ISSN : 1978-774X

DETERMINATION OF FAILURE RISK FOR TRANSFORMER SYSTEM BASED ON CLASSIFICATION TECHNIQUE

Iveline Anne Marie[1], Anung B Ariwibowo[2], Docki Saraswati[3], Amal Witonohadi[4]
[1,3,4)] Faculty of Industrial Technology, Industrial Engineering Department, Trisakti University
Jl. Kyai Tapa No. 1, Grogol, Jakarta Barat 11440
e-mail addresses ivelineannemarie@yahoo.com ; docki@trisakti.ac.id ; awitonohadi@yahoo.com

[2)] Faculty of Industrial Technology, Informatics Engineering Department, Trisakti University
Jl. Kyai Tapa No. 1, Grogol, Jakarta Barat 11440
e-mail addres anung@trisakti.ac.id

ABSTRACT

The most essential role of transformers in electrical power distribution network is to maintain voltage level requirements. In this study, the risks of transformer failure is studied using the Decision Theory classification techniques and Naive Bayes Algorithm. The risk of failure of the transformer system become the basis for the development of a Risk Matrix Model in the Decision Support System (DSS) for transformer maintenance policy at the company. Interviews were conducted to obtain data on the risks of failure of the transformers. Decision Tree and Naive Bayes are built based on empirical data transformer maintenance involving six data attributes. The Decision Tree and Naive Bayes Algorithm are used to determine the rules that can be used to determine the Risk Class Attribute . Next , intuitively Bayesian formula calculates the probability value of data belonging to the class of a particular risk attribute. The study used Learning Algorithm J48 and Naive Bayes Classifier Algorithm as a comparison. Both algorithms are implemented in software package WEKA.

Key words: transformer, risk matrix, classification technique, decision tree, naive bayes

1. INTRODUCTION

The decision making process in the industry that provides many services to the community, concerns the interests of many parties and requires substantial investment tend to be highly risky. The decision making process should be done with careful consideration and must consider many aspects, given the magnitude of the risk to be borne if wrong in taking the decision (Marie, 1999).

The right decision in management and utilization of resources appropriately is a strategic approach to improving the performance of the organization. The right decisions are influenced by the completeness and accuracy of information and knowledge that is involved in the decision-making process itself (Kadarsah, 1998).

Division X is the parent unit Owned Company (Persero) PT Y, with the main task is to operate and maintain an electric power transmission system JB area and to plan the development of the power system JB. In the execution of its duties to provide electrical power for the needs of society, division X PT Y depends on the availability and reliability of facilities and infrastructure of electricity transmission system owned. (Marie, 2014).

Maintenance is all about keeping, caring for, maintaining the functions of the equipment to ensure the equipment is able to do what the user wants and when users want the equipment did so. In the equipment maintenance system, the approach Reliability Centered Maintenance (RCM) is used to determine how to maintain an asset correctly or the system can do smoothly in the current operation as desired by the user (Moubray, 1997).

The transformer is the most important equipment for the agencies that manage the transmission network. Decision making for transformer maintenance system at Division

ISSN : 1978-774X

X PT Y requires many attention of decision makers. Security and reliability system is affected by the status of the data transformer during operation. There are several types of system failures on the transformer. The impact on any kind of failure is different from each other. There is a type of failure that often occurs but the impact is almost non-existent, on the contrary there are kinds of failures that are rare but the impact is very serious. This paper discusses the use of risk analysis with risk marix approach to know the factors that influence the likelihood of system failure and the consequences faced.

Classification Techniques can find a model (functions) that describe and distinguish classes or concepts. This model can be used to predict the class or object that has a class label that is unknown. The model is derived based on the analysis of data training (the data object has a class label and can be represented in various forms such as if-then rules, decision trees, mathematical formulas or neural network) (Tan, 2006).

2. THEORETICAL BACKGROUND

Risk analysis procedures covers all activities aimed at identifying threats, estimating risks and impacts. The occurrence of an unusual event can cause a state of emergency and may even result in harmful effects. An assessment of the risk posed by an event can be determined by the applicable criteria (Rak , 2006).

Decision Tree algorithm works based on the entropy of the information in the observation data. Information entropy is defined (Shannon, 1948) can describe the content of the information in a communication channel. Training data used in the Decision Tree algorithm can determine which attributes have the best "driving force" (decisive power). Determination of the decisive power is based on how pure or homogeneous data on each attribute of the data in the Attribute Class. The more homogeneous data on an attribute of the data in the Attribute Class, the greater the ability of these attributes to be taken as an attribute of decision makers. Entropy formula used in the Decision Tree algorithm to determine how homogeneous data on an attribute is the following

$$E(S) = \sum_{i=1}^{c} - p_i \log p_i \qquad (1)$$

where S is an attribute, c is the number of attributes available in the data set, and p_i is the proportion of the value of i in the training data. Based on this formula, the data will have a homogeneous entropy value 0, and the data is non-homogeneous distributed equally have entropy values 1.

Naïve Bayes algorithm is based on the probability of frequency of occurrence of data values for each attribute and the Class Attributes (Rish, n.d.). From a given set of observation data, grading process is computed using a formula based on the observed data:

$$P(C \mid v_i) = \frac{P(C) \cdot P(v_i \mid C)}{P(v_i)} \qquad (2)$$

where P(C) is the prior probability of class C, $P(v_i|C)$ is the likelihood probability of the attribute value, and $P(v_i)$ is observed probability.

As a classifier, Naive Bayes has an advantage in determining the class attributes without having received a complete observational data. The Decision Tree algorithm may result in "the wrong decision" if the data used in the training phase (training) is not complete. Knowledge about the feasibility of the data may be used as a training data. This is one step in the pre-processing of data. In this study, the collected data is assumed to already sufficiently representative to serve as training data.

Decision Tree is the method in Data Mining which is classified into Classifier Models Group. This model can be used to determine the class or group of an object of observation. Classifier model becomes a guided (supervised) learner algorithm because the data of the observation object

ISSN : 1978-774X

has been given as a guide for Decision Tree algorithm. The advantage of the Decision Tree algorithm is the ability of this algorithm to classify the data that is not found in the training data based on the inference that formed from the training data.

3. RESEARCH METHOD

Based on structured interviews with the Manager of Asset Management and Maintenance System as well as the related staff, expert studies were conducted to determine factors that cause failure risk on a transformer system. By using Classification Techniques in Decision Tree algorithm and Naive Bayes algorithm we process the determinant factor of failure risks in order to determine the level of risk of failure of the transformer system. The results of data processing will be used as the basis for development of decision models in the Risk Matrix Maintenance System Decision Support Systems at the Power Transmission Division X PT Y.

4. RESULT AND DISCUSSION

4.1. Maintenance System at the Division X PT Y

Maintenance System at Division X PT Y is currently done based on Condition-Based Maintenance (CBM) Approach which is a maintenance strategy that utilizes equipment conditions as a basis for determining the equipment maintenance activities. Checking the status of the condition of the equipment can be done by using the results of the monitoring (Off / On Line) and the previous maintenance, refer to the equipment failure modes and the equipment aging process.

In this study, the observed equipment restricted to the Power transformer equipment is often called the Daya Transformer. This transformers have a very important role in the distribution network to deliver electricity from the substation to the load. Power distribution system is a system of channeling electricity from power plants on the required voltage level.

4.2. Development of the Decision Model: Determination of the Risk Level

Decision model in this study was developed based on the condition assessment. With this model we can know: the relative importance or urgency of assets, the speed rate of causing problem, and the critical level of the asset problems so that it can affect the outage and failure. At the risk matrix, the results of the risk assessment has 4 levels category as follows: Low, Moderate, High and Extreme considering Likelihood and Impact Criteria which is determined based on five following criteria: Safety, Extra Fuel Cost, System Reliability, Equipment Cost, Customer Satisfaction & Social. Table 1 shows the six attributes of the data from Level 1 to Level 5.

Table 1. Attribute Data for Each Level

Level.	Like. of Occur.	Safety	Extra Fuel Cost (Rp)	Syst, Rel.	Equip. Cost (Rp)	Cust. Satisfact.
1	1 in more than 100 years	Near miss	< 100 Mil	<100 MWh	<50 Mil	Compl. from VIP cust.
2	1 in 10 years to 1 in 100 years	First aid injury, medical aid injury	0,1 – 1 Bil	0,1 – 1 GWh	50 - 500 Mil	Compl. from industrial cust.
3	1 in 1 year to 1 in 10 years	Lost time injury temp. disab.	1 – 10 Bil	1 -10 GWh	0,5 – 5 Bil	Compl. from comm.
4	1 - 10 in 1 year	Perman. disab.	10 - 100 Bil	10 – 100 GWh	5 - 50 Bil	Compl. from comm.that have potential riot
5	> 10 in 1 year	Fatality	>100 Bil	> 100 GWh	>50 Bil	High potential riot

One of the outputs of Matrix Risk Decision Model is the Level of Risk (Low, Moderate, High and Extreme) in accordance with the data in Table 2.

Table 2. Risk Level

Level	Risk
1	Low
2	Moderate
3	High
4	Extreme

Next we develop a Decision Tree based on Classification Techniques Approach. This study use a learning algorithm J48 (Quinlan , 2007) which is one of several algorithm variants of Decision Tree that has been developed and widely used. In addition to the J48 algorithm, we also use Naive Bayes classifier algorithm as a comparison. Both of these algorithms are implemented in software package Weka (Waikato Environment for Knowledge Analysis) (Hall et al., N.d.).

The figure 1 below show the Decision Tree obtained from J48 algorithm applied on the data set.

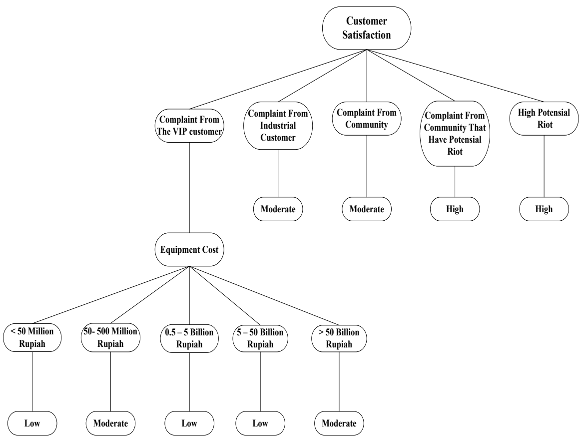


Figure 1 Decision Tree for Transformer Failure Risk

Decision Tree and Naive Bayes is built based on transformer maintenance empirical data. The Decision Tree output can be used to help determining the decisions that need to be taken in regard to each transformer. The attributes of the data used can be seen from Table 1, includes the Likelihood of Occurrence, Safety, Extra Fuel Cost, System Reliability, Equipment Cost, and Customer Satisfaction. Based on the six attributes, Decision Tree algorithm and Naive Bayes is used to determine the rules that can be used to determine the Risk Attribute Class. Next, intuitively Bayesian formula calculates a probability value of data belonging to the certain attribute class, based on the observations.

Each attribute is an nominal attribute. Total possible combinations that can be formed from the entire attribute to determine the decision risk is 5 × 5 × 5 × 5 × 5 × 5 × 4 = 62500 data. Based on interviews with the parties involved, we collected only 125 data.

Table 3 shows a sample of data used in this study.

Table 3. Sample Data for Decision Tree

No.	*Like. of occur.*	*Safety*	*Extra fuel cost*	*Sys. Rel.*	*Equip. cost*	*Cust. Satisfac.*	*Risk*
1	4	5	1	1	1	1	2
2	4	5	1	1	1	2	2
3	4	5	1	1	1	3	3
4	4	5	1	1	1	4	3
5	4	5	1	1	1	5	4
...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...
123	1	2	3	1	1	1	1
124	1	2	3	1	1	2	1
125	1	2	3	1	1	3	2

The application of J48 algorithm to the data set results in a classifier in the form of decision tree. The classifier obtained is as follows:

CSSoc = complaint from the customer vip
| EC = <50 million rupiah: low (18.0 / 2.0)
| EC = 50-500 million rupiah: moderate (2.0)
| EC = 0.5 - 5 billion rupiah: low (2.0 / 1.0)
| EC = 5 -50 billion rupiah: low (2.0 / 1.0)
| EC => 50 billion rupiah: moderate (1.0)

CSSoc = complaint from industrial customer: moderate (25.0 / 6.0)

CSSoc = complaint from community: moderate (25.0 / 9.0)

CSSoc = complaint from the community that have a potential riot: high (25.0 / 7.0)

CSSoc = high potential riot: high (25.0 / 10.0)

4.3. Illustration Implementation

For example, calculations in Table 4 show a subset of transformer maintenance data used in the study. Examples of Bayesian calculations using two of the eight existing attributes (Safety and Risk), to describe the process of Bayesian calculation. Bayesian formula can be used to help calculate the probability of risk if we observe Safety factor = Fatality.

Table 4. Subset Data Transformer Maintenance

No.	Safety	Risk
1	Fatality	Extreme
2	Fatality	Extreme
3	Fatality	Extreme
4	Fatality	Low
5	Near miss	Extreme
6	Near miss	Extreme
7	Near miss	Low
8	Near miss	Low
9	Near miss	Low
10	Near miss	Low
11	Near miss	Low
12	Near miss	Low
13	Near miss	Low
14	Near miss	Low
15	Near miss	Low
17	Near miss	Low
18	Near miss	Low

Based on data from Table 4, known P (Risk = Extreme) = 5/17 = .294 and P (Safety = Fatality) = 4/17 = 0.235. Factors likelihood P (Safety = Fatality | Risk = Extreme) is calculated based on reduced set like data in the Table 5.

Based on the set reduced, likelihood value is calculated from the number of occurrences Safety = Fatality among all occurrences Risk = Extreme, equals 3/5 = 0,600

$$P(\text{Risk} = \text{Extreme} \mid \text{Safety} = \text{Fatality}) = \frac{P(\text{Risk} = \text{Extreme}) \cdot P(\text{Safety} = \text{Fatality} \mid \text{Risk} = \text{Extreme})}{P(\text{Safety} = \text{Fatality})}$$

$$P(\text{Risk} = \text{Extreme} \mid \text{Safety} = \text{Fatality}) = \frac{0{,}294 \cdot 0{,}600}{0{,}235} = 0{,}750$$

From this example we can observe that the Extreme risk factor increased from 0.294 to 0.750 if we observe that there is a fatality safety factor.

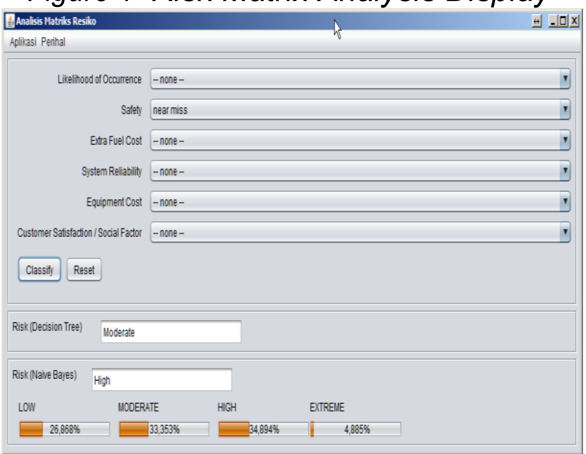
Table 5. Reduced Set for The Observation of extreme = risk

No.	Safety	Risk
1	Fatality	Extreme
2	Fatality	Extreme
3	Fatality	Extreme
4	Near miss	Extreme
5	Near miss	Extreme

Based on the results of the implementation of Decision Support System (DSS) prototype has been built, Naive Bayes classifier can be used to help determine risk based on situations observed.

A prototype application is shown in Figure 1. The picture show a case where the observation data collected only a safety factor that is worth "near miss". Decision Tree algorithm provides the results in risk level "moderate", while the Naïve Bayes algorithm gives the results in the risk of level "high". There is a small difference value between the Bayesian probability of the risk of "moderate" (0.333) and the risk of "high " (0.349). From the probabilistic point of view, it may be in fact that the level of risks arising is classified as "moderate". The result of this computation is given to system users to make final decision.

Figure 1 Risk Matrix Analysis Display



5. CONCLUSION

The Risk Matrix Decision Model which had been developed into a system of applications utilizing Decision Tree algorithm and Naive Bayes algorithm can help decision maker to acquire the risk level of any failure of a transformer system in order to determine a further solution. Development of the decision model and the system support the policy for transformer maintenance system at the Division X PT Y.

6. REFERENCES

(a) Friedman-Hill, Ernest. 2003. JESS in Action: Rule-based Systems in Java. USA : Manning Publications.
(b) Hall, M., National, H., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I. H. (n.d.). The WEKA Data Mining Software : An Update, *11*(1), 10–18.
(c) Kadarsah, Suryadi, Dr.Ir. dan Ir. M. Ali Ramdhani, MT. (1998) Sistem Pendukung Keputusan : Suatu Wacana Struktural Idealisasi dan Implementasi Konsep Pengambilan Keputusan. Bandung : PT Remaja Rosdakarya .
(d) Marie, I. A. (1999) Perancangan Sistem Pendukung Keputusan Untuk Penggantian Peralatan (Studi Kasus Penggantian Sistem Peralatan Tol di PT X) [Tesis]. Jakarta: Program Pascasarjana, Universitas Indonesia.
(e) Marie, I.A., Saraswati, D. dan Witonohadi, A. (2014) Perancangan Model Keputusan Pemeliharaan Sistem Transmisi Tenaga Listrik di Divisi X PT Y. Prosiding Seminar Nasional Teknik Industri BKSTI 2014, Bukittinggi, 3 September 2014, ISBN 978-602-9081-11-4, VI-68 –VI-72.
(f) Rak, J., dan Tchorzewska-Cieslak, B., (2006) Five-Parametric Matrix to Estimate the Risk Connected with Water Supply System Operation, *Environment Protection Engineering*, Vol.32, h. 37-46.
(g) Moubray, John. *(1997)* Reliability Centered Maintenance*. London :* Butterworth-Heinemann*, British Library.*
(h) Quinlan, J. R. (2007) Induction of Decision Trees, 81–106.
(i) Rish, I. (n.d.). An empirical study of the naive bayes classifier. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/summary;jsessionid=042074214DC4FB6BE0DC8C0223FA1519?doi=10.1.1.330.2788

AUTHOR BIOGRAPHIES

Iveline Anne Marie is a lecturer in Department of Industrial Engineering, Faculty of Industrial Technology, Trisakti University, Jakarta. She received her Master of Industrial Engineering from Indonesia University in 1999. Next, in 2012, she received her Doctoral Degree from Bogor Agriculture Institute. She is a member of the Production System Laboratory. Her email address is <ivelineannemarie@yahoo.com

This paper was supported by the Directorate General of Higher Education, Ministry of Education and Culture of the Republic of Indonesia Number 180 / K3 / KM / 2014. The authors are grateful for all the help and support of data from the management and staff of the Division X PT Y and Trisakti University Research Institute for the implementation of this research.