

Application of Classification Algorithm C 4.5 for Predicting Asset Maintenance

<https://doi.org/10.3991/ijes.v7i4.11829>

Muhammad Rizki ^(✉), Deni Mahdiana
Universitas Budi luhur, Kebayoran Lama, South Jakarta
rizkiakak@gmail.com

Abstract—In asset management, determining maintenance actions is one of the problems faced by the company. The importance of maintenance to accelerate the production or performance of a company is now a necessity that must be run. The problem faced by Astra Daihatsu Motor is the difficulty in determining the maintenance action that must be chosen because of information delays when there are assets that are damaged, failed or failure. With the proposal using a decision tree with C4.5 algorithm can predict failures and damage that occur so that it can determine more accurate maintenance actions. Decision tree is a prediction model using tree structure or hierarchical structure. The concept of a decision tree is to transform data into decision trees and decision rules. The main benefit of using a decision tree is its ability to break down complex decision-making processes to be simpler so that decision makers will better interpret the solution of the problem. Using the decision tree method with the C4.5 algorithm can help the problems faced by Astra Daihatsu Motor in determining maintenance. This is shown from the test results of 98.20%. And it can be concluded that the application of C4.5 algorithm is able to produce asset maintenance patterns with better accuracy.

Keywords—Assets, Maintenance, Decision Tree, C4.5 Algorithm, Classification.

1 Introduction

Assets are goods which in the legal sense are called objects, which consist of immovable and moving objects. The intended goods include immovable property (land or building) and movable property, both tangible and intangible, which are included in the assets or assets of a company, business entity, institution or individual, and in the sense of state assets or HKN (State Assets) also consist of the items or objects mentioned above. Including foreign aid that was legally obtained (Siregar. 2004).

Maintenance is all activities carried out to maintain the condition of an item or equipment, or return it to certain conditions. Mentioned in his book defines care or maintenance as a conception of all activities needed to maintain or maintain the quality of facilities / machinery in order to function properly as the initial conditions. (Dhillon, 2006).

Assets have a close problem with maintenance. Some of the problems that often occur within a company are caused by users being late in getting information about asset damage. User delays in getting damaged assets make it difficult for them to control assets so it is difficult to decide on maintenance actions to be taken. As a result can not prevent damage to asset that occurs.

There are several research and asset maintenance prediction techniques conducted by researchers such as those conducted by P. Bastos, I. Lopes and L. Pires (2014) making decision tree models with the C4.5 algorithm. Research conducted by Gian Antonio Susto, Andrea Schirru, Simone Pampuri (2016) discusses the implementation of machine learning with classification methods using SVM and k-NN, they compare the accuracy of these 2 methods. All of the above algorithm models and methods are used to predict damage to assets so as to determine more accurate maintenance actions.

To overcome the above problems, in this study using the C4.5 algorithm decision tree model to form an asset maintenance classification model.

2 Literature Review

2.1 Data mining

Data mining is the process of taking knowledge from large volumes of data stored in databases, data warehouses, or information stored in repositories (Han & Kamber, 2012). Data Mining (DM) is the core of the Knowledge Discovery in Database (KDD) process, which involves algorithms in exploring data, developing models and discovering patterns that were not previously known (Maimon, 2010). This model is used to understand phenomena from data, analysis and prediction. KDD is an organized process to identify patterns that are valid, new, useful, and understandable from a large and complex dataset.

2.2 Data mining classification algorithm

Classification is determining a new data record to one of several categories that have been previously defined or called supervised learning (Hermawati, 2009) The classification process is based on four components (Gorunescu, 2011):

- **Class:** The dependent variable is in the form of a categorical that represents the "label" contained in the object. For example: credit risk, customer loyalty, earthquake type.
- **Predictor:** The independent variable is represented by the characteristics (attributes) of the data. For example: savings, assets, salaries.
- **Training the dataset:** A data set that contains the values of the two components above which are used to determine the suitable class based on.
- **Predictor:** Testing dataset Contains new data to be classified by the model that has been made and classification accuracy is evaluated.

2.3 Classification algorithm C4.5

One of the most popular classification techniques used in the data mining process is classification and decision trees. Decision trees are used to predict object membership for different categories (classes), taking into account values that correspond to their attributes or predictor variables (Gorunescu, 2011). C4.5 algorithm or a decision tree resembles a tree where there are internal nodes (not leaves) that describe the attributes, each branch represents the result of the attribute being tested, and each leaf represents the class. Decision trees can easily be converted to classification rules. In the attribute testing process, new branches that are formed will be considered from the attribute type (Han & Kamber, 2012). There are 3 types of branches that may appear in the decision tree, namely:

1. If the attribute has a discrete value, then the branch formed will always be the same as the number of variations in the value contained in the attribute.
2. If the branch value is continuous, it will be solved according to the split point, while the split point is calculated with each decision tree algorithm. The split branch formed will be patterned like the \leq attribute, and one more branch $>$ attribute.
3. If the attribute being tested is binary, then the branch formed must be two and involve a yes or no value.

The steps in making a decision tree with the C4.5 algorithm (Gorunescu, 2011) are:

- a. Prepare training data, can be taken from historical data that has happened before and has been grouped in certain classes
- b. Determine the root of a tree by calculating the highest gain value of each attribute or based on the lowest entropy index value. Previously calculated the entropy index value, with the formula:

$$Entropi (S) = \sum_{j=1}^k - p_j \log_2 p_j$$

S: is a case set

K: is the number of S partitions

Pj: is the probability obtained from Sum (Yes) divided by Total Cases.

4. Calculate the gain value using the following formula

$$Gain (A) = Entropi (S) - \sum_{i=1}^k \frac{|S_i|}{|S|} \times Entropi(S_i)$$

S = space (data) sample used for training.

A = attribute.

| Si | = number of samples for V.

| S | = the sum of all sample data.

Entropy (Si) = entropy for samples that have a value of i

5. Repeat step 2 until all records are partitioned. As for the partitioning process in the decision tree, it will stop if:
 - a. All tuples in the record in node m get the same class
 - b. There are no attributes in the partitioned record anymore
 - c. There are no records in the blank branch

2.4 Pengujian K-fold cross validation

Cross Validation is a validation technique by dividing data randomly into k sections and each part will be classified (Han & Kamber, 2012). By using cross validation an experiment of k . Each trial will use one testing data and the $k-1$ part will become training data, then the testing data will be exchanged for one training data so that for each experiment different testing data will be obtained. Training data is data that will be used in learning while testing data is data that has never been used as learning and will function as data testing the truth or accuracy of learning outcomes (Witten & Frank, 2011). The data used in this experiment is training data to find the overall error rate value. In general, the k -value test is carried out 10 times to estimate the accuracy of the estimate. In this study the value of k used amounted to 10 or 10-fold Cross Validation.

2.5 Confusion matrix

Confusion Matrix is a visualization tool commonly used in supervised learning. Each column in the matrix is an example of a prediction class, while each row represents the actual event in the class (Gorunescu, 2010). Confusion matrix contains actual and predicted information on the classification system.

3 Methodology

The steps of this research are explained in Figure 1 below:



Fig. 1. Research steps

3.1 Data collection technique

This research data uses secondary data directly taken from PT Astra Daihatsu Motor's database. The data used in this study are asset data with the type of machinery and equipment with active status and have undergone maintenance processes from 2012 to 2018.

The data obtained is then carried out cleaning so that it can be suitable for research needs. The large amount of asset data and maintenance data obtained after cleaning the data, so the data obtained will be manifold. 222. Table 1 is a sample of data that will be used as research material.

Table 1. Examples of data to be used in research

ASSETNUM	WONUM	Description	ASSETYEAR	FAILURESTAT	LOWERWARNING	LOWERACTION	METER
AS9039	WO2034	Tekanan Filter Unit CJC 5A Line	1	yes	2344	3000	GAUG
AS3072	WO12345	Tekanan Filter Unit CJC 4A Line	2	yes	1980	1234	GAUG
AS3072	WO2455	Refraksi Oli Washing Unit 4A Line	4	yes	2300	2455	GAUG
AS3072	WO0987	Viscositas Washing Oli 4A Line	3	yes	4000	2000	GAUG
AS9039	WO3456	Check Visco Oil Washing Unit 5A	2	no	2400	2344	GAUG
AS2974	WO2466	CLEARANCE BRAKE 3B - 1	1	no	2300	1233	GAUG
AS3015	WO3577	CLEARANCE BRAKE 3B - 2	3	no	1200	2340	GAUG
AS2972	WO8766	CLEARANCE BRAKE 3B - 3	1	no	1200	1233	GAUG
AS3016	WO5466	CLEARANCE BRAKE 3B - 4	3	no	1200	3455	GAUG
AS2941	WO3422	CLEARANCE BRAKE 2A - 1	1	no	1200	2333	GAUG
AS2942	WO7655	CLEARANCE BRAKE 2A - 2	2	no	1200	2345	GAUG
AS2943	WO9877	CLEARANCE BRAKE 2A - 3	3	no	1200	1234	GAUG
AS2963	WO7789	CLEARANCE BRAKE 2A - 4	1	no	3400	2200	GAUG
AS5320	WO1267	Motor Hidrolik Hemming 6	4	no	3400	3200	GAUG
AS2941	WO9800	Motor Hidrolik Hemming 7	5	no	3700	1230	GAUG
AS2942	WO7655	Motor Hidrolik Hemming 8	6	yes	2300	2344	GAUG
AS2943	WO5899	Motor Hidrolik Hemming 9	7	yes	2331	1222	GAUG

3.2 Preprocessing

At this stage, determining the dataset that will be created in the data mining application process, then the data set is carried out cleaning data (cleaning). Besides cleaning the dataset, it is also transformed to fit the needs of the data mining application that will be built.

3.3 Model development

At this stage, training data was explored with 3 classifiers, namely C4.5 algorithm, Naïve Bayes (NB) and K-Nearest Neighbor (K-NN) using Rapid Miner. Model development uses training data with 10 fold cross validation.

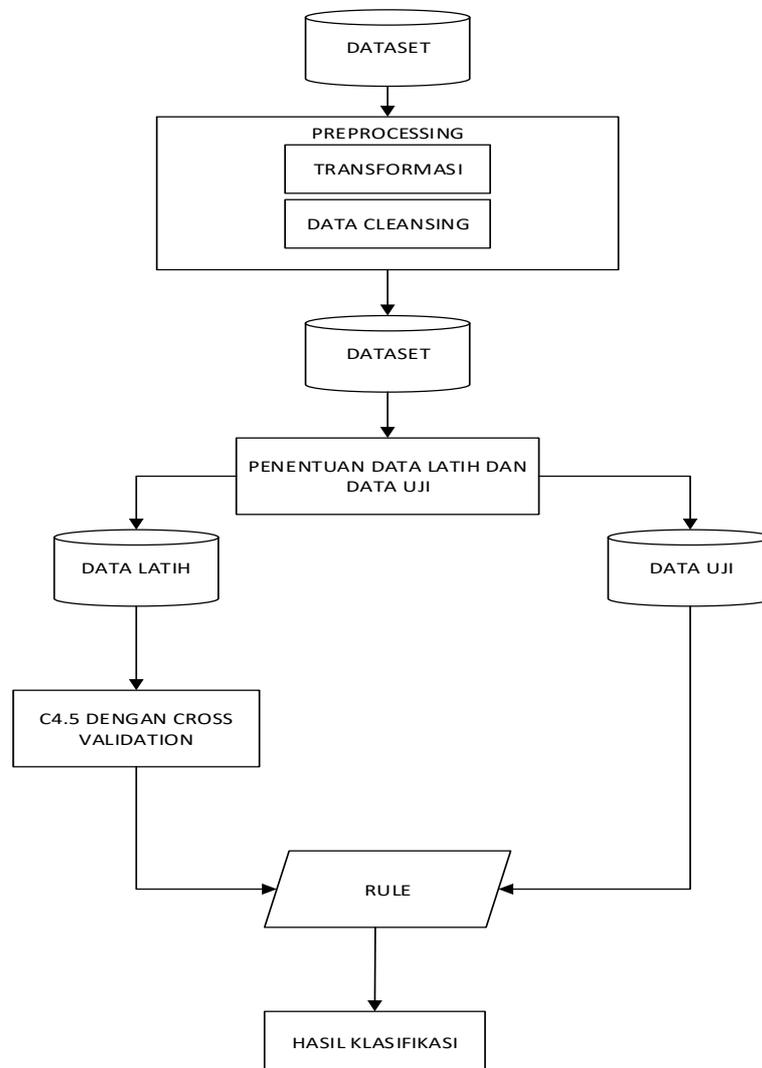


Fig. 2. Proposed model

3.4 Pengujian model

The resulting model is tested using a confusion matrix. The test aims to determine the accuracy, precision, recall using equations.

Table 2. Confusion Matrix 2 class

Classification	Predicted class	
	Class=Yes	Class=No
Class=Yes	a (true positive)	b (false negative)
Class=No	c (false positive)	d (true negative)

In the confusion matrix table above, true positive (TP) is the number of positive records classified as positive, false positive (FP) is the number of negative records classified as positive, false negatives (FN) is the number of positive records classified as negative, true negatives (TN) is the number of negative records classified as negative. After the test data is classified, a confusion matrix will be obtained so that the amount of sensitivity, specificity, and accuracy can be calculated.

Sensitivity is the proportion of class = yes that is correctly identified. Specificity is the proportion of class = no that is correctly identified. For example in the classification of computer customers where class = yes is a customer who buys a computer while class = no is a customer who does not buy a computer. The sensitivity is 95%, meaning that when a classification test is performed on a customer who purchases, then that customer has a 95% chance of being positive (buying a computer). If a specificity of 85% is produced, meaning that when a classification test is performed on customers who do not buy, then the customer has a 95% chance of being negative (not buying). The formula for calculating accuracy, specificity, and sensitivity in a confusion matrix is as follows (Gorunescu, 2011)

$$Akurasi = \frac{TP + TN}{TP + TN + FP + FN} = \frac{a + d}{a + b + c + d}$$

$$Sensitivitas = \frac{TP}{TP + FN} = \frac{a}{a + b}$$

$$Spesifisitas = \frac{TN}{TN + FP} = \frac{d}{d + c}$$

3.5 Prototype development

In this section, we will explain the prototype description that will be built in the form of a use case diagram below:

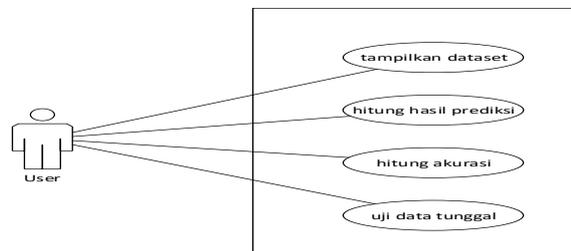


Fig. 3. Use Case Prototype

Actors from this system are only users. The main function that can be done by users is to do classification. Users can also import data and measure accuracy of existing data and prediction results. Amount of data classified correctly and incorrectly from processing with the C4.5 algorithm.

3.6 Prototype testing

The prototype that has been completed is then tested for its accuracy. Testing is done in two ways, namely overall data testing and single data testing. Testing is done by testing the rules made from the classification results.

4 Experimental Result

4.1 Data collection

In this study the data used comes from a combination of asset data and maintenance data. In this study looking for relationships between several attributes of the two data. The following data sources are used in this study:

Assets data: Asset data is data originating from PT. Astra Daihatsu Motor after the data is registered and input into the database. The asset data contains the asset identity data

Table 3. Asset Data

Atribut	Description
Assetnum	The asset identification number when registered for the first time
Description	Description of each asset
Assetclass	Classes in assets consist of 5 classes, namely machinery & equipment, office tools, vehicles, buildings, low value assets
Failurestat	Failure status of the asset, if the asset is damaged
Location	Location of the place where the asset is located
Assetyear	This attribute is information about the age of an asset, which contains a number and is calculated from the first time the asset is registered
Vendor	This attribute is information about the vendor from which the asset was first purchased

Maintenance data: Maintenance data is asset data that has been declared damaged or has failed so that it can hamper production.

Table 4. Data Maintenance

atribut	Description
WONUM	An ID when the supervisor carries out maintenance for damaged assets
Lower warning	Is a minimum warning for some assets that are indicated when they need to be maintained, the lower warning value, the smaller the value, the more dangerous the asset is
Lower action	It is an action taken by the supervisor or the related PIC when the lower warning is indicated
Upper warning	It is a maximum warning for some assets that have to be maintained, the higher the value of this warning the more dangerous the asset is
Upper action	It is the supervisor's or PIC's action when the assets have been indicated to be maintained and there is an upper warning
Supervisor	This attribute is the supervisor who is responsible for the maintenance of the asset
Meter type	This attribute is a measure when assets must be maintained Example: Celsius is a measure when the assets have gone too far the temperature limit and must be maintained

4.2 Preprocessing

Preprocessing activities carried out in this study are as follows:

1. **Transformation:** Data transformation is the process of converting data into other formats that suit research needs. At this stage data transformation will be carried out in the form, format or other data structures, adjusted to the needs from the analysis side.
2. **Asset age** is used as a basis in determining whether the asset title must be maintained or not, the previous asset age data is only a register date when the asset is registered, then changed for research needs by calculating the life span of the asset from the asset being registered. The asset life transformation can be seen in table 5

Table 5. Transforming the age of assets

Register Date	New Atribut
2013-02-22	5 tahun
2014-11-01	4 tahun

3. **Data cleaning:** The cleaning process in this research includes removing data duplication, checking inconsistent data, and correcting errors in the data. The cleaning process is done manually in collaboration with a Supervisor who understands the condition of the assets and is fully responsible for the asset data that is in PT. Astra Daihatsu Motor.

assetnum	description	lowerwarning	loweraction	upperwarning	upperaction	siteid	orgid
AS3072	Viscositas Washing Oli 4A Line	4.000	3.500	7.000	8.000	ADMP1	ADM
AS9039	Check Visco Oil Washing Unit 5A Line	6.000	3.000	7.000	13.000	ADMP1	ADM
AS2974	CLEARANCE BRAKE 3B - 1	NULL	NULL	NULL	NULL	ADMP1	ADM
AS3015	CLEARANCE BRAKE 3B - 2	NULL	NULL	NULL	NULL	ADMP1	ADM
AS2972	CLEARANCE BRAKE 3B - 3	NULL	NULL	NULL	NULL	ADMP1	ADM
AS3016	CLEARANCE BRAKE 3B - 4	NULL	NULL	NULL	NULL	ADMP1	ADM
AS2941	CLEARANCE BRAKE 2A - 1	NULL	NULL	NULL	NULL	ADMP1	ADM

Fig. 4. Data before cleaning

In Figure 4 you can see a lot of data containing "NULL" because the user is lazy to fill in or as long as he does the filling. So that existing data can be used in research it is necessary to correct inconsistent data by filling in according to the conditions in the field. If indeed the data is not found in the field then the data will be deleted. Besides fixing inconsistent data, some duplicated data were found to be 500 records.

assetnum	description	lowerwarning	loweraction	upperwarning	upperaction
ASP50...	CM FDBA - 10 - 1 - 3 Bulan__PBS 1...	5.000	4.000	99998.000	99999.000
ASP50...	CM FDBA - 11 - 3 Bulan__PBS 1 - Fr...	19.000	17.000	99998.000	99999.000
ASP50...	CM FDBA - 11 - 1 - 3 Bulan__PBS 1...	5.000	4.000	99998.000	99999.000
ASP50...	CM FDBA - 12A - 3 Bulan__PBS 1 - ...	19.000	17.000	99998.000	99999.000
ASP50...	CM FDBA - 12A - 1 - 3 Bulan__PBS...	5.000	4.000	99998.000	99999.000
ASP50...	CM FDBA - 12B - 3 Bulan__PBS 1 - ...	19.000	17.000	99998.000	99999.000
ASP50...	CM FDBA - 12B - 1 - 3 Bulan__PBS...	5.000	4.000	99998.000	99999.000

Fig. 5. Duplicate data on data maintenance

4.3 Data set creation

At this stage, determining the data set that will be used in the data mining application process, then the data set is carried out cleaning data to accelerate the data mining process as explained in the previous process. The data set used must be in accordance with the data you want to use in the data mining process. The data are asset number, asset year, failure stat, lower warning, lower action, upper warning, upper action. The data is taken from the asset data table and maintenance data that is uploaded into the system in Excel format. In this study the data used database-based files and to speed up the process, the data used when processing data mining uses only one table, as shown in Figure 6.

TABEL ASSET
ASSETNUM
ASSETYEAR
FAILURESTAT
LOWERWARNING
LOWERACTION
UPPERWARNING
UPPERACTION
MAINTENANCE ACTION

Fig. 6. Data to be used for the data mining process

4.4 Determination of class data values

The process of determining the label class in this study is still done manually

Table 6. Asset age classes

Umur aset	Status maintenance
>4 tahun	Harus dimaintenance
≤4 tahun	Tidak harus di maintenance

From the table above, the asset data that must be maintained using the asset lifespan attribute can be categorized into 2 parts:

- Maintenance, if the age of the asset is more than 4 years
- Does not have to be maintained, if the age of the asset is still under 4 years

As for the second maintenance parameter, the attributes taken from the maintenance table. The process of selecting data contained in this maintenance table is based on the type of asset that has been determined at the beginning. Because not all types of assets perform the same maintenance actions. According to the research needs maintenance data taken is maintenance data with the type of "machinery and equipment" assets. The class values of the attributes in this maintenance table are explained in table 7.

Table 7. Warning and maintenance classes

Atribut	value	Maintenance status
Lower warning	≤ 2000	Maintenance
Lower action	< 2000	Maintenance
Upper warning	≥ 5000	Maintenance
Upper action	> 5000	Maintenance

The data class used for data mining is prepared, so it has binominal and polynomial classes according to the rules that have been created based on the value of the data. Table 8 is the division of variables and data classes used in data mining analysis.

Table 8. Division of data types

nama field	Jenis kelas data	Kelas Data yang digunakan
ASSETYEAR	Polynominal	> 4 harus di maintenance, <4 tidak harus di maintenance
FAILURESTAT	binominal	yes, no
LOWERWARNING	Polynominal	≤ 2000 harus di maintenance, > 2000 tidak harus dimaintenance
LOWERACTION	Polynominal	≤ 2000 harus di maintenance, > 2000 tidak harus dimaintenance
UPPERWARNING	Polynominal	≥ 5000 harus di maintenance, < 5000 tidak harus dimaintenance
UPPERACTION	Polynominal	≥ 5000 harus di maintenance, < 5000 tidak harus dimaintenance

4.5 Model testing

The purpose of this study is to analyze asset maintenance predictions by applying data mining classification techniques using the decision tree C4.5 algorithm. The testing stage of this model the data used is compared with the decision tree algorithm (C4.5), the Naïve Bayes algorithm (NB), and the K-Nearest Neighbor (K-NN) algorithm, and then tested using cross validation. The cross-validation method is used to avoid overlapping the testing data. The stages of cross-validation are as follows:

- Divide the data into k subsets of the same size
- Use each subset of testing data and the rest for training data

The standard evaluation method is a 10-fold cross-validation stratified. Why 10? The results of extensive experiments and theoretical evidence show that 10-fold cross-validation is the best choice to get accurate validation results. 10-fold cross-validation will repeat the test 10 times and the measurement result is the average value of 10 times the test. Model design for testing uses decision tree algorithms (c4.5), naïve bayes (NB) algorithm, and K-Nearest Neighbor (K-NN) algorithm shown in the figure below.

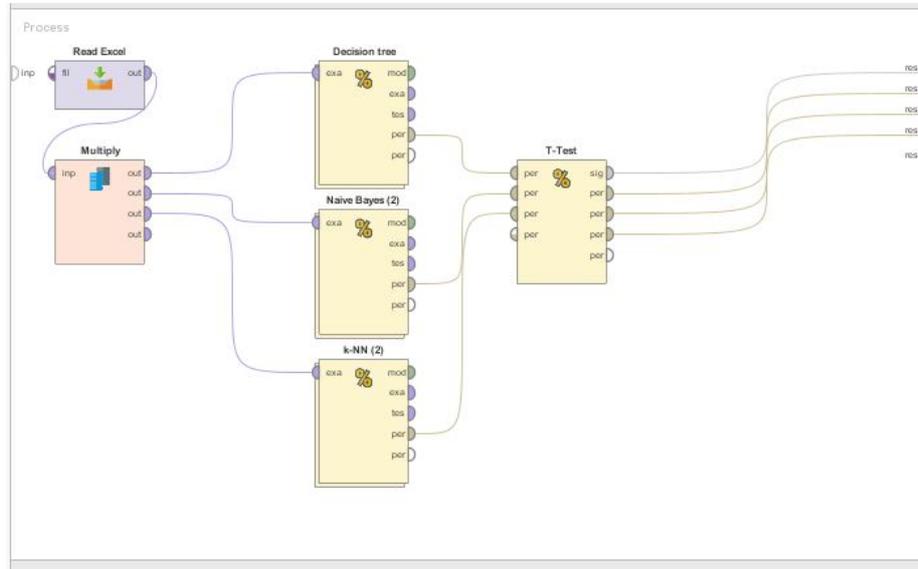


Fig. 7. Testing Model Design

The results of testing the three algorithms above using 10-fold cross validation are shown in the table

Table 9. Results of T-test Differences

Algoritma	C4.5	NB	K-NN
Accuracy	98,66%	90,51%	88,72%
AUC	0,895	0,995	0,891

4.6 Validation and evaluation

The purpose of this research is to analyze asset maintenance predictions by applying data mining classification techniques to the decision tree C4.5 algorithm. At the testing stage of this model, the data used has passed the preprocessing stage. The design model to be used is shown in Figure 8.

- **Read excel:** This operator is used to import the dataset to be used, in this study the data is imported from the xls file
- **Validation:** The validation method used in this study is a sampling technique
- **Decision tree:** The classification method used in this study
- **Apply model:** Operator used in C4.5 research
- **Performance:** The operator used to measure the performance accuracy of the model

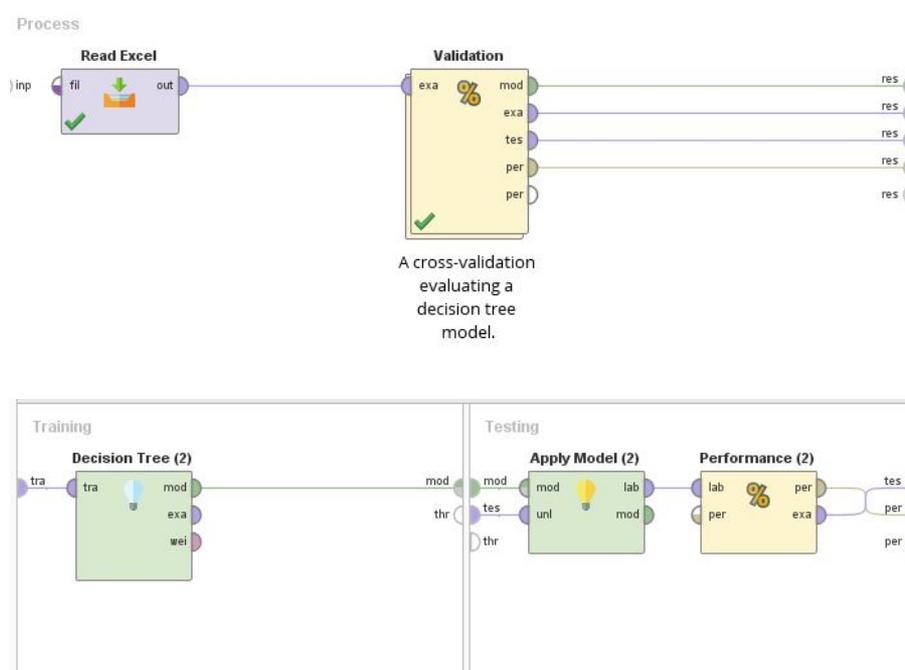


Fig. 8. Model Design Process Validation

Testing will be carried out from the training data population. The amount of training data to be used is 222 with an error rate of 5%, both maintenance predictions and simple random sampling. Judging from the results of each method of attribute selection, the results show similarity. From the test results above, the level of accuracy will be evaluated using a model that is using a confusion matrix.

4.7 Confusion matrix evaluation

After the tested data is entered into the confusion matrix, calculate the values that have been entered to calculate the amount of sensitivity, specificity, precision and accuracy. Sensitivity is used to compare the number of true positives to the number of tuples that are positives while the specificity is to compare the number of true negatives to the sphere of negative tuples. Seen in the picture ... this shows the value of accuracy, recall, and precision produced by rapidminer using the confusion matrix model:

PerformanceVector

```
PerformanceVector:
accuracy: 98.20% +/- 3.00% (micro average: 98.20%)
ConfusionMatrix:
True:  MAINTENANCE    NO
MAINTENANCE:    176    1
NO:             3     42
precision: 95.89% +/- 8.97% (micro average: 93.33%) (positive class: NO)
ConfusionMatrix:
True:  MAINTENANCE    NO
MAINTENANCE:    176    1
NO:             3     42
recall: 98.33% +/- 5.00% (micro average: 97.67%) (positive class: NO)
ConfusionMatrix:
True:  MAINTENANCE    NO
MAINTENANCE:    176    1
NO:             3     42
AUC (optimistic): 1.000 +/- 0.000 (micro average: 1.000) (positive class: NO)
AUC: 0.988 +/- 0.025 (micro average: 0.988) (positive class: NO)
AUC (pessimistic): 0.976 +/- 0.050 (micro average: 0.976) (positive class: NO)
```

Fig. 9. Confusion Matrix results

The accuracy generated from the C4.5 algorithm modeling is 98.20%. With the number of true positive (tp) of 176 records, false positive (fp) of 3 records, the number of true negative (tn) of 42 records, and the number of false negative (fn) of 1 record.

4.8 Testing result

Testing of asset data using the decision tree method produces a classification tree. These results can be used as a strategic information that can be converted into knowledge. This knowledge can be used as a supporter of strategic decisions or policies for an organization.

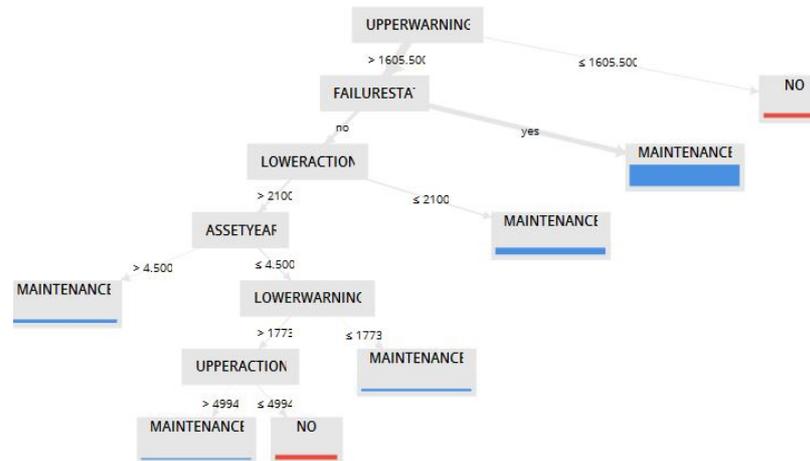


Fig. 10. Decision Tree Results

The following are some asset criteria that can be applied as a strategic policy for Astra Daihatsu Motor based on research interpretation.

1. If upperwarning has exceeded the maximum limit: These assets must be maintained immediately before a failure occurs on these assets. This can be prevented by doing preventive maintenance and annual maintenance
2. If upperwarning has not passed the minimum threshold: Asset conditions must still be considered, conditions have not passed the upper warning not necessarily the assets are free from maintenance, before that happens PIC must pay attention to other conditions such as lower warnings and the age of assets that continue to grow from time to time
3. If the age of the asset is still under 4 years: Assets that are under 4 years of age are not necessarily free from damage, in fact from the results of the classification of assets under 4 years of age there is still a lot of damage.

Based on the results of system testing that has been done on data taken from PT Astra Daihatru Motor, it can be seen that there are advantages and disadvantages to the old system and the new system.

By using the old system, still using the estimated system so that the error rate for predicting maintenance is still large. Whereas by using this data mining technique, the error rate in predicting maintenance can be reduced by an error rate of 5%.

By using the old system, the decision making is still complex and global. Meanwhile, after using a new system of decision making areas that were previously complex and very global, can be changed to be more simple and specific.

By using the old system, a tester still has difficulty in analyzing to estimate either the high dimensional distribution or certain parameters of the class distribution.

Whereas with the new system in analysis, with very many criteria and classes a lot of testers usually need to estimate either the high dimensional distribution or certain parameters of the class distribution. The decision tree method can avoid this problem by using fewer criteria at each internal node without greatly reducing the quality of the resulting decisions.

4.9 Prototype making

In this section, we will discuss making prototypes that have been described using the use case diagram in chapter 3.5.

1. **Home page:** The home page is the first page that appears when the system starts. This page does not have the main function of the system, but only displays an image of the name of the system, or can be called a welcome screen page.

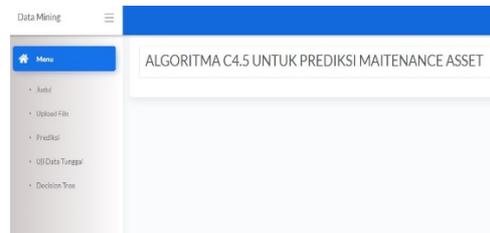


Fig. 11. Prototype Home page

2. **Data upload page:** On this page is a page to enter the dataset by uploading excel data format (xls) then select the data that previously had to be preprocessed.

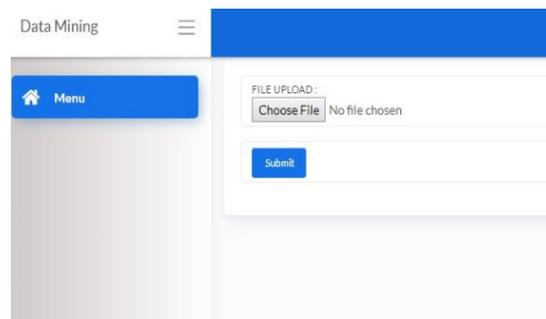


Fig. 12. Data Upload Page

3. **Data set display page:** On this page is a dataset display page that has been uploaded based on an excel file that has been selected by the user, the prediction column is empty because it has not yet made the prediction process.

Show 10 entries

JM	ASSET YEAR	FAILURE STAT	LOWER WARNING	LOWER ACTION	UPPER WARNING	UPPER REACTION	MAINTENANCE ACTION	PREDIKSI
	1	no	1200	2333	4890	4988	MAINTENANCE	
	5	no	3700	1230	5311	4687	MAINTENANCE	
	4	no	1222	2345	4777	7000	MAINTENANCE	
	4	no	2344	1234	4233	4600	MAINTENANCE	
	2	no	1200	2345	4777	5200	MAINTENANCE	

Fig. 13.Data Set Display Page

4.10 Prototype testing

The dataset that appears on the previous data display page has not released the prediction results, to calculate and display the predicted results the user must press the process button to calculate the overall data.

Show 10 entries

	ASSET YEAR	FAILURE STAT	LOWER WARNING	LOWER ACTION	UPPER WARNING	UPPER REACTION	MAINTENANCE ACTION	PREDIKSI
	1	yes	3489	2200	5600	10000	MAINTENANCE	maintenance
	4	yes	3489	1234	4890	4922	MAINTENANCE	maintenance
	3	no	3489	2345	4476	4977	NO	no
	1	no	3489	2333	4777	4788	NO	no
	3	no	3489	3455	0	0	NO	no
	1	no	3489	1233	4444	4577	MAINTENANCE	no
	6	yes	3489	2340	4776	6700	MAINTENANCE	maintenance

Fig. 14.Overall Data Prediction Results

The picture above shows the appearance of the dataset with predicted results, then what percentage of accuracy will be generated by comparing the amount of correct prediction data and the total data overall.

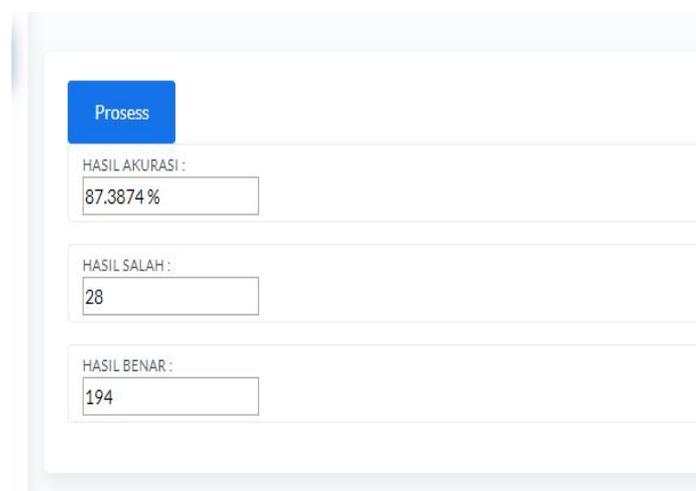


Fig. 15.Prototype Accuracy Results

Figure 15 is the result of accuracy produced with the prototype model that was made. Accuracy results are obtained from a comparison between the value of the correct accuracy compared to the overall data then multiplied by 100%.

5 Conclusion

Based on research that has been done, C4.5 algorithm can be applied to predict asset maintenance. By using the C4.5 algorithm to determine maintenance users can find out before damage occurs. This can be seen in the results of testing using rapid miners, predictions using the C4.5 algorithm produce an accuracy of 98.20%.

The advice given for further research is the use of the C4.5 algorithm combined with other methods. This is done with the hope that the accuracy to make predictions will improve. The process of determining the label class on the data is still done manually with the help of related supervisors, it is necessary to do further research by combining clustering and classification methods. Prototyping must be developed to improve the accuracy of prediction results, by testing 10 fold cross validations on the prototype.

6 Reference

- [1] Akdon & Riduwan. (2010). Rumus dan Data Analisis Statistika, Cet 2. Alfabeta.
- [2] Aprilla, Dennis, Donny Aji Baskoro, Lia Ambarwati, dan I Wayan Simri Wicaksana. (2012). Belajar Data Mining dengan RapidMiner. Jakarta: Gramdia Pustaka Utama.
- [3] Barry, Render dan Jay Heizer. (2001). Prinsip-prinsip Manajemen Operasi: Operations Management. Jakarta: Salemba Empat.
- [4] Berry, Michael J.A dan Linoff, Gordon S. (2004). Data Mining Techniques For Marketing, Sales, Customer Relationship Management Second Editon. United States of America: Wiley Publishing, Inc.

- [5] Bertalya. (2009). *Konsep Data Mining, Klasifikasi: Pohon Keputusan*. Jakarta: Universitas Gunadarma.
- [6] Corder, Antony dan kusnul Hadi. (1992). *Teknik Manajemen Pemeliharaan* Jakarta: Erlangga.
- [7] Dhillon, B. S., (2006), *Maintainability, Maintenance, and Reliability for Engineers*, Taylor & Francis Group, New York.
- [8] Efraim Turban, dkk. (2005). *Decision Support Systems and Intelligent Systems*. Yogyakarta: ANDI
- [9] Fayyad, Usama. 1996. *Advances in Knowledge Discovery and Data mining*. MIT Press.
- [10] Flavio Trojan. (2017). "Maintenance-types Classification to Clarify Maintenance Concepts in Production and Operations Management".
- [11] Gian Antonio Susto, Andrea Schirru, Simone Pampuri. (2016). "Machine Learning for Predictive Maintenance: a Multiple Classifier Approach".
- [12] Gorunescu, F. (2011). *Data Mining Concept Model and Techniques*. Berlin: Springer.
- [13] Han, J, Kamber, M, & Pei, J. (2006). *Data Mining: Concept and Techniques*, Second Edition. Waltham: Morgan Kaufmann Publishers.
- [14] Han, J, Kamber, M, & Pei, J. (2012). *Data Mining: Concept and Techniques*, Third Edition. Waltham: Morgan Kaufmann Publishers.
- [15] Haryono, A. (2007). *Modul Prinsip dan Teknik Manajemen Kekayaan Negara*. Tangerang: Badan Pendidikan dan Pelatihan Keuangan, Pusdiklat Keuangan Umum.
- [16] Kusnawi. (2007). *Pengantar Solusi Data Mining*. Seminar Nasional Teknologi 2007 (SNT). Yogyakarta: STMIK AMIKOM Yogyakarta.
- [17] Kusrini, luthfi taufiq Emha, (2009), *Algoritma Data Mining*, Penerbit Andi, Yogyakarta.
- [18] Larose D, T., (2005), *Discovering knowledge in data: an introduction to data mining*, Jhon Wiley & Sons Inc.
- [19] P. Bastos, I. Lopes dan L. Pires. (2014). "Application of data mining in a maintenance system for failure prediction".
- [20] Purba, Yugi Trianto. (2008). *Penerapan Data Mining Untuk Mengetahui Pola Antara Nilai Ujian Saringan Masuk (USM) Terhadap Indeks Prestasi (IP)*.
- [21] Ruben Sipos, Dmitriy Fradkin, Fabian Moerchen, Zhuang Wang. (2014). "Log-based Predictive Maintenance".
- [22] Sibaroni. (2008), *Analisis Dan Penerapan Metode Klasifikasi Untuk Membangun Perangkat Lunak Sistem Penerimaan Mahasiswa Baru Jalur Non Tulis*, Tesis, S2 Institut Teknologi Bandung.
- [23] Siregar, D. D. (2004). *Manajemen Aset. Strategi Penataan Konsep Pembangunan Berkelanjutan Secara Nasional dalam Konteks Kepala Daerah Sebagai CEO' spada Era Globalisasi dan Otonomi Daerah*. PT. Gramedia Pustaka Utama. <https://doi.org/10.20885/jsb.vol1.iss9.art1>
- [24] Sugiyono. (2010). *Metode Penelitian Pendidikan Pendekatan Kuantitatif, kualitatif, dan R&D*. Bandung : Alfabeta.
- [25] Swasti R. Khuntia, José Luis Rueda, Mart Meijden dan Sonja Bouwman. (2015). "Classification, domains and risk assessment in asset management "
- [26] Witten, I. H., Frank, E., Hall, M. A. (2011). *Data Mining Practical Machine Learning Tools and Techniques* (3rd ed). USA: Elsevier.

Article submitted 2019-09-06. Resubmitted 2019-10-21. Final acceptance 2019-10-22. Final version published as submitted by the authors.