International Journal of Mechanical, Industrial and Aerospace Sciences
ISSN: 2517-9950
Vol:7, No:7, 2013

Estimation Model of Dry Docking Duration Using Data Mining

Isti Surjandari and Riara Novita

***Abstract*—**Maintenance is one of the most important activities in the shipyard industry. However, sometimes it is not supported by adequate services from the shipyard, where inaccuracy in estimating the duration of the ship maintenance is still common. This makes estimation of ship maintenance duration is crucial. This study uses Data Mining approach, i.e., CART (Classification and Regression Tree) to estimate the duration of ship maintenance that is limited to dock works or which is known as dry docking. By using the volume of dock works as an input to estimate the maintenance duration, 4 classes of dry docking duration were obtained with different linear model and job criteria for each class. These linear models can then be used to estimate the duration of dry docking based on job criteria.

***Keywords*—**Classification and regression tree (CART), data mining, dry docking, maintenance duration.

I. Introduction

MAINTENANCEis one of the most important activities in the shipping industry because it can determine the airworthiness of the ship. However, it is often occurs that this activity is not supported by adequate facilities as compared to the increases in marine transport activities. The capacity of national shipyard for maintenance activities is very limited. It makes the estimated duration of ship maintenance becomes crucial. If the estimated duration of maintenance is too long, then shipyard becomes uncompetitive. On the other hand, if the estimated duration is too short, it is sometimes constrained by the availability of spare-parts which are required for the ship maintenance process [1]. Usually, the duration of ship maintenance is estimated based on operator experiences. There is no standard method that is used to estimate the duration of ship maintenance. The objective of this study is to model the estimated duration of ship maintenance that is performed on dock (i.e., dry docking maintenance)by using Data Mining approach.

Data mining is an analysis method of large amounts of data to find a hidden pattern, in this case, the relationship between maintenance data and its duration. So that, by knowing the volume of each type of dry docking work to be done, the operator can directly estimate the dry docking duration.

Isti Surjandari is Professor at the Industrial Engineering Department, Faculty of Engineering, University of Indonesia, Kampus UI, Depok 16424, Indonesia (phone: +6221 78888805; e-mail: isti@ie.ui.ac.id).

Riara Novita is graduate student in the Industrial Engineering Department, Faculty of Engineering, University of Indonesia, Kampus UI, Depok 16424, Indonesia (e-mail: riara.novita@gmail.com).

II. Methods

Data Mining is the analysis of (often large) observational data sets to find unsuspected relationships and to summarize the data in novel ways that are both understandable and useful to the data owner [2]. A given Data Mining project has a life cycle consisting of six phases, those are (1) business understanding, (2) data understanding, (3) data preparation, (4) modeling, (5) evaluation, and (5) deployment phase.

There are six main tasks that Data Mining is usually called upon to accomplish: Description, Estimation, Clustering, Association, Classification, and Prediction [3]. *Classification And Regression Tree* (CART) is one of Data Mining method used for estimation.

CART is one of the decision tree algorithms for classification by constructing a flowchart-like structure where each internal node represents a test on an attribute, each branch denotes an outcome of the test, and each external node means a class prediction [4]. The characteristic of CART is to use a set of "if-then" conditions to perform predictions or classification of cases such that CART is very suitable to either large problems or small data set with both continuous and categorical variables [4], [5]. Moreover, the attribute that is not appeared in the tree is assumed to be irrelevant in the analysis because CART ranks the attributes by giving their respective weights [4]. Therefore, the set of attributes appearing in the tree forms the reduced subset of attributes.

CART analysis consists of four basic steps. The first step consists of tree building, during which a tree is built using recursive splitting of nodes. Based on *goodness of split* criterion, the tree-growing algorithm finds the best attribute as a parent node. Parent node is an attribute with the purest node among the others. Mostly, CART uses Gini index to get node purity of each attribute, which is defined by (1).

$$\Delta i(\tau) = i(\tau) - \{p_L i(\tau_L) + p_R i(\tau_R)\} \qquad (1)$$

TABLE I
CART Classification

Split	Class of Y		Row Total
	1	0	
$x_i \le c$	n_{11}	n_{12}	n_{1+}
$x_i > c$	n_{21}	n_{22}	n_{2+}
Column Total	n_{+1}	n_{+2}	n_{++}

where

$$i(\tau) = 2\left(\frac{n_{+1}}{n_{++}}\right)\left(1 - \frac{n_{+1}}{n_{++}}\right) = 2\left(\frac{n_{+1}}{n_{++}}\right)\left(\frac{n_{+2}}{n_{++}}\right) \quad (2)$$

$$i(\tau_L) = 2\left(\frac{n_{11}}{n_{1+}}\right)\left(1 - \frac{n_{21}}{n_{1+}}\right) = 2\left(\frac{n_{11}}{n_{1+}}\right)\left(\frac{n_{12}}{n_{1+}}\right) \quad (3)$$

$$i(\tau_R) = 2\left(\frac{n_{21}}{n_{2+}}\right)\left(1 - \frac{n_{21}}{n_{2+}}\right) = 2\left(\frac{n_{21}}{n_{2+}}\right)\left(\frac{n_{22}}{n_{2+}}\right) \quad (4)$$

The second step consists of stopping the tree building process, at this point a "maximal" tree has been produced or the number of cases for each terminal node is less than minimum required, which is 5 cases or 10% of all cases in training set.

The third step consists of tree "pruning," which results in the creation of a sequence of simpler and simpler trees based on re-substitution error rate in every terminal node, as in (5).

$$R(\tau_l) = r(\tau_l)p(\tau_l) \quad (5)$$

where

$$r(\tau_l) = 1 - max(p, 1-p) = min(p, 1-p) \quad (6)$$

$p(\tau)$is the probability that an observation falls into node τ.

The fourth step consists of optimal tree selection, during which the tree that fits the information in the learning dataset, but does not overfit the information, is selected from among the sequence of pruned trees by using testing data [6], as in (7) and cross validation, as in (8).

$$R^{ts}(T_*) = min_k R^{ts}(T_k) \quad (7)$$

$$R^{CV/V}(T_*) = min_k R^{CV/V}(T_k) \quad (8)$$

where

$$R^{CV/V}(T_k) = R^{CV/V}\left(T(\alpha_k)\right) \quad (9)$$

$$R^{CV/V}\left(T(\alpha)\right) = N^{-1}\sum_{i=1}^{J}\sum_{j=1}^{J} n_{ij}(\alpha) \quad (10)$$

$$n_{ij}(\alpha) = \sum_{v=1}^{v} n_{ij}^{(v)}(\alpha), i,j = 1,2,\dots,J \quad (11)$$

For the purpose of this study, an empirical data will be obtained from one of the largest shipyard in Jakarta, i.e., the capital city of Indonesia. Indonesia is an archipelago country, where sea transportation is important, and Jakarta port is the largest and busiest seaport in Indonesia.

Docking is a ship displacement from water/sea to the dock with the help of docking facilities. Dry docking is performed in a large pond located on the seafront where its construction material is composed of concrete and steel. Generally, the docking process begins with the preparation of a maintenance works list (repair list) that is obtained from survey conducted by the shipyard operator (i.e., observation and interview) and also based on the owner requests, which is done either the ship is still on sailing or leaning at the port (at least 2 months before entering the dock). After the job list is approved by both parties, then the maintenance works can be done immediately.

Dry docking works consist of: (1) the plate hull replacement (re-plating), (2) rudder maintenance and repair, (3) propeller maintenance and repair, (4) hull maintenance and repair, (5) anchors and anchor chains maintenance and repair, (6) the maintenance and repair of suction valve and scupper valve, (vii) maintenance of ship tank.

A. Dry Docking Attributes

Every shipyard usually has satisfaction note, a report of maintenance works that have already done. Actually, the data in satisfaction note can be used not only to determine the maintenance costs but also to estimate the maintenance duration. Table II shows the data specification for each attribute of dry docking activities.

TABLE II
DRY DOCKING ATTRIBUTES SPECIFICATION

Attributes	Data Specification
size, tank, anchor chain tub	Volume (m³)
grt, plate	Mass (ton)
scraping, sandblasting, washing, painting	Wide (m²)
zinc, seal-prop, ring-prop, packing-prop, chrome, ring-rud, packing-rud, seal-rud, chest, valve, scrupper, manhole, plug	Amount
ut, weld	Amount (point)
grease	Mass (kg)
propeller, shaft-prop, rudder, shaft-rud	A = no maintenance, B = recondition, C = balancing, D = change, E = recondition and balancing, F = change and balancing, G = change and recondition, H = change, recondition, and balancing
bearing-prop, bearing-rud	A = no maintenance, B = recondition, C = change, D = balancing
shaftseal, anchor	Options (YES, NO)
weldlength	Length (m)

m = meter and kg = kilogram.

B. Attribute Reduction

To avoid the negative effect of irrelevant, useless, and redundant of attributes on the historical data, then an attributes reduction must be done. Attributes reduction is done based on interview with the shipyard operator who is an expert in the maintenance activities. The attributes which are identified as irrelevant are those whose influence on the overall works on dry docking duration were considered as insignificant by the operator. Redundant attributes are those attributes that contain information which already included in the information contained in some other attribute(s). While attributes that have the same values for all or most of the data are marked as useless for the available dataset, because such attributes will give only very small effect on the learning model [1]. Eventually, 28 significant attributes out of 35 attributes were obtained and will be used for this study. Those 28 attributes are: tank, plate, scraping, sandblasting, washing, painting, zinc, seal-prop, ring-prop, packing-prop, chrome, ring-rud, packing-rud, seal-rud, chest, valve, scrupper, ut, welding, grease, propeller, shaft-prop, rudder, shaft-rud, bearing-prop,

International Journal of Mechanical, Industrial and Aerospace Sciences
ISSN: 2517-9950
Vol:7, No:7, 2013

bearing-rid, shaftseal, and weldlength.

C. Model Validation

Validation of the developed model is done by using the same procedure as was done in the development of the model; it is just that in this validation step will be done using the K-Nearest Neighbor (k-NN) method [7]. The difference between CART and k-NN method lies in the number of attributes that will be used. CART generalizes all attributes in the data set, so that the linear models only consist of significant attributes. On the other hand, k-NN does not generalize the attributes, so that it uses all attributes to estimate dry docking duration.

III. Analysis

By using Weka (Waikato Environment for Knowledge Analysis) software, the classification tree is depicted in Fig. 1.

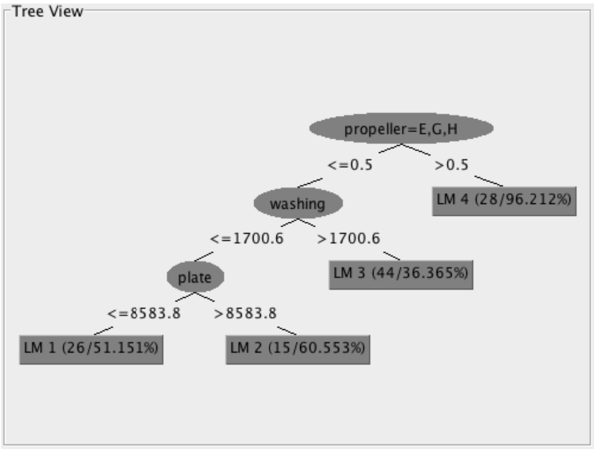

Fig. 1 Classification tree of dry docking duration

Based on classification tree in Fig. 1, it can be seen that dry docking duration is classified based on three attributes; those are propeller, washing, and plate. The first classification is propeller. If there is work to be done on the propeller, then the linear model to estimate the dry docking duration is LM4; otherwise the estimation of dry docking duration should be classified based on washing works. If the washing surface is more than 1700.6 m², then the linear model is LM3;otherwise the estimation of dry docking duration should be classified based on plate works. If the weight of plateis more than 8583.8 ton, then the linear model is LM1; otherwise use LM2. Fig. 2 shows some equations of the linear models.

```
LM num: 1
duration =
   -0.001 * scraping
   - 0.0007 * sandblasting
   + 0.001 * washing
   + 0.082 * zinc
   + 0.0007 * ut
   + 0.0002 * plate
   + 0.7393 * propeller=E,G,H
   + 8.5688

LM num: 2
duration =
   -0.0036 * scraping
   - 0.0004 * sandblasting
   - 0.0006 * washing
   + 0.0499 * zinc
   + 0.0007 * ut
   + 0.0002 * plate
   + 0.7393 * propeller=E,G,H
   + 14.6214

LM num: 3
duration =
   0.0008 * sandblasting
   - 0.0006 * washing
   + 0.0007 * ut
   + 0.0001 * plate
   + 0.7393 * propeller=E,G,H
   + 8.2669

LM num: 4
duration =
   -0.0006 * washing
   + 0.0007 * ut
   + 0.0002 * plate
   + 1.7194 * propeller=E,G,H
   + 13.2962
```

Fig. 2 Dry docking linier models

These models represent linear combination of variables. Each linear model has attributes with different sign. Positive sign shows that the dry docking duration increases as the work volume increases, whereas negative sign shows that the dry docking duration decreases as the work volume increases. It means that works with negative sign must be finished as soon as possible because there are other works that need more time to be finished.

The validation results in the following indicator numbers: (1) Coefficient of Correlation (CC) = 0.5804; MAE = 4.799; RMSE = 6.3664; RAE = 78.16%; RRSE = 80.80%. The CC value was good enough, because its value is above the median value. However, the resulted error values (MAE RMSE, RAE, RRSE) were relatively high. Therefore, the validation process should be done using different method that is k-NN method.

The following values are obtained based on the k-NN method: CC = 0.4639; MAE = 5.3009; RMSE = 7.2589; RAE = 86.334%; RRSE = 92.1362%. Comparing the measurement performance values between CART and k-NN, it can be seen that CC of CART is higher than that of the k-NN and its error values are lower than that of the k-NN. Hence, it can be concluded that CART classification model was good and can be used to estimate dry docking duration

IV. CONCLUSION

The results of this study show that there are three dry docking works that can be used to classify the dry docking duration; which are propeller, washing, and plate. Based on this classification, there are 4 classes of dry docking duration with different linear model for each class.

The validation process for this model shows relatively high error values. This might be due to the data used to develop the model is too varied. However, the CART error values were lower than that of the k-NN method. Hence, it can be concluded that the developed CART model is good and can be used to estimate the dry docking duration.

ACKNOWLEDGMENT

The authors would like to express their sincere gratitude to University of Indonesia for financing this research, and also to Dr. Srdoc for her help in analyzing the attributes for dry docking works.

REFERENCES

[1] A. Srdoc, I. Bratko,and A. Sluga,"Machine learning applied to quality management - a study in ship repair domain,"*Computers in Industry*vol. 58, pp. 464-473, 2007.

[2] D.J. Hand, H. Mannila, and P. Smyth, *Principles of data mining*. Cambridge, MA: MIT Press, 2001.

[3] D.T. Larose, *Discovering knowledge in data: an introduction to data mining*. New Jersey: John Wiley & Sons, Inc., 2005.

[4] Shih-Yen Lin, Jo-Ting Wei, Chih-ChienWeng, and Hsin-Hung Wu, "A case study of using classification and regression tree and LRFM model in pediatric dental clinic," *International Conference on Innovation, Management and Service*,IPEDR vol.14. Singapore: IACSIT Press, 2011.

[5] I. H. Witten and E. Frank, *Data mining: practical machine learning tools and techniques – 2ndEdition*, San Francisco: Elsevier, 2005.

[6] R.J. Lewis, "An introduction to classification and regression tree (CART) analysis," *Annual Meeting for Academic Emergency Medicine* in San Francisco, 2000.

[7] P. Cunningham and S.J. Delany, "K-nearest neighbor classifiers," *Technical Report UCD-CSI-2007-4*, 2007.

Isti Surjandari is Professor and Head of Statistics and Quality Engineering Laboratory in the Department of Industrial Engineering, Faculty of Engineering, University of Indonesia. She holds a bachelor degree in Industrial Engineering from University of Indonesia and a Ph.D. degree from the Ohio State University. Her areas of interest are industrial management, quality engineering and applied statistics. She has a vast experience in industrial and manufacturing systems and has published papers in national and international journals. She is a senior member of American Society for Quality (ASQ), and Indonesian Association of Industrial Engineering and Management (ISTMI). She is also Editorial Board for International Journal of Technology.

Riara Novita is a graduate student and research assistant in the Department of Industrial Engineering, Faculty of Engineering, University of Indonesia. She holds a bachelor degree in Industrial Engineering from University of Indonesia.