Application of Machine Learning Method to Predict Reliability in Lubricating Oil System Components

Nurvita Arumsari¹, Feby Agung Pamuji², Bianda Devi Puspita³
¹,³ Marine Engineering Department, Shipbuilding Institute of Polytechnic Surabaya, Surabaya, Indonesia
² Electrical Engineering Department, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia

E-mail: arum@ppns.ac.id, feby@ee.its.ac.id

Abstract. In predicting the reliability and failure of components, classical methods are often used by determining the distribution between failure times. Sometimes, the determination of this distribution does not always match the data pattern that is owned because of the limited data records and the many types of distribution that must be chosen. In addition, how much influence the time series has on components cannot be clearly analyzed. Therefore, in this study a prediction will be carried out by combining classical methods with machine learning, namely Support Vector Regression (SVR) and Least Square Support Vector Regression (LSSVR). Both methods are considered capable of improving the prediction accuracy of a series of data. The results showed that the classical combined method with LSSVR had better accuracy than SVR.

1. Introduction
In carrying out component maintenance or maintenance on the lubricating oil system or in the main engine, it is necessary to predict the appropriate time for maintenance or component replacement beforehand. The time in question is the failure time between components. Reliability is an opportunity for the component or system to fulfill a predetermined task without failing for a certain period of time if it is operated properly in a certain environment[1].

The reliability value pattern based on the failure time between components will be adjusted to the distribution of the data which will then be estimated using machine learning (ML). According to research conducted by [2], the ML method is considered to be more effective and accurate in estimating a continuous data. In machine learning data is needed as training before issuing output then the training results will be tested with the same or contradictory data. There are two main categories in machine learning, namely classification and prediction. There are many forecasting techniques that are applied in various studies. For example, to predict the life span of a component, the classical time series method is often used. However, the method is largely limited by the classification of failure types so that it is difficult to validate all the assumptions that exist [3]

Machine learning is an application of computer science, specifically Artificial Intelligence (AI), which provides automatic system performance without being explicitly programmed. The learning process begins with data observation or searching for data patterns and making better decisions. Machine learning algorithms have been around for a long time, with the ability to automatically apply complex mathematical calculations to big data and faster. Broadly speaking, machine learning techniques or ways
of working are divided into 4 types, namely supervised learning, unsupervised learning, semi supervised learning, and reinforcement learning.

The application of machine learning in this study focuses on supervised learning by comparing two methods between Support Vector Regression (SVR) and Least Square Support Vector Regression (LS-SVR). The purpose of using both methods is to predict the reliability value of lubricating oil system components. The two methods are then hybridized by obtaining the results of the distribution of the failure data parameters generated through the software. The research object was selected from the lubricating oil system which has the most frequent interval between failure times in the last 5 years, namely Lubrication Oil (LO) Cooler and Strainer.

Both ML methods are the development of the Support Vector Machine (SVM) for numerical data types in the regression case. In the SVR algorithm, the response variable has a range in the form of real number values with the limitation function in the form of inequality [4]. Meanwhile, LS-SVR is a modification of SVR for numerical data with a boundary function in the form of an equation.

1.1 Lubricating Oil System

The following is a figure of the Lubricating Oil System and how it works

![Lubricating Oil System With Strainer and LO Cooler Component](image)

Figure 1. Lubricating Oil System With Strainer and LO Cooler Component

The lubrication system used is a wet sump with the type of main with the engine type is niigata. The principle of work start from the process of supplying lubricating oil from the storage tank to the sump tank with the help of a gear type hand pump. Before the lubricating oil flows into the main diesel engine, the lubricating oil will pass through the strainer in the sump tank for the filtering process, and then it will be pumped with a lube oil pump to the lube oil cooler. The temperature of the lubricating oil coming out of the cooler is automatically controlled at a constant level determined to obtain the desired viscosity of the main diesel engine. Then the lubricating oil will be divided into two lines, some of the lubricating oil will be flowed to the top of the main diesel to lubricate the components on the head such as the rocker arm, camshaft, etc. At the same time, some of the other lubricating oil will flow to the lube oil filter for the filtering process for the separation of lubricating oil. With water and unwanted substances. Furthermore, the lubricating oil will go to the main engine bearing and also flow back to the sump tank.
1.2 Support Vector Regression (SVR)

The SVR method aims to minimize risk so as to be able to overcome cases that are overfitting. The general function of the SVR is as follows [5].

\[ f(X) = w^T \varphi(X) + b \]  

where \( w \) is the weight vector, \( \varphi(X) \) is a function that maps \( x \) to a dimension, and \( b \) represents the error value. Meanwhile, to minimize the risk the following functions are used.

\[ R_{emp}(f) = \frac{1}{N} \sum_{i=1}^{N} \varnothing_\varepsilon (Y_i, f(X_i) + b) \]  

Turns into an equation

\[ R_{emp}(f) = \frac{1}{N} \sum_{i=1}^{N} \varnothing_\varepsilon (Y_i, W^T \varphi(X_i) + b) \]  

Where \( \varnothing_\varepsilon (Y, f(X)) \) is the \( \varepsilon \)-insensitive loss function which is defined as a function

\[ \begin{cases} |f(X) - Y| - \varepsilon, & \text{jika}|f(X) - Y| \geq \varepsilon \\ 0, & \text{untuk yang lain} \end{cases} \]  

The focus of SVR is to find the optimum hyperplane and minimize errors between training data and \( \varepsilon \)-insensitive loss as shown in the following equation.

\[ \min_{w,b,\xi^*,\xi} R_\varepsilon(W, \xi^*, \xi) = \frac{1}{2} W^T W + C \sum_{i=1}^{N} (\xi^*_i + \xi_i) \]  

With limitations

\[ \begin{align*} 
Y_i - W^T \varphi(X_i) - b & \leq \varepsilon + \xi^*_i, \quad i = 1, 2, \ldots, N \\
-Y_i + W^T \varphi(X_i) - b & \leq \varepsilon + \xi_i, \quad i = 1, 2, \ldots, N \\
\xi^*_i & \geq 0 \\
\xi_i & \geq 0 
\end{align*} \]

\( C \) is defined as a parameter with a maximum tolerance limit. Whereas \( W \) is a parameter vector with equations

\[ W = \sum_{i=1}^{N} (\beta^*_i - \beta_i) \varphi(X_i) \]  

By using the quadratic and lagrange program solving function, the following SVR final model is obtained.

\[ f(X) = \sum_{i=1}^{N} (\beta^*_i - \beta_i) K(X_i, X) + b \]  

With limitations

\[ \sum_{i=1}^{N} (\beta^*_i - \beta_i) = 0; 0 \leq \beta_i \leq C; 0 \leq \beta_i \leq C \]

And \( K(X_i, X) \) is a kernel function.
1.3 Least Square Support Vector Regression (LS-SVR)

The LS-SVR method was first developed by [6]. LS-SVR is a modification of the SVR method with the same goal of minimizing risk. This LS-SVR limitation function is in the form of an equation so it is predicted that it will produce a unique and more efficient solution. The objective function of the LS-SVR is shown the same as the SVR as follows [7].

\[ f(X) = w^T \varphi(X) + b \]

Based on the principle of minimizing risk, the optimum function and limits of the LS-SVR are as follows.

\[
\min J(w, e) = \frac{1}{2} w^T w + \frac{1}{2} c \sum_{i=1}^{N} e_i^2 \tag{8}
\]

\[ s. t \quad y_k = w^T \varphi(x_i) + b + e_k, \quad k = 1, 2, ..., N \tag{9} \]

Where \( J \) is the loss function and \( C \) is the maximum tolerance limit. With the lagrange equation, the optimum function is transformed as follows.

\[
L(w, b, e, \alpha) = J(w, e) - \sum_{k=1}^{N} \alpha_k \{w^T \varphi(x_k) + b + e_k - y_k\} \tag{10} 
\]

With \( \alpha_k \) are Lagrange multipliers.

In LS-SVR, the optimization problem can be simplified by solving linear equations rather than the more complex SVR. Thus, the final LS-SVR model is given as follows.

\[ y(x) = \sum_{k=1}^{N} \alpha_i K(X, X_k) + b \tag{11} \]

Where \( K(X_i, X) \) is the kernel function and \( \alpha_i \) is in vector form.

1.4 Time To Failure Distribution

To measure the Reliability \( R(t) \) of each components, first determine the type of distribution and its parameters. The following is the formula of the 2 parameter exponential and 2 parameter Weibull distribution [14].

The formula of \( R(t) \) 2 parameter exponential distribution

\[ R(t) = \exp \left[ -\frac{x - \gamma}{\theta} \right], \quad x > \gamma, \theta > 0, -\infty < \gamma < \infty \tag{12} \]

The formula of \( R(t) \) 2 parameter weibull distribution

\[ R(t) = \exp \left[ - \left( \frac{x - \gamma}{\alpha} \right)^{\beta} \right], \quad x > \gamma, \theta > 0, -\infty < \gamma < \infty \tag{13} \]

With
\( \theta \) is scale parameter
\( \gamma \) is threshold parameter
\( \alpha \) is scale parameter
\( \beta \) is shape parameter
1.5 Goodness of Fit Test Distribution

Determination of the distribution of Time to failure (TTF) components is based on the output Goodness of Fit value. In this case, the calculation of the value of Goodness of Fit uses the Anderson-Darling statistic [8].

Let \( Z = F(X) \), where \( F(X) \) is the cumulative distribution function (CDF). Suppose that a sample \( X_1, X_2, \ldots, X_n \) gives values \( Z(i) = F(X_i), \ i = 1, \ldots, n \).

Rearrange \( Z(i) \) in ascending order \( Z(1) < Z(2) < \cdots < Z(n) \).

Then, the Anderson-Darling Statistic (\( A^2 \)) is calculated as follows.

\[
A^2 = -n - \left( \frac{1}{n} \right) \sum_i \left[ (2i - 1) \log Z(i) + (2n + 1 - 2i) \log (1 - Z(i)) \right]
\] (14)

Besides Anderson-Darling statistics, goodness of fit test distribution is also determined by p-value exact.

1.6 Evaluation of Predicted Results

In evaluating the accuracy of the prediction results of reliability, two types of errors are used, namely MAE (Mean Absolute Error) and RMSE (Root Mean Square Error) [9] which are stated as follows.

\[
MAE = \frac{\sum_{n=1}^{N} |\hat{r}_n - r_n|}{N}
\] (15)

\[
RMSE = \sqrt{\frac{\sum_{n=1}^{N} (\hat{r}_n - r_n)^2}{N}}
\] (16)

Where \( \hat{r}_n \) is the predicted value, \( r_n \) is the observed value and \( N \) is the total predicted data.

1.7 The Discussion about Reliability Prediction Research

Basically, many researches on reliability forecasting have been done. One of the studies on reliability was carried out by [10] who predicted the reliability concrete of components using conventional methods through determining the Weibull parameter distribution based on experimental data. Further research that combines reliability with machine learning methods, among others, was conducted by [11] and [12].

The two researchers used ridge regression and support vector regression methods in predicting the component reliability values. In ridge regression, the predicted data pattern has a linear pattern, while in SVR it can be used to predict non-linear data patterns because it involves kernel functions. In this study, the degenerate data in the SVR method was failure time data so that the results obtained were only nonparametric values without knowing the appropriate distribution of failure time data. This study tries to combine the results obtained from conventional methods and machine learning, both SVR and LSSVR. Research that has been conducted by [2] and [13] has proven that the SVR method is better at estimating non-linear patterns with the RBF kernel model. Meanwhile, according to [7] the LSSVR method is considered to have a more accurate prediction result.
2. Research Methodology

The data used in this study are data on the failure time of the LO Cooler and strainer components in the lubricating oil system with a numerical scale. Data on component failure times are in the form of data series for the last 5 years from 2013 to 2018. As previously explained, the proposed method in this study is to combine conventional methods with machine learning.

![Research Flowchart](image)

3. Result and Discussion

In applying the machine learning method to the LO Cooler and Strainer components, the parameter distribution will first be determined. The parameter distribution is determined based on the failure time data record between components. The results are stated in Table 1 below.

| Component | Distribution | Shape  | Scale   | Threshold |
|-----------|--------------|--------|---------|-----------|
| LO Cooler | 2 Parameter  | Eksponential | 2057, 08 | 3558,87   |
|           |              |        |         |           |
| Strainer  | 2 Parameter  | Weibull | 10,92   | 3503,4    |

From the distribution parameter results obtained in Table 1, then the reliability prediction for each component with formula 12 and 13 are carried out at certain time intervals.
Based on Figures 3 and 4, it can be seen that the reliability pattern of LO Cooler tends to sharpen after $t = 3000$ hours, while in the strainer after $t = 2880$ hours. This shows that the strainer component fails more frequently than the LO Cooler. In the Strainer, after a time interval of 3650 hours the component has a reliability value of less than 0.2 and ideally component replacement should be carried out. Meanwhile, if you want to treat the strainers, it should be done at intervals of every 3290 hours or about 4.5 months. This is because after $t$ the reliability value of the strainer is less than 0.6.

Meanwhile, the LO Cooler has a less frequent failure than the strainer. In LO Cooler, component replacement should be done after $t = 6040$ hours because it has a reliability below 0.2. Meanwhile, if you want to do maintenance on the LO cooler, it should be done at intervals of 4500 hours or about 6 - 6.5 months. This is because, after an interval of 4500 hours the LO cooler has a reliability of less than 0.6.

From the prediction results of reliability with conventional methods, it can be seen that the pattern formed is non-linear because it follows an exponential and weibull pattern. Thus, the next recommended method is a method that is able to overcome the nonlinearity problem so that it can produce more accurate predictions. In this study, the two recommended methods for hybridization with conventional methods are SVR and LSSVR.

After obtaining the predictive value of classical reliability, namely by adjusting the distribution patterns and parameters obtained, then it will be hybridized with the SVR and LSSVR machine learning methods. Before prediction with the two methods, first the data reliability will be displaced into training and testing data. In this study, the percentage of training data selection was 90% and testing was 10%.
The kernel function was selected based on the results of previous research, namely the RBF (Radian Basis Function) Kernel. The selection of the RBF kernel function is considered the most appropriate in estimating non-linear patterns. The prediction results of reliability can be seen in Table 2 below.

| t | Actual | Hybrid SVR | Hybrid LSSVR |
|---|--------|------------|--------------|
|   | Cooler | Strainer | Cooler | Strainer | Cooler | Strainer |
| 1 | 0.3401 | 0.6044 | 0.3371 | 0.6103 | 0.3400 | 0.6045 |
| 2 | 0.3348 | 0.5943 | 0.3322 | 0.6107 | 0.3347 | 0.5942 |
| 3 | 0.3295 | 0.5839 | 0.3274 | 0.5932 | 0.3294 | 0.584 |
| 4 | 0.3243 | 0.5735 | 0.3226 | 0.585 | 0.3242 | 0.574 |
| 5 | 0.319 | 0.563 | 0.318 | 0.5769 | 0.3191 | 0.5631 |
| 6 | 0.3139 | 0.5523 | 0.3134 | 0.569 | 0.3138 | 0.552 |
| 7 | 0.3087 | 0.5416 | 0.3089 | 0.5613 | 0.3088 | 0.5415 |
| 8 | 0.3036 | 0.5307 | 0.3045 | 0.5539 | 0.3035 | 0.5306 |
| 9 | 0.2985 | 0.5197 | 0.3002 | 0.5468 | 0.2984 | 0.5198 |
| 10 | 0.2934 | 0.5087 | 0.296 | 0.5401 | 0.2935 | 0.5086 |

**Figure 5.** Reliability Prediction of Strainer Component with Hybrid SVR

**Figure 6.** Reliability Prediction of LO Cooler Component with Hybrid SVR
Based on the results of the analysis of the reliability prediction with hybrid SVR for the Strainer and LO Cooler components in the lubricating oil system, it is known that the predictive value obtained by using training data is almost close to the true value. In this case, the reliability value is actually the result of conventional prediction by generating distribution. Based on Figures 5 and 6, it can be seen that the prediction results follow the reliability distribution graph pattern. In the strainer component, the reliability with the observation index above 250 has a fairly sharp pattern of decline, while in the LO Cooler, a significant decrease in the predicted number occurs after the 175th observation index. In this case, the observation index does not represent the actual time. The time interval given can differ according to the needs of the interpretation of the researcher. A sharp decrease in the number of reliability indicates that the use of components is predicted to exceed the lifespan. The prediction results of the two components of the lubricating oil system are then compared with the LSSVR hybrid method as shown in Figures 7 and 8 below.

![Figure 7. Reliability Prediction of Strainer Component with Hybrid LSSVR](image7.png)

![Figure 8. Reliability Prediction of LO Cooler Component with Hybrid LSSVR](image8.png)

The results of descriptive analysis through graphic observations for the two components of the lubricating oil system show that the predictive value of the reliability obtained closely coincides with the actual predicted value. The axes show the time intervals in hours. Descriptively, when compared with the predicted numbers using the hybrid SVR method, it can be seen that the hybrid LSSVR method tends to be better because it is very similar to the original pattern of the actual prediction results.
Finally, the following of contrast error between actual and prediction result with Hybrid SVR and LSSVR can be see in Figure 9 and 10.

![Figure 9. The Contrast Error of Reliability Prediction LO Cooler](image)

![Figure 10. The Contrast Error of Reliability Prediction Strainer](image)

To ensure a descriptive description of the results of the analysis of the two recommended methods, model validation was carried out by calculating the selected errors, namely MAE and RMSE, as shown in Table 3. The calculation of the error value comes from the prediction results with training data which is then compared with the testing data. Overall validation results show that the hybrid LSSVR method is better than the hybrid SVR. This is shown in the error acquisition of the two components using both MAE and RMSE which is smaller than the hybrid SVR. Meanwhile, when compared between components using both methods, it can be seen that the LO cooler has a smaller error value compared to the Strainer.
Table 3. Error Validation With Hybrid SVR dan LSSVR

| Component | Hybrid SVR | Hybrid LSSVR |
|-----------|------------|--------------|
|           | LO Cooler  | Strainer     | LO Cooler  | Strainer     |
| MAE       | 0.00948    | 0.1005       | 0.000019   | 0.000011     |
| RMSE      | 0.0022     | 0.1254       | 0.00000197 | 0.0000019    |

4. Conclusion

In this study, the application of predictive method selection is a method commonly used for nonlinear cases. In this case, the SVR and LSSVR methods were chosen because in addition to being compatible with the nonlinear case, the two methods were considered to be able to improve the accuracy of the component prediction values. The choice of kernel function in this study is based on the results of previous studies which correspond to the nonlinear case, namely the RBF kernel. The contribution given to this paper is in the form of combining classic reliability prediction methods with machine learning. The results obtained, the combined method can improve the prediction results of component reliability, especially in LSSVR. For further research, we can apply the Hybrid LSSVR and SVR method to conduct predictive maintenance in LO Cooler component.

Acknowledgements

The author would like to thank those who cannot be named one by one (anonymous) who have provided various suggestions and references for a significant improvement in paper quality.

References

[1] H. Al Haiany 2016 “Reliability Centered Maintenance Different Implementation Approaches,” Eng. Maint.
[2] E. K. Adhitya, R. Satria, and H. Subagyo 2015 “Komparasi Metode Machine Learning dan Metode Non Machine Learning untuk Estimasi Usaha Perangkat Lunak,” J. Softw. Eng., vol. 1, no. 2, pp. 109–113.
[3] Y. H. Liang 2011 “Analyzing and forecasting the reliability for repairable systems using the time series decomposition method,” Int. J. Qual. Reliab. Manag.
[4] A. J. Smola and B. Schölkopf 2004 “A Tutorial on Support Vector Regression. Statistics and Computing archive,” Stat. Comput.
[5] G. F. Fan, L. L. Peng, W. C. Hong, and F. Sun 2016 “Electric load forecasting by the SVR model with differential empirical mode decomposition and auto regression,” Neurocomputing.
[6] J. A. K. Suykens, T. Van Gestel, J. De Brabanter, B. De Moor, and J. Vandewalle 2002 “Basic Methods of Least Squares Support Vector Machines,” in Least Squares Support Vector Machines.
[7] G. Y. -., Z. L. -., L. X. -., R. C. -., and M. J. 2013 “A New Time Series Prediction Method Based on LS-SVR and AGO,” Int. J. Adv. Comput. Technol., vol. 5, no. 6, pp. 542–551.
[8] L. N. Harris 1987 “Goodness-of-fit techniques, Ralph B. D’Agostino and Michael A. Stephens (eds). Marcel Dekker Inc. 1986. No. of pages: 560. Price: $79.75 (U.S.A. and Canada); $95.50 (elsewhere),” Qual. Reliab. Eng. Int.
[9] W. Wang and Y. Lu 2018 “Analysis of the Mean Absolute Error (MAE) and the Root Mean Square Error (RMSE) in Assessing Rounding Model,” in IOP Conference Series: Materials Science and Engineering.
[10] Q. S. Li, J. Q. Fang, D. K. Liu, and J. Tang 2003 “Failure probability prediction of concrete components,” Cem. Concr. Res., vol. 33, no. 10, pp. 1631–1636.

[11] M. D. C. Moura, E. Zio, I. D. Lins, and E. Droguett 2011 “Failure and reliability prediction by support vector machines regression of time series data,” Reliab. Eng. Syst. Saf., vol. 96, no. 11, pp. 1527–1534.

[12] S. Mangalathu and J. S. Jeon 2018 “Classification of failure mode and prediction of shear strength for reinforced concrete beam-column joints using machine learning techniques,” Eng. Struct., vol. 160, no. September 2017, pp. 85–94.

[13] N. Arumsari and F. A. Pamuji 2017 “Peramalan Irradiance Cahaya Matahari pada Sel Surya untuk Memenuhi Kebutuhan Energi Listrik dengan Metode Support Vector Regression (SVR),” J. Nas. Tek. Elektro, vol. 6, no. 1, p. 18.

[14] https://support.minitab.com/en-us/minitab/18/help-and-how-to/quality-and-process-improvement/quality-tools/how-to/individual-distribution-identification/methods-and-formulas/distributions/. 35 L74