

Case-Based Reasoning using The Nearest Neighbor Method for Detection of Equipment Damage to PLN Power Plant

Riska Amalia Praptiwi*¹, Nur Rokhman², Wahyono³

¹Master Program of Computer Science, FMIPA UGM, Yogyakarta, Indonesia

^{2,3}Department of Computer Science and Electronics, FMIPA UGM, Yogyakarta, Indonesia
e-mail: *¹riskamaliatiwi93@mail.ugm.ac.id, ²nurrokhman@ugm.ac.id, ³wahyo@ugm.ac.id

Abstrak

Predictive Maintenance (PdM) pada Pembangkit PLN merupakan suatu kegiatan pemantauan peralatan secara periodik sebelum peralatan tersebut mengalami kerusakan pada kondisi yang lebih parah. Menurut seorang pakar atau owner PdM bahwa analisis pemeliharaan kurang tepat dan efisien berdampak pada biaya pemeliharaan yang tidak sedikit. Pada kondisi yang nyata, owner PdM menganalisis kerusakan peralatan berpedoman pada kasus - kasus kerusakan peralatan yang terjadi sebelumnya. Maka diperlukan suatu sistem cerdas berbasis komputer yang dapat membantu deteksi kerusakan peralatan.

Berdasarkan Kajian Pustaka yang telah dilakukan, Penalaran Berbasis Kasus dapat menyelesaikan masalah baru dengan menggunakan jawaban atau pengalaman dari masalah lama seperti meniru kemampuan manusia. Proses Penalaran Berbasis Kasus terdapat tahapan paling penting yaitu mencari nilai similaritas atau tingkat kesamaan paling tinggi antara kasus baru dengan kasus lama dengan mengadaptasi solusi dari kasus-kasus lama yang sudah terjadi (Sankar, 2004). Pada penelitian ini proses similaritas atau pendekatan menggunakan Nearest Neighbor.

Pengujian pada sistem menggunakan data uji sebanyak 20 data dan pengukuran kinerja sistem menggunakan confusion matrix. Evaluasi pengujian menggunakan confusion matrix dapat diketahui seberapa akurat sistem dapat mengklasifikasikan data secara benar yaitu sebesar 97,98%. Kemudian nilai presisi 95% merupakan jumlah data kategori positif yang diklasifikasikan secara benar dibagi dengan total data yang diklasifikasi positif. Selanjutnya hasil pengujian terhadap data uji deteksi kerusakan peralatan pada pembangkit PLN dengan nilai threshold 0,75 menggunakan nearest neighbor, sistem memiliki unjuk kerja dengan tingkat sensitivitas 95%.

Kata kunci— *Predictive Maintenance* Pembangkit PLN, Penalaran Berbasis Kasus, Nearest Neighbor

Abstract

Predictive Maintenance (PdM) at the PLN Power Plant is a periodic monitoring of equipment activities before the equipment is damaged in more severe conditions. According to an expert or PdM owner that maintenance analysis is not appropriate and efficiency has an impact on maintenance costs that are not small. In real conditions, the PdM owner analyzes equipment damage based on previous cases of damage equipment. Then we need a computer-based intelligent system that can help detect damage to equipment.

Based on the Literature Review that has been done, Case-Based Reasoning can solve new problems using answers or experiences from old problems such as imitating human

abilities. *Case-Based Reasoning Process* there is the most important step, which is to find the highest similarity value or the level of similarity between new cases and old cases by adapting solutions from old cases that have occurred (Sankar, 2004). In this study the process of similarity or approach using Nearest Neighbor.

Testing on the system uses 20 test data and the measurement of system performance uses confusion matrix. Evaluation of testing using confusion matrix can be seen how accurately the system can classify data correctly that is equal to 97.98%. Then the precision value of 95% represents the number of positive categorized data that is correctly divided by the total data classified as positive. Furthermore, the test results of the equipment damage detection test data at the PLN plant with a threshold value of 0.75 using the nearest neighbor, the system has a performance with a 95% sensitivity level.

Keywords— *Predictive Maintenance at PLN Power Plant, Case-Based Reasoning, Nearest Neighbor*

1. INTRODUCTION

Predictive Maintenance (PdM) at the National Electric Power Plant (PLN) is a periodic monitoring of equipment to detect the beginning of damage to an equipment before the equipment is damaged in more severe conditions. Equipment that is damaged even to the point of failure (failure) can disrupt the production process and increase maintenance costs. The PdM activity also serves to maintain the condition of the condition based maintenance equipment [1].

In the process of PdM equipment will provide a signal, then the results will appear in the form of reports for each of these equipment, then experts analyze one by one equipment and in large quantities through existing technologies. Experts have integrated vibration technology and thermography technology for equipment damage analysis at the PLN power plant. There are types of reports that are generated from signals using vibration technology, among others, in alarm condition A shows that the equipment is still in a condition with no symptoms of damage such as the state of the machines are still new, alarm B shows the condition of the equipment is normal and sometimes the symptoms of slight damage, alarm condition C show symptoms of damage that occurs that cause the equipment can only operate in the short term and alarm D shows symptoms of severe damage. As for the thermography technology report, it is a temperature with a unit of °C on each piece of equipment. [1].

In real conditions, the predictive maintenance team analyzes equipment failures based on previous cases of equipment failures. Because there is equipment that has almost the same damage as the previous damage case with the same damage analysis, a different case with the same damage analysis, the same case with a different handling solution, and so forth. Equipment in large quantities (1 equipment = more than 1 machine) can damage at the same time and must be analyzed immediately so that no damage can occur that is getting worse. In this case the damage analysis can be less than optimal and efficient so that there can be an analysis that is less precise and accurate. If there is a lot of severe damage, then the maintenance costs incurred are very expensive or not small. So that PdM experts have not been able to handle damage optimally.

These conditions, a computer-based intelligent system that is able to analyze the handling of damage to generating equipment will be very helpful especially for an expert. Complex problems require a fast, precise and accurate method. So we need Case-Based Reasoning (CBR) methods. CBR is one approach that can provide damage detection and treatment of new damage cases based on previous damage cases that have similar features or criteria. This CBR system is used to complement or strengthen a knowledge-based expert system that has developed previously that is rule-based, this system produces conclusions based on similarities to the cases that occur [2].

The CBR process has the most important step, which is to find the highest similarity value between new cases and old cases by adapting solutions from old cases that have already occurred [3]. In this study the process of similarity or approach using Nearest Neighbor (NN). This CBR system is used to complement or strengthen a knowledge-based expert system that has developed previously that is rule-based, this system produces conclusions based on similarities to the cases that occur [4]. NN was chosen because it is one of the classification and pattern recognition techniques and is widely used in the CBR system[5]. Damage detection is expected to be an experienced assistant for experts to help conduct early handling of equipment damage to the PLN Power Plant.

In the search for similarity values, most use symbolic or continue data, but in this study have used numerical (discrete) data and symbolic data (continue). Data contained in numerical and symbolic data are called mixed attributes[6].

The CBR process has the most important step, which is to find the highest similarity value between new cases and old cases by adapting solutions from old cases that have already occurred[7]. In this study the process of similarity or approach using Nearest Neighbor (NN). This CBR system is used to complement or strengthen a knowledge-based expert system that has developed previously that is rule-based, this system produces conclusions based on similarities to the cases that occur[4]. NN was chosen because it is one of the classification and pattern recognition techniques and is widely used in the CBR system[5]. Damage detection is expected to be an experienced assistant for experts to help conduct early handling of equipment damage to the PLN Power Plant.

2. METHODS

2.1 Description System

The system built is a damage detection system at the power plant at PLN. The output of the system can be used for consideration by an expert or owner of PLN's Predictive Maintenance as a detection and handling solution solution for damage to generating equipment. The model used in this study is a case-based reasoning model (CBR). The basic idea in selecting this model is a report on the case history of handling damage to the power plant in the PLN (Jambi Sector). The CBR model used in this study is the detection of equipment damage to the generator by applying the Nearest Neighbor algorithm.

The first stage of the system in this study begins by entering data on the age and temperature features (thermographic measurement results) and the damage experienced by the device. This stage aims to get a history of tool damage as a system input. Then the data to be entered on a case basis is obtained from the history of tool damage analysis at the Jambi PLN power plant. The data will then be stored into a case base, in one data there will be information for the type of equipment damage and tool handling solutions for the generator.

The next stage is entered in the retrieval process. The retrieve process in this study uses the Nearest Neighbor. Retrieve is the process of determining the level of closeness or similarity of the target case to the case base by using Nearest Neighbor. So that it can be seen the similarity values of the target cases against each case base. Terdapat dua proses perhitungan pada proses pencarian similaritas kasus. Setelah memasukkan data uji dilakukan perhitungan similaritas lokal terlebih dahulu. Apabila similaritas berupa numerik maka akan memakai rumus similaritas numerik. Kemudian untuk similaritas diskrit maka akan dihitung menggunakan similaritas diskrit yang jika sama maka bernilai 1 dan jika tidak sama maka akan bernilai 0. Tahap kedua yaitu pencarian similaritas Global.

Limits on the similarity value determined by the threshold value. The threshold value used is 0.75. The similarity value is more than equal (the threshold value will be continued in the verification process. If there is more than 1 data that has a similarity value above the threshold, then the highest similarity value and pass the verification process will be used as a solution. Then the other similarity values or those that are does not pass verification, is entered into the reuse process, if the similarity value is less than (<), then the revised data entered by the expert is then stored in the database, the process can also be called retain.

2. 2 Representation Case

A case needs to be represented in the form of a particular interest for storage on a case basis and retrieval process. According to [8], the representation of a case must include the problem at hand and the solution to the problem.

Cases can be represented in various forms, such as prepositional representations, frames, semantic networks, and combinations of the three [9]. The choice of case representation models depends on the domain and case data structure available. Table 1 shows a model representation of the case in the form of frames.

Tble 1 Model representation of the case in the form of frames.

| Case 1 | |
|--------------------------|---|
| Feature | Value |
| Usia Peralatan | 9 tahun |
| Report Termografi (Suhu) | 115°C |
| Report Vibrasi | - Terdapat kenaikan nilai vibrasi - Indikasi spectrum subharmonik - Terdapat unbalance pada shaft |
| Jenis Kerusakan | Rolling Element Bearing Wear |
| Solusi penanganan | Lakukan pembersihan pada fan |

2. 3 Retrieval

CBR is based on the hypothesis that solutions to cases that have never happened before can help solve new cases that occur on condition that there are similarities between these cases[9]. Measurements are made by comparing the similarity of features in a target case (case target) with similar features that exist on a case basis. A case is identical to another case if it has a similarity value of one, whereas if the similarity value is less than one, it can be said that the case is similar.

The more case data stored in the base case, the longer it takes for scanning / matching features between the test data and the base case. To overcome this, indexing is performed on each data record contained in the database. Nearest Neighbor Retrieval is a retrieval technique that is often used in CBR systems. How the NN algorithm works by comparing each feature in the new case against the features found in the old case in the case base, then the comparison is calculated using the similarity function. The similarity value is in the range of 0 to 1. The value of 1 indicates that the cases found have 100% similarity, while the value of 0 indicates no similarity.

2. 3.1 Nearest Neighbor

Nearest Neighbor is one Mechine Learning algorithms are relatively simple to determine the classification of an object. This algorithm included in the category of supervised learning which requires a number of reference data that has a class attribute. The workings of this algorithm are to compare each attribute of the target case (target case) with the attributes

contained in the base case, then the comparison is calculated using the proximity value function (similarity). There are two types of similarity used in the Nearest Neighbor algorithm, that is:

1) Local Similarity

Local similarity is defined as the level of similarity contained in data / features. Local similarity can be divided into two, namely numerical (discrete) and symbolic (continuous). Discrete data is data whose value is a natural number, while continuous data is data whose value is in a certain interval. In this case features such as the age of the equipment at the plant and the temperature of the thermographic report are included in the numeric type. For numerical data types with a range of data, normalization can be carried out with the aim of scaling the process value to fall within the specified range shown in Equation (1):

$$f(s_i, t_i) = 1 - \frac{|s_i - t_i|}{\max(s_i, t_i) - \min(s_i, t_i)} \quad (1)$$

Where $f(s_i, t_i)$ is the local similarity of the $-i$ between source case and feature target case. s_i is the value feature $-i$ from source case. t_i is the value feature i from target case. $\max(s_i, t_i) =$ maximum value feature i between source case and feature target case. $\min(s_i, t_i) =$ minimum value feature i between source case and feature target case.

The value on the local similarity using Boolean symbolic of data. Then for Boolean symbolic data types shown in Equation (2)[9]:

$$(s_i, t_i) = \begin{cases} 1 & \text{jika } s_i = t_i \\ 0 & \text{lainnya} \end{cases} \text{ where } s_i, t_i \{true, false\} \quad (2)$$

where s_i is source case and t_i is target case. For example, the symbolic value in the source case is the same as the target case, so it's worth 1. Conversely, if the symbolic value in the source case is the same as the target case, then it's worth 0.

2) Global Similarity

Global similarity is defined as the level of similarity contained in an object or contained in a case. Global similarity is used to calculate the similarity between new cases and cases that are stored in the base case. There are many ways to measure the distance between cases, such as Euclidean Distance, Manhattan Distance, and Nearest Neighbor. Nearest Neighbor has several advantages, namely the resilience of training data that has a lot of noise and is very effective if it has a large amount of training data. Calculating the value of similarity (similarity cosine) between test data and all training data can be used Equation (3) [10].

$$SimNN(s, t) = \left(\frac{\sum_{i=1}^n f(s_i, t_i) \times w_i}{\sum_{i=1}^n w_i} \right) \quad (3)$$

where $SimNN(s, t)$ is global similarity between source case and target case t . $f(s_i, t_i)$ is local similarity feature $-i$ between source case and feature target case. n is the amount of feature. w_i is weight value feature i . In 2012 Uki Mancasari has modified the formula used to overcome the value of trust and treatment in the target case. In equation (4) is a modification of nearest neighbor.

$$SimNN(s, t) = \left(\frac{\sum_{i=1}^n f(s_i, t_i) \times w_i}{\sum_{i=1}^n w_i} \right) \times P(S) \times \frac{J(S_i T_i)}{J(T_i)} \quad (4)$$

where $P(S)$ is the level of confidence case in source case, $J(S_i, T_i)$ is the number of symptoms found in the target case that appear in the source case symptoms, $J(T_i)$ is the number of symptoms found in the target case.

Given the weight of the features used in this case have the same weight, which means all of the features considered essential by experts as the global similarity formula does not need to use weights. So that used the formula global similarity equation (5) [11]

$$SimNN(s, t) = \left(\frac{\sum_{i=1}^n f(s_i, t_i)}{\sum_{i=1}^n i} \right) \quad (5)$$

2.4 Flow Model System

The model is built based on the system flow analysis system needs. Models built workflow system consists of several parts with each function as illustrated in Figure 1.

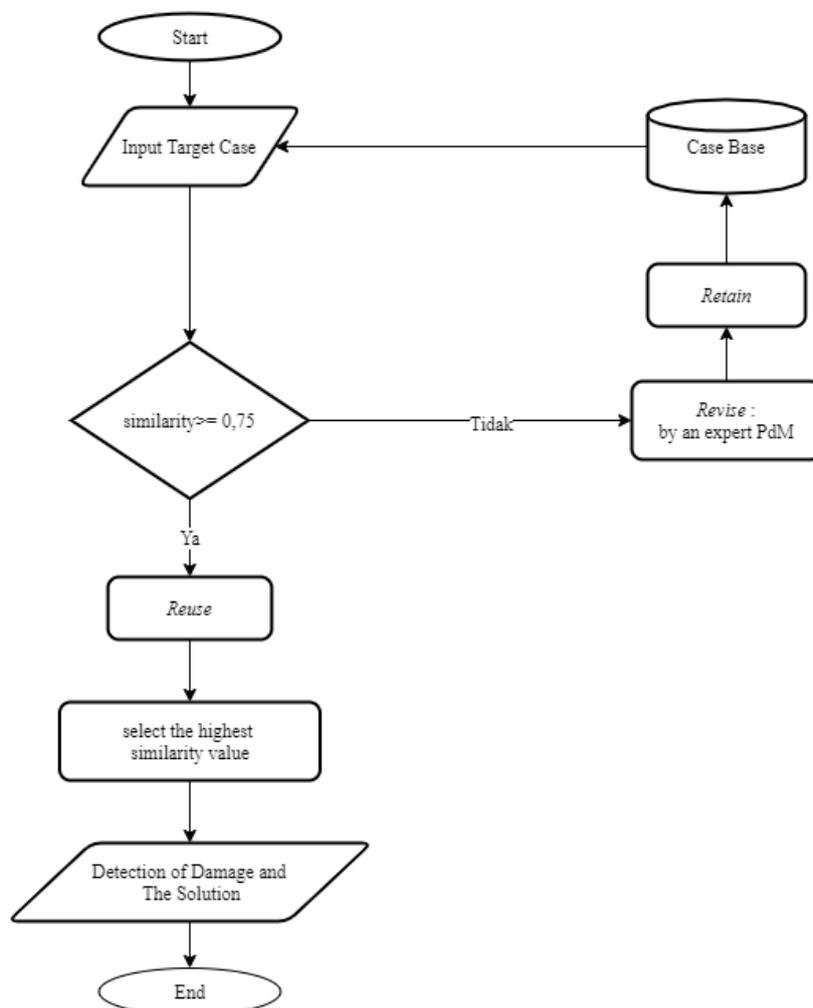


Figure 1 Architectural Design of CBR for Damage Detection Equipment on Generating PLN

2.5 Database Design

Database designed in the form of the model database diagram as shown in Figure 2

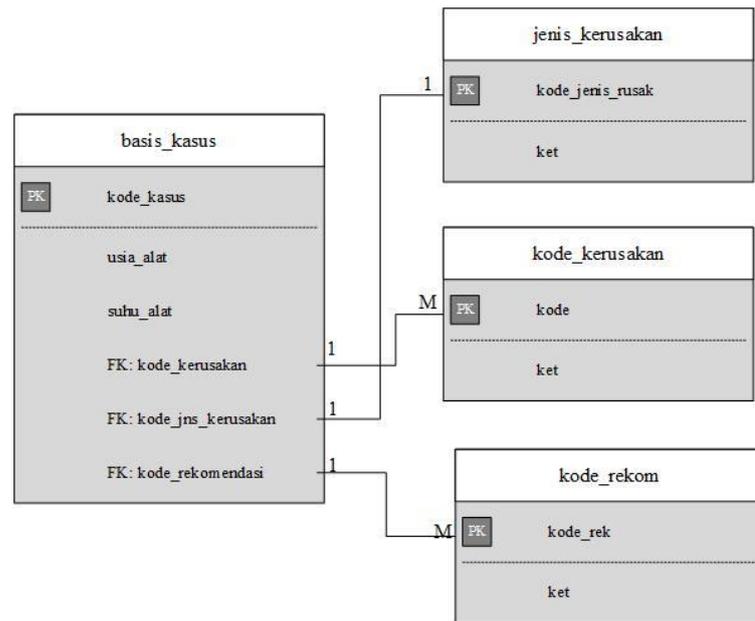


Figure 2 Design of the model database diagram

2. 6 Data and Methods of Testing Systems

At this stage, the testing will be conducted using data report cases of damage to Predictive Maintenance. The test scheme is to provide 100 data which are then chosen 80% (80 data) as training data and 20% (20 data) as test data.

System testing in this research uses diagnostic tests. Diagnostic test is a technique to measure the ability of a system to detect damage. In the diagnostic test there are terms such as sensitivity. Sensitivity values can be used to determine the accuracy of annotated test (Tempola, 2018). Then the accuracy is the level of measurement of the proximity of the quantity to the actual value. Analysis of the test results was carried out using 4 parameters namely True Positive (TP), False Positive (FP), True Negative (TN), False Negative (FN). According to Han and Kamber in 2012 confusion matrix is a useful way to analyze how well tuple systems are from different classes. TP and TN provide correct diagnosing system information, while FP and FN provide information when the diagnosis system is wrong. Thus testing uses the Confusion Matrix which is a matrix to analyze the accuracy of the classification results of a classifier algorithm (Han et al, 2012). The following confusion matrix table in Table 2.

Table 2 Confusion Matrix

| Actual | | Prediction | |
|--------|--|------------|----|
| | | Yes | No |
| Yes | | TP | FN |
| No | | FP | TN |

Table 2 is a confusion matrix table. Sensitivity can be designated as true positive (TP), which is the proportion of positive tuples that are correctly identified. The sensitivity function can be shown with Equation 6 and the accuracy level with Equation 7.

$$\text{Sensitivity} = \frac{TP}{TP+FN} \times 100 \quad (6)$$

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \times 100 \quad (7)$$

where TP is the number of positive / true diagnosis results for positive test data, while FP is the number of negative / false diagnosis results for positive test data. Then TN is the number of positive / true diagnosis results for negative test data, and FN is the number of negative / false diagnosis results for negative test data.

3. RESULTS AND DISCUSSION

Case test data used is data from the damage history report called the Predictive Maintenance (PdM) report from the PLN UPDK Jambi Power Plant. Case history obtained as many as 100 cases or data divided into two parts, 80 data as training data and 20 data as test data.

In this study the type of damage is divided into 5 types of damage according to table 3 including Ski Slope (K01), Imbalance (K02), Misalignment (K03), Looseness (K04) and Roolling Element Bearing Wear (K05). Based on the results of the type of damage output can be tested for accuracy from the results of its classification.

Table 3 Details of Research data

| Type of damage | Total Data | Number of case base | Number of test case |
|-------------------------------------|------------|---------------------|---------------------|
| Ski Slope (K01) | 13 | 9 | 4 |
| Imbalance (K02) | 18 | 15 | 3 |
| Misalignment (K03) | 17 | 13 | 4 |
| Looseness (K04) | 33 | 28 | 5 |
| Roolling Element Bearing Wear (K05) | 19 | 15 | 4 |
| Total | 100 | 80 | 20 |

Evaluation of the results of detection tests with CBR in detecting damage to equipment at the power plant in PLN is done by calculating the value of similarity and accuracy. Evaluation is important to know whether the detection results with the CBR that have been built are feasible to be applied in detecting equipment damage to the power plant at PLN. The process of calculating the value of sensitivity and accuracy can be done by first making a confusion matrix. The following table 4 is a testing table using confusion matrix.

Table 4 Confusion Matrix for CBR Damage Detection Equipment

| | | <i>Prediction</i> | | | | |
|---------------|-----|-------------------|-----|-----|-----|-----|
| | | K01 | K02 | K03 | K04 | K05 |
| <i>Actual</i> | K01 | 3 | 0 | 0 | 0 | 0 |
| | K02 | 0 | 2 | 0 | 0 | 1 |
| | K03 | 0 | 0 | 4 | 0 | 0 |
| | K04 | 0 | 0 | 0 | 5 | 0 |
| | K05 | 0 | 0 | 0 | 0 | 4 |

Furthermore, the test results obtained from Table 4 can be used to calculate the value of sensitivity and accuracy of every class. sensitivity and accuracy value can be found using equation (6) and (7). Results calculated from the value of sensitivity and accuracy are shown in Table 5.

Table 5 Results of testing sensivity and accuracy value each class

| Type of damage | sensivity | accuracy |
|------------------------------------|-----------|----------|
| Ski Slope (K01) | 75% | 100% |
| Imbalance (K02) | 100% | 94,12% |
| Misalignment (K03) | 100% | 100% |
| Looseness (K04) | 100% | 100% |
| Rooring Element Bearing Wear (K05) | 100% | 93,75% |

After testing the sensivity and accuracy value of each class is done, then the next calculation of the overall average grade/group conflict based on the value of sensivity and accuracy. The accuracy of calculation according to the equation.

Table 6 System performance test results with threshold 0.75

| sensivity | accuracy | Presicion |
|-----------|----------|-----------|
| 95% | 97,98% | 95% |

Based on the graph of the test results in table 6 shows that the CBR system to detect equipment damage to power plants in PLN using nearest neighbors produces a sensitivity value of 95% and an accuracy rate of 98.89%.

Testing is done by using a threshold of 0.75. Giving the threshold value is based on the results of a literature review has been done. This shows that the equipment damage detection system at the power plant at PLN can analyze precisely and accurately based on the results of accuracy.

4. CONCLUSIONS

The conclusions of the research conducted are based on tests carried out on 20 test data are:

- 1) This research produces a CBR system to detect damage to equipment at power plants at PLN so as to produce precise and accurate detection.
- 2) The test results of equipment damage test data at the PLN power plant indicate that the system is able to correctly identify the type of damage and damage detection using the *nearest neighbor* method by 97.98%.
- 3) The test results of the equipment damage detection test data at PLN generators with a *threshold* value of 0.75 using the *nearest neighbor*, the system has performance with an accuracy level of 95%.

ACKNOWLEDGEMENTS

The suggestions that can be given in this research are:

- 1) Further research needs to be done to handle the *reuse* process if there are 2 or more cases that have the same similarity value as the *target case* (more than one *reuse* solution).
- 2) Adding the number of objects so that research can be developed in other generators on a national basis.
- 3) Ranking of each feature in order to know the priority of the feature in cases of equipment damage.

REFERENCES

- [1] P. Sumatera and B. Selatan, "Team Predictive Maintenance Sektor Pengendalian Pembangkitan Jambi Pembangkitan Sumatera Bagian Selatan," 2017.
- [2] S. Wulandari, *Case Based Reasoning untuk Mendeteksi Kerusakan pada Mesin Kapal Nelayan*. Yogyakarta: Universitas Gadjah Mada, 2017.
- [3] S. Mulyana and I. Sahputra, "The Determination of the Action towards the Patient's Psychological Therapy in the Post-accident Using Case-based Reasoning," *IJCCS (Indonesian J. Comput. Cybern. Syst.*, vol. 12, no. 1, p. 11, 2018.
- [4] U. A. Mancasari, *Sistem Pakar Menggunakan Penalaran Berbasis Kasus untuk Mendiagnosis Penyakit Syaraf pada Anak*. Yogyakarta: S1 Ilmu Komputer UGM, 2012.
- [5] [6] N. Rokhman, "A Survey on Mixed-Attribute Outlier Detection Methods," vol. 13, no. 1, pp. 39–44, 2019.
- [6] S. Mulyana, S. Hartati, R. Wardoyo, and E. Winarko, "Case-Based Reasoning for Selecting Study Program in Senior High School," vol. 6, no. 4, pp. 136–140, 2015.
- [7] F. Tempola, "Case Based Reasoning For Determining The Feasibility Of Scholarship Grantees Using Case Adaptation," *2018 5th Int. Conf. Inf. Technol. Comput. Electr. Eng.*, pp. 372–376, 2018.
- [8] S. K. Pal and S. C. K. Shiu, *Foundations of Soft Case-Based Reasoning*. 2004.
- [9] J. L. Kolodner, "An introduction to case-based reasoning," *Artif. Intell. Rev.*, 1992.
- [10] M. Muhammad, "Combination Of Case-Based Reasoning And Nearest Neighbour For," pp. 348–352, 2017.