
<td>25000</td>
<td>21740</td>
<td>100</td>
<td>23045</td>
</tr>
<tr>
<td>April</td>
<td>2019</td>
<td>15400</td>
<td>14230</td>
<td>590</td>
<td>12780</td>
</tr>
<tr>
<td>May</td>
<td>2019</td>
<td>15500</td>
<td>14760</td>
<td>330</td>
<td>23544</td>
</tr>
<tr>
<td>June</td>
<td>2019</td>
<td>25300</td>
<td>23090</td>
<td>240</td>
<td>25060</td>
</tr>
<tr>
<td>July</td>
<td>2019</td>
<td>60000</td>
<td>34590</td>
<td>650</td>
<td>21304</td>
</tr>
<tr>
<td>August</td>
<td>2019</td>
<td>35500</td>
<td>28830</td>
<td>220</td>
<td>27118</td>
</tr>
<tr>
<td>September</td>
<td>2019</td>
<td>45200</td>
<td>30400</td>
<td>320</td>
<td>25603</td>
</tr>
<tr>
<td>October</td>
<td>2019</td>
<td>40500</td>
<td>28240</td>
<td>580</td>
<td>23071</td>
</tr>
<tr>
<td>November</td>
<td>2019</td>
<td>38600</td>
<td>27610</td>
<td>270</td>
<td>25056</td>
</tr>
<tr>
<td>December</td>
<td>2019</td>
<td>55000</td>
<td>31630</td>
<td>640</td>
<td>21641</td>
</tr>
<tr>
<td>January</td>
<td>2020</td>
<td>42000</td>
<td>29440</td>
<td>400</td>
<td>26690</td>
</tr>
<tr>
<td>February</td>
<td>2020</td>
<td>56000</td>
<td>32820</td>
<td>380</td>
<td>25826</td>
</tr>
<tr>
<td>March</td>
<td>2020</td>
<td>45000</td>
<td>30120</td>
<td>260</td>
<td>27565</td>
</tr>
<tr>
<td>April</td>
<td>2020</td>
<td>53000</td>
<td>30760</td>
<td>300</td>
<td>26568</td>
</tr>
<tr>
<td>May</td>
<td>2020</td>
<td>52000</td>
<td>30030</td>
<td>370</td>
<td>25481</td>
</tr>
<tr>
<td>June</td>
<td>2020</td>
<td>57000</td>
<td>33430</td>
<td>440</td>
<td>24816</td>
</tr>
<tr>
<td>July</td>
<td>2022</td>
<td>49000</td>
<td>27790</td>
<td>220</td>
<td>?</td>
</tr>
<tr>
<td>August</td>
<td>2022</td>
<td>55000</td>
<td>32560</td>
<td>260</td>
<td>?</td>
</tr>
<tr>
<td>September</td>
<td>2022</td>
<td>48000</td>
<td>32120</td>
<td>340</td>
<td>?</td>
</tr>
</tbody>
</table>

Table 1 is a table of total production with unknown production from May to September 2022, so this data is used as test data.

2. Create Rules

Before presenting the automatic C 4.5 generation rules using WEKA, here are the general working steps of the C 4.5 algorithm in building decision trees;

a. Select a variable as root.
b. Create a branch for each value.
c. Divide cases into branches.
d. Repeat the process for each branch until all cases in the branch have the same class.

The process of making rules using WEKA by changing the production output value in the dataset is a fuzzy set (Sheena et al., 2017). This study divides 3 fuzzy sets consisting of Little, Enough, and Many (Tundo, 2022). The initial process for changing the value of production output by determining the minimum, middle, and maximum values (Tundo & Nugroho, 2020). Then it is assumed that the minimum value until it approaches the midpoint is Little, approaches the midpoint until it approaches the maximum value is Enough and the rest is Many. The following are the maximum rules obtained after experiencing 5x test trials, where the maximum rules obtained produce an accuracy of 80%, as shown in Figure 1.
Based on the resulting accuracy and detail accuracy for each class based on Figure 1 above, the rules formed by selecting the C 4.5 visual tree from the process produce rules that look like those in Figure 2.
Based on Figure 2, the IF...THEN rule is obtained by writing the classification for each parameter first to make it easier to translate the rules that are formed (Selvachandran et al., 2019). The following classification for each parameter looks like the following.

Supply:
> 600 = Many  
300 – 600 = Enough  
< = 300 = Little

Land:
> 45200 = Many  
42000 – 45200 = Enough  
< = 42000 = Little

Demand:
> 32350 = Many  
< = 32350 = Little

Based on the classification above, the IF...THEN rules that are formed can be seen in Table 2 (Tundo & Kurniawan, 2019).

<table>
<thead>
<tr>
<th>Code</th>
<th>IF ... THEN</th>
</tr>
</thead>
<tbody>
<tr>
<td>R1</td>
<td>IF Supply Enough And Land Enough THEN Production Enough</td>
</tr>
<tr>
<td>R2</td>
<td>IF Supply Little And Land Many THEN Production Many</td>
</tr>
<tr>
<td>R3</td>
<td>IF Supply Enough And Land Many And Demand Little THEN Production Enough</td>
</tr>
<tr>
<td>R4</td>
<td>IF Supply Enough And Land Many And Demand Many THEN Production Little</td>
</tr>
<tr>
<td>R5</td>
<td>IF Supply Enough And Land Little THEN Production Little</td>
</tr>
<tr>
<td>R6</td>
<td>IF Supply Many And Land Enough THEN Production Little</td>
</tr>
</tbody>
</table>

3. FIS Tsukamoto

Tsukamoto’s FIS is a method in which each consequence is in the form of an IF...THEN rule must be represented in the concept of a fuzzy set with a monotonous membership function (Rahmalia, 2021). As for the stages in this method, there are 3 processes to get the output, namely:

a. Formation of Fuzzy Sets
This process requires a rule model that must exist Tundo (2020), such as fuzzy sets, and representation of membership functions and domains (Haghpanah & Taheri, 2017). This study implements the concept based on Table 3, which is in the form of rules, criteria in fuzzy sets, and membership function representations as shown in Figure 3 and Figure 3.
Table 3. Fuzzy Set Rules Model

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Criteria</th>
<th>Fuzzy Set</th>
<th>Domain</th>
</tr>
</thead>
<tbody>
<tr>
<td>Land</td>
<td>Little</td>
<td>Little</td>
<td>[15400 – 27050]</td>
</tr>
<tr>
<td></td>
<td>Enough</td>
<td>Enough</td>
<td>[15400 – 62000]</td>
</tr>
<tr>
<td></td>
<td>Many</td>
<td>Many</td>
<td>[50350 – 62000]</td>
</tr>
<tr>
<td>Demand</td>
<td>Little</td>
<td>Little</td>
<td>[14230 – 19515]</td>
</tr>
<tr>
<td></td>
<td>Enough</td>
<td>Enough</td>
<td>[14230 – 35370]</td>
</tr>
<tr>
<td></td>
<td>Many</td>
<td>Many</td>
<td>[30085 – 35370]</td>
</tr>
<tr>
<td>Supply</td>
<td>Little</td>
<td>Little</td>
<td>[95 – 290]</td>
</tr>
<tr>
<td></td>
<td>Enough</td>
<td>Enough</td>
<td>[95 – 875]</td>
</tr>
<tr>
<td></td>
<td>Many</td>
<td>Many</td>
<td>[680 – 875]</td>
</tr>
<tr>
<td>Output</td>
<td>Production</td>
<td>Little</td>
<td>[12780 – 18210]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Enough</td>
<td>[12780 – 34500]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Many</td>
<td>[29070 – 34500]</td>
</tr>
</tbody>
</table>

Figure 3. Membership Function

Based on Figure 3, the formula for the degree of membership for each set on the criteria looks like this (Geman, Chiuchisan, & Aldea, 2017):

\[
\mu_{[\text{Little}]}(x) = \begin{cases} 
0 & x \geq b \\
\frac{b-x}{b-a} & a \leq x \leq b \\
1 & x \leq a 
\end{cases} \tag{1}
\]

\[
\mu_{[\text{Many}]}(x) = \begin{cases} 
0 & x \leq c \\
\frac{x-c}{d-c} & c \leq x \leq d \\
1 & x \geq d 
\end{cases} \tag{2}
\]

\[
\mu_{[\text{Enough}]}(x) = \begin{cases} 
0 & x \leq a, x \geq d \\
\frac{x-a}{b-a} & a \leq x \leq b \\
\frac{d-x}{d-c} & c \leq x \leq d \\
1 & b \leq x \leq c 
\end{cases} \tag{3}
\]

b. Application Function Implication

Tsukamoto’s FIS method in applying this concept uses the implication function Min (Sumitra & Supatmi, 2019). The form of this function is If.....Then.... And using the AND
operation, where the rules formed in this study use the decision tree C 4.5. The following is an illustration of the implication function application.

$$\mu A \cap B = \min (\mu A[x], \mu B[y]) \quad (4)$$

with:
$$\mu A[x] = \text{the membership degree value of the fuzzy solution up to the rules } i;$$
$$\mu B[y] = \text{value of the degree of membership of consequent fuzzy rules } i;$$

c. Deffuzy

At the defuze stage, changes are made to the results of concluding the Tsukamoto fuzzy inference system to output in the form of crisp values or firm values using the Weighted Average Method process, in general, it is formulated (Jalota & Agrawal, 2019):

$$z^* = \frac{\sum (a \cdot z)}{\sum a} \quad (5)$$

4. Accuracy

Accuracy at this stage is used to measure the success rate of a method that has been used (Hamsa, Indiradevi, & Kizhakkethottam, 2016). In this method, the accuracy calculation uses the Average Forecasting Error Rate (AFER) method with the formula:

$$\text{AFER} = \frac{\sum |A_i - F_i|}{A_i} \times 100 \% \quad (6)$$

Where $A_i$ is the actual data in the data and $F_i$ is the predicted value for the $i$ data. Meanwhile, $n$ is the amount of data (Tuan et al., 2020).

C. RESULT AND DISCUSSION

1. Calculation of FIS Tsukamoto

The following is an example of calculating the prediction of the amount of manual production using FIS Tsukamoto based on test data, which will be used as an example of the manual calculation, namely, in August 2022, with a land inventory of 49,000 Kg, a demand of 27,790, and an existing supply of 220.

Step 1:

Look for the membership degree value of each criterion based on the modeling that has been made in Table 3 and Figure 3. The following is the value of each membership degree of each criterion.

$$\mu_{\text{land–little}[49000]} = 0$$
$$\mu_{\text{land–enough}[49000]} = 1$$
$$\mu_{\text{land–many}[49000]} = 0$$
$$\mu_{\text{demand–little}[27790]} = 0$$
$$\mu_{\text{demand–enough}[27790]} = 1$$
$$\mu_{\text{demand–many}[27790]} = 0$$
\[
\mu_{\text{supply--little}[220]} = \frac{290 - 220}{290 - 95} = 0.3590 \\
\mu_{\text{supply--enough}[220]} = \frac{220 - 95}{195} = 0.6410 \\
\mu_{\text{supply--many}[220]} = 0
\]

**Step 2:**
The implication function application uses the MIN function, for each rule, to find the a and z values for each rule. Where the value of a and z is a parameter to produce a weighted average. Each Rule will have a and z as many rules as are formed like the following calculation;

R1: IF Supply Enough And Land Enough THEN Production Enough
\[
\alpha_1 = \mu_{\text{supply--enough}} \cap \mu_{\text{land--enough}} = \min(0.6410; 1) = 0.6410
\]

Due to R1, R3 for THEN Production Enough so that it uses a rising linear membership degree representation. So there is a Z value that occurs 2 times.

\[
\begin{align*}
\text{R2: IF Supply Little And Land Many THEN Production Many} & \\
\alpha_2 = \mu_{\text{supply--little}} \cap \mu_{\text{land--many}} = \min(0.3590; 0) = 0
\end{align*}
\]

Due to R2 for THEN Production Many, it uses the increasing linear membership degree representation in finding z.

\[
\begin{align*}
\text{R3: IF Supply Enough And Land Many And Demand Little THEN Production Enough} & \\
\alpha_3 = \mu_{\text{supply--enough}} \cap \mu_{\text{land--many}} \cap \mu_{\text{demand--little}} = \min(0.6410; 0; 0) = 0
\end{align*}
\]
R4: IF Supply Enough And Land Many And Demand Many THEN Production Little
\[ \alpha_4 = \mu_{\text{supply-enough}} \cap \mu_{\text{land-many}} \cap \mu_{\text{demand-many}} \]
\[ = \min(0, 6410; 0; 0) = 0 \]

Due to R4, R5, and R6 for THEN Production Little it uses a linear membership degree representation down in finding z.

R5: IF Supply Enough And Land Little THEN Production Little
\[ \alpha_5 = \mu_{\text{supply-enough}} \cap \mu_{\text{land-little}} \]
\[ = \min(0, 6410; 0) = 0 \]

R6: IF Supply Many And Land Enough THEN Production Little
\[ \alpha_6 = \mu_{\text{supply-many}} \cap \mu_{\text{land-enough}} \]
\[ = \min(0; 1) = 0 \]
Step 3:
Results or output can be obtained from the calculation of the weighted average, namely;

\[ Z^* = \frac{a_1 z_1 + a_2 z_2 + a_3 z_3 + a_4 z_4 + a_5 z_5 + a_6 z_6 + a_7 z_7 + a_8 z_8}{0.6410 + 0.6410 + 0 + 0 + 0 + 0} \]

\[ Z^* = \frac{0.6410 \times 16261 + 0.6410 \times 31019 + 0 \times 29070 + 0 \times 12780 + 0 \times 34500 + 0 \times 18210 + 0 \times 18210}{1,282} \]

\[ Z^* = \frac{3,030,650}{1,282} \approx 23,640 \]

After all of the test data from May to September 2022 is calculated, predictions for tile production are produced as shown in Table 4.

<table>
<thead>
<tr>
<th>Month</th>
<th>Land</th>
<th>Demand</th>
<th>Supply</th>
<th>Production Predictions</th>
</tr>
</thead>
<tbody>
<tr>
<td>May</td>
<td>49000</td>
<td>27790</td>
<td>220</td>
<td>23640</td>
</tr>
<tr>
<td>June</td>
<td>55000</td>
<td>32560</td>
<td>260</td>
<td>22461</td>
</tr>
<tr>
<td>July</td>
<td>48000</td>
<td>32120</td>
<td>340</td>
<td>22315</td>
</tr>
<tr>
<td>August</td>
<td>56000</td>
<td>32740</td>
<td>300</td>
<td>21059</td>
</tr>
<tr>
<td>September</td>
<td>53000</td>
<td>32600</td>
<td>350</td>
<td>22785</td>
</tr>
</tbody>
</table>

2. Analysis of Comparative Results
The predicted results are compared directly with actual production, in detail shown in Figure 4. Real production results are shown in a yellow bar chart, while predictions are shown in green. Predictions have less value and are closer to their real value, as shown in Figure 4.

![Figure 4. Comparison of Predictive Results](image)

The prediction comparison results (Table 4) were tested using the Average Forecasting Error Rate (AFER) method. The values obtained have error and truth values as shown in Table 5 and Figure 5.
Based on the calculation, the error value obtained with AFER is 29.34%, so the accuracy of the truth obtained is 70.66%.

D. CONCLUSION AND SUGGESTIONS

The results of the study show that FIS Tsukamoto – Decision tree C 4.5 can predict tile production at TH ABADI Kebumen. The prediction results with actual production from May to September 2022 using AFER have an error percentage of 29.34% with a truth value of 70.66%. The overall results of predictions on actual production did not exceed anything, so it can be concluded that the FIS Tsukamoto - Decision tree C 4.5 method is sufficiently optimum in providing predictive estimates. It is said to be quite optimum because all customer requests are met, either generated by the production prediction itself or the prediction results are added up with inventory data, and all predictions are close to actual production. This research also shortens the time in making rules because it is sufficiently processed with decision tree C 4.5 using WEKA, so there is no need for expert intervention. The rules that are formed from decision tree C 4.5 using WEKA can be accounted for due to the resulting accuracy, where the accuracy of the rule formation reaches 80%. Suggestions for future researchers can make comparisons with other FIS, namely Mamdani and Sugeno, as well as the decision tree that is formed can use other than C 4.5 to find out the differences in each of the methods used.

REFERENCES


