Spare part supplier selection model using decision tree classification techniques: J48 Algorithm

Aulia Ishak* and Tommy Wijaya
Industrial Engineering Department, Faculty of Engineering, Universitas Sumatera Utara, Medan, Indonesia

*E-mail: aulia.ishak@gmail.com

Abstract. Spare parts are goods that consist of several components that form a single unit and have certain functions. To facilitate companies in selecting suppliers, a supplier selection model is needed to make it easier for companies to select suppliers and make it easier to see strategic directions to take several criteria from suppliers to achieve priorities. This research was conducted at a manufacturing company engaged in car spare parts with rubber raw materials. In this study the problems that occur are the difficulty of the company in selecting suppliers based on criteria that are in accordance with the company and the delay in the receipt of raw materials, and the lack of raw materials supplied from each supplier. This study aims to classify suppliers based on desired criteria by the company and to design a supplier selection model in the long term. The J48 algorithm is produced by a supplier selection model with 2 rule selection models to classify efficient suppliers and inefficient suppliers. The accuracy of the Decision Tree model is 90.8547%, the MAE error value is 0.1256. From the J48 algorithm the biggest gain is the criteria for quality, price, delivery, and warranty and complaint services.

1. Introduction

Spare parts are goods that consist of several components that form a single unit and have certain functions. The development of the number of motorized vehicles in Indonesia from 2012-2017 has always increased, with a growth rate of 31.88%. Along with the increasing number of Indonesian motorized vehicles, the demand for motor vehicle spare parts will increase. Supplier have an important role in the availability of raw materials for the ongoing production activities of a company[1]. The selection of the right supplier is not only profitable for the company but also can increase customer satisfaction.[2]

Therefore in the selection of suppliers the company must have a system of selection and evaluation of suppliers of raw materials and components. The selection of suppliers is a strategic activity in purchasing management in the supply chain [3]. Many criteria need to be considered in supplier selection. In the selection of suppliers there are important attributes that provide a large contribution, namely the quality, quantity of raw materials, reliability and price of materials [4].

This research was conducted at a manufacturing company engaged in car spare parts with rubber raw materials. In this study the problems that occur are the difficulty of the company in selecting suppliers based on criteria that are in accordance with the company and the delay in the receipt of raw materials, and the lack of raw materials supplied from each supplier. The shortage of raw material supply from April 2018-March 2019 is 21.21%. The shortage of raw material supply for each supplier is the biggest problem that occurs in the company so that it will have an impact on the production process activities, namely the production capacity so that the company cannot meet consumer demand and will affect the company's sales profit.Developing assessment criteria and methods that provide sustainability measurement are a prerequisite for choosing the best alternative, identifying
requirements, informing the integrated design of alternative show manufacturers [5]. The importance of indicative of conceptual and methodological with the emergence of a variety of criteria and the latest measurement tools in determining best alternative. Innovation and selection of criteria requires parameters related to suitability, reliability, practicality and size limits [6].

Research conducted by Wu regarding supplier selection by combining DEA method, Decision Tree and Neural Network to classify suppliers into efficient and inefficient groups and applying supplier selection prediction models so that companies can implement supplier selection in the long run [7]. To make it easier for companies to select suppliers based on criteria that are in accordance with the company, a framework or model of supplier selection is needed so that suppliers can carry out their obligations in providing raw materials needed by the company for their production processes so that there is no delay in meeting consumer demand and shortages of raw materials by using the Decision Tree namely J48 Algorithm.

2. Method
Data Mining is a process of exploring and analyzing automatically or semi-automatically towards large amounts of data with the aim of finding meaningful patterns or rules. The advantages of data mining techniques include: being able to explore databases to reveal hidden patterns, find information to predict, can handle large amounts of data, high levels of flexibility and can be applied to complex problems [8].

The disadvantage of data mining techniques is that it cannot analyze data directly, and the obstacles to the database used.

Decision Tree is one technique that can be used to classify a set of objects. This research classification used is classification tree, because it uses a classification method using the J48 algorithm. Decision tree consists of internal nodes that describe the data being tested, branches describe the output value of the data being tested, while leaf nodes describe the distribution of classes of data used [9].

This research was conducted from February 2019 to April 2019. The object of the research observed was the selection of suppliers of spare parts raw materials. The variables used in the design of supplier selection models are identified using the delphi method. Then the next variable is identified based on the criteria that most influence supplier selection. The variables in this study are:
1. The dependent variable in this study is the assessment of supplier capabilities and performance, as well as the supplier selection model.
2. The independent variables in this study are supplier selection criteria, namely the criteria of quality, delivery, price, warranty and complaint services.

The data in this study were conducted using a questionnaire consisting of open questionnaires, closed questionnaires, supplier ability assessment questionnaires and supplier performance appraisals, to assess supplier management capabilities in supplying raw materials to companies while assessing supplier ability to assess performance or supplier performance against the company. Steps to collect data about supplier selection models in this case study. The stages of conducting research are:
1. A preliminary study is conducted to identify problems found in the company, production processes, observations and other information.
2. Study the study of literature and supporting theories to get solutions to problem solving.
3. Data collection evaluating the ability of suppliers to assess supplier management capabilities in supplying raw materials to the company while assessing the ability of suppliers to be used to assess supplier performance or performance against the company.
4. Conducting groupings of each supplier criteria attribute.
5. Determination of Split Points by calculating split info (D).
6. Gain calculation for each supplier selection criteria.
7. Processing questionnaire data evaluating supplier capabilities and supplier performance using the J48 Algorithm with WEKA software.
8. Analysis of the results of Decision Tree data processing with the J48 algorithm.
9. Drawing conclusions.

Problem solving analysis begins with obtaining criteria and subcriteria, then classifying supplier selection models using the J48 Algorithm.

### 3. Result and discussions

#### 3.1. Supplier determination criteria results

This questionnaire was given to respondents who represented research among experts in the field of raw material supplier selection. The questionnaire was distributed based on the Delphi principle by gathering opinions from a group of experts through a questionnaire where feedback was carried out through gathering in a complete collection of questions from related literature and supported by experts. Expert data can be seen in Table 1.

| NO | RESPONDENT                      | TOTAL |
|----|--------------------------------|-------|
| 1  | Production manager              | 1     |
| 2  | Head of Production Section      | 1     |
| 3  | Head of Raw Materials Section   | 1     |
| 4  | Staff of Production             | 1     |
| 5  | Staff of Raw Materials Section  | 1     |
|    | **TOTAL**                       | **5** |

From the table above, it can be seen that there are 5 respondents who serve as respondents in the selection criteria and subcriteria for the selection of spare parts suppliers, namely, production manager, head of production section, head of raw materials section, staff of production, and staff of raw materials section.

#### 3.2. Grouping of each supplier criteria attribute

Decision tree is a classification technique in data mining that is used to see the relationship between each subcriteria with the criteria in determining the supplier selection model. A decision tree consists of internal nodes that represent decisions related to split points, and leaf nodes that represent regions or data partitions are referred to as classes or classes. Grouping of supplier criteria attributes can be seen in Table 2.

| Quality | Efficient | Not Efficient | Total |
|---------|-----------|---------------|-------|
| 1       | 0         | 0             | 0     |
| 2       | 9         | 12            | 21    |
| 3       | 4         | 38            | 42    |
| 4       | 17        | 5             | 22    |
| 5       | 69        | 10            | 79    |
| **Total** | **99**      | **65**        | **164** |

Table 2. Number of quality criteria in supplier selection.
In the table above can be seen the classification of quality criteria where the quality with the fulfillment of sub-criteria 5 there are 69 efficient data and 10 inefficient data.

3.3. Determination of split points
The first step in classifying is to choose a split point that provides the best separation or discrimination between supplier selection criteria. This can be done by calculating the entropy (H) and information gain values.

\[
Info(D) = -\sum_{j=1}^{k} \frac{freq(D_j, D)}{|D|} \times \log_2 \left( \frac{freq(D_j, D)}{|D|} \right) = 0.9693
\]  

From the Info calculation (D) explain the overall classification makes information a total of 0.9693 which means it is very good at conducting classifications. Then the info value of each criterion is calculated.

\[
Info_{Kualitas}(D) = \sum_{j=1}^{y} \frac{|D_j|}{|D|} \times Info(D_j)
\]

After calculating the info value of each criterion then calculated the Gain value of each criterion to see how much influence each criterion has on the model to be produced

\[
Gain(Kualitas) = Info(D) - Info_{Kualitas}(D)
\]

| Criteria                  | Gain  |
|---------------------------|-------|
| Quality                   | 0.4155|
| Delivery                  | 0.1622|
| Warranty and Complaint    | 0.0027|
| Price                     | 0.2097|

After obtaining all supplier Gain criteria, the criteria attribute with the largest Gain value 0.4155 is the root of the decision tree, namely quality. Branches followed with the order of Gain values from the largest to the smallest, namely price, delivery, warranty and complaint services.

3.4. Classification using WEKA software
Set data obtained is classified using WEKA software using the J48 algorithm. The classification results using decision tree are as follows: Set data obtained is classified using WEKA software using the J48 algorithm. The classification results using the decision tree are as follows:

J48 unpruned tree

```
Quality <= 3
|   Price <= 5
|   |   Delivery <= 3: not_efficient (49.0/1.0)
|   |   Delivery > 3: efficient (3.0)
|   Price > 5
|   |   Delivery <= 1: not_efficient (2.0)
|   |   Delivery > 1: efficient (9.0)
Quality > 3
|   Price <= 3
|   |   Delivery <= 2: not_efficient (8.0/1.0)
|   |   Delivery > 2: efficient (5.0)
```
Based on the rule assessment above, an example can be taken as follows, suppliers that have quality criteria > 3, price criteria > 3, and delivery criteria > 1 are efficient suppliers with values (83.0 / 5.0) which means that there are 83 data that are appropriate (pure) with the supplier criteria attribute above and there are 5 other data that have missing attributes on the subtree of quality criteria, price criteria, and delivery criteria. The level of accuracy or data classified appropriately amounts to 149 data with an accuracy percentage of 90.8537% of the total data, and there are data that are not classified correctly with 15 data with an error percentage accuracy of 9.1463%. Kappa Statistic obtained from processing results of 0.8084 which indicates that the closeness of the agreement of the results (strength of agreement) is worth fair.

Based on the Decision Tree above, two rules of supplier selection model were obtained to classify suppliers into efficient and inefficient suppliers. The first rules are quality criteria > 3, Price criteria > 3, Delivery Criteria > 1, then suppliers are classified as efficient suppliers. The second rule of quality criteria is <= 3, Price <= 5, Delivery <= 3 so suppliers are classified as not efficient suppliers.

Figure 1. Classification result of decision tree.

4. Conclusions
The criteria for supplier selection which have the greatest weight are quality criteria attributes with the
highest gain value and these criteria are the root of the decision tree, then followed by branches of the criteria of price, delivery, warranty and complaint services. Suppliers that have quality criteria > 3, price criteria > 3, and delivery criteria > 1 are efficient suppliers with values (83.0 / 5.0) which means that there are 83 data that are appropriate (pure) with the supplier criteria attribute. The first rules are quality criteria > 3, Price criteria > 3, Delivery Criteria > 1, then suppliers are classified as efficient suppliers. The second rule of quality criteria is <= 3, Price <= 5, Delivery <= 3 so suppliers are classified as not efficient suppliers. The accuracy of the classification with the J48 algorithm is 90.8537% and the kappa statistical value is 0.8084 which indicates that the agreement on classification results is fair. Error MAE value for Decision Tree classification is 0.1256 and MSE error is 0.2864. The classification results of Decision Tree produce two rules of supplier selection model to classify suppliers so that companies can easily choose suppliers according to the criteria desired by the company.

References

[1] Celebi D and Bayraktar D 2008 Hybrid Analytical Hierarchy Process Model For Supplier Selection Expert Systems with Applications 35 pp 1698–1710.
[2] Chen CT, Lin CT and Huang SF 2006 A fuzzy approach for supplier evaluation and selection in supply chain management Int J Prod Econ 102 pp 289–301.
[3] Chopra S and Meindl P 2001 Supply Chain Management: Strategy, Planning, and Operation (New Jersey Prentice Hall).
[4] Dickson G W 1966 An analysis of vendor selection systems and decisions Journal of purchasing 21 pp 5-17.
[5] Wu D 2009 Supplier Selection : A Hybrid Model Using DEA, Decision Tree and Neural Network. Expert System with Application 36 5.
[6] Ishak A 2017 Decision Support Model for Selection Technologies in Processing of Palm Oil Industrial Liquid Waste IOP 277 pp 2.
[7] Ishak A 2019 Identification criteria and indicators of palm oil industrial liquid waste processing technology IOP 505 pp 2.
[8] Suyanto 2007 Artificial Intelligence (Bandung Penerbit INFORMATIKA)
[9] Kuo RJ, Wang YC and Tien FC 2010 Integration Of Artificial Neural Network And MADA Methods For Green Supplier Selection Journal of Cleaner Production 18 pp 1161-1170.