International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 8, Issue 4, pp 51-56, April-2024

https://doi.org/10.47001/IRJIET/2024.804006

Modeling of Capacity Factor in Rembang Coal-Fired Steam Power Plant Using Regression Modeling

[1]*Ery Perdana, [2]Sulardjaka, [3]Budi Warsito

[1]Master of Energy, School of Postgraduate, Diponegoro University, Semarang, Indonesia
[2]Department of Mechanical Engineering, Faculty of Engineering, Diponegoro University, Semarang, Indonesia
[3]Department of Statistics, Faculty of Science and Mathematics, Diponegoro University, Semarang, Indonesia
*Corresponding Author's E-mail: eryperdana2005@gmail.com

Abstract **- Coal-Fired Steam Power Plant (PLTU) Rembang is an important power plant in the Central Java electricity system. Like other coal-fired steam power plants, fuel cost is the most significant expense when operating the PLTU Rembang. During the 2019-2021 period, the average fuel cost was 73.88% of total costs. One of the ways to reduce fuel costs is by improving the accuracy of fuel demand planning. Fuel procurement planning is very dependent on the projected amount of electricity sales from power plant, which is largely determined by the power plant's Capacity Factor (CF). However, PLTU Rembang does not have any CF prediction modeling. This research developed and compared four prediction models: random forest regression, support vector regression, multiple polynomial regression, and multiple linear regression. Based on the comparison of validation from the four prediction model with MAPE and R-squared parameters, the multiple linear regression models is the best model, with the lowest MAPE of 7.83% and the highest R-squared of 0.8814. This multiple linear regression model can be used to predict the CF of PLTU Rembang in the future so that fuel demand planning is more accurate.**

Keywords: capacity factor, regression, multiple linear regression, MAPE, R-squared.

I. INTRODUCTION

Coal-fired steam power plant (PLTU) Rembang is a crucial power plant in the Central Java electricity subsystem. It maintains voltage stability and provides active power sources in the eastern area of Central Java. Before the operation of PLTU Rembang, the Northern Coastal Area of Java, spread along Central Java Province from Kudus – Jekulo – Pati – Rembang – Blora - Cepu, constituted a 'weak area' in the Java-Bali electricity system due to rare active power sources and relative voltage fluctuation between normal load periods and peak load periods [1].

PLTU Rembang utilizes both medium-rank coal and low-rank coal as fuel. During the period of 2019-2021, the fuel cost expense in PLTU Rembang was notably high, as depicted in Figure 1.

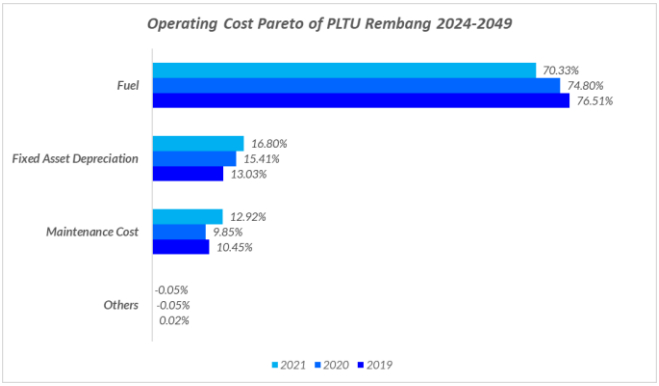

Figure 1: Pareto of Operating Cost 2019-2022

According to the Financial Report of PLTU Rembang from 2019-2021, shown in figure 1, fuel cost in 2021was Rp. 1,456 billion, or about 70.33% of total cost. During 2019-2021, average fuel cost was about 73,88% of total cost [2]. This average fuel cost is still considered very high. A small saving in fuel demand planning will significantly reduce PLTU Rembang's operational costs.

The coal purchasing plan is very dependent on the projected electricity sales. The projection of electricity sales is determined by the power plant's Capacity Factor (CF) prediction. CF is a ratio between the actual and maximum electricity produced in a certain period [3]. Currently, coal demand planning relies on annual operation planning (ROT) published by the load dispatcher, PLN P2B, for annual planning, and previous CF for monthly coal demand planning.

During 2016-2021, the CF gap between PLN's Annual Operation Plan (ROT) and actual CF is quite large. In 2017-2022, CF gap was -18,22%, -4,99%, -34,48%, 1,39%, and 14,3% respectively. A negative CF gap condition means that it needs additional coal demand. Additional fuel costs that had not been budgeted for at the beginning of the year need to be

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 8, Issue 4, pp 51-56, April-2024

https://doi.org/10.47001/IRJIET/2024.804006

allocated. A positive CF gap condition means that electricity sales decrease by less than the initial sales plan. It will have a significant impact on the company's financial performance. One of the causes of these CF gaps is that PLTU Rembang does not have accurate CF predictions. Currently, PLTU Rembang only uses the CF ROT plan and CF Monthly Operation Plan (ROB) in predicting CF.

In this research, four regression models will be developed for CF prediction, namely random forest regression, support vector regression, multiple linear regression, and multiple polynomial regression. This research aims to obtain the best prediction model by comparing the four CF prediction models based on the Mean Absolute Percentage Error (MAPE) and R-squared criteria.

II. MATERIAL AND METHOD

This research utilizes both primary and secondary data. Primary data are obtained from field observations and field data inventory, specifically at the Rembang Coal-Fired Steam Power Plant (PLTU Rembang). Primary data are sourced from the Business Production Report of PLTU Rembang. These data consist of CF (Capacity Factor), SOF (Scheduled Outage Factor), and EFOR (Equivalent Forced Outage Rate) of PLTU Rembang. SOF (Scheduled Outage Factor) is the ratio of the total planned outage and maintenance outage hours of the generating unit to the total hours in a given period, while EFOR is the ratio of the total unplanned outage hours (forced outage) to the total hours in a given period[4].

Secondary data consist of the Java-Bali load, Central Java load, Net Capability (DMN) of Java-Bali, Net Capability (DMN)of Central Java, sourced from PLN P2B load monitoring website, and merit order data and Incremental Fuel Cost (IFC) sourced from monthly releases by PLN P2B. DMN represents the net capability of power plants according to the power purchase agreement between the generating companies and PLN and may be revised with a declaration letter from the Regional Operations Division of PLN [4]. Merit order is the simplest method in economic dispatch, ordering power plants from the lowest to the highest operating costs [5]. IFC (Incremental Fuel Cost) is defined as the change in fuel cost resulting from changes in generated power output [6].

The process of developing the CF prediction model consists of several stages: data preprocessing, selection of significant variables, model development, and model validation. These process stages are carried out using the Python programming language. This research utilizes the Python programming language because its simple syntax yields exceptional results [7]. The flow of process stages is illustrated in Figure 2.

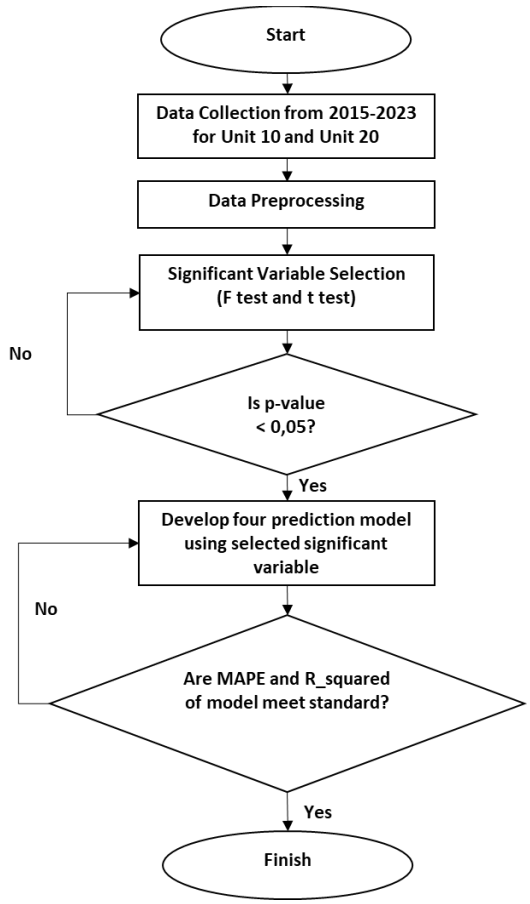

Figure 2: Research Flowchart

Based on Figure 2, the research begins by collecting data from 2015 to 2023. This data is presented in Table 1.

Table 1: Data for CF Prediction Modeling

Month	Merit Order	IFC (Rp/ kWh)	EFOR (%)	SOF (%)	Java-Bali Load (MW)	Central Java Load (MW)	Java-Bali DMN (MW)	Central Java DMN (MW)	CF (%)
Jan 2015	2	305	2.72	0	22375	3608	31189	5193	94.24
Feb 2015	2	293	4.83	0	21755	3570	31189	5193	85.2
Mar 2015	3	284	0.51	0	22356	3730	31189	5193	93.31
Apr 2015	1	270	1.56	0	22953	3730	31166	5193	95.41
…	…	…	…	…	…	…	…	…	…
Sep 2023	20	463	0.76	0	29975	4996	43349	11094	90.58
Oct 2023	13	422	1.08	0	31082	5203	43351	11096	95.31
Nov 2023	11	434	1.14	0	31515	5248	43489	11096	94.75
Dec 2023	17	458	0.23	32.03	30905	5086	44368.7	11096	64.01

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 8, Issue 4, pp 51-56, April-2024

https://doi.org/10.47001/IRJIET/2024.804006

Data is further subjected to initial preprocessing to detect any missing data. For missing data, imputation is carried out using linear interpolation method. The next process involves selecting significant variables using hypothesis testing in the form of F-test and t-test [8]. The t-test process is performed using the backward elimination method. Significant variables are those with p-value < 0.05 based on the t-test results [9]. Using these selected variables, prediction modeling is conducted. There are four models developed, namely random forest regression, support vector regression, multiple polynomial regression, and multiple linear regression. These four regression models are among the seven most commonly used methods[10].

Validation of the model is conducted using the MAPE (Mean Absolute Percentage Error) and R-squared parameters. The best model is the one with the smallest MAPE and the largest R-squared. MAPE is the average absolute difference between the predicted and actual values, expressed as a percentage of the actual value. MAPE can be calculated using equation (1)[11].

$$MAPE = \frac{100}{n}\sum\frac{|A_t - F_t|}{A_t} \quad (1)$$

Where

A_t = actual value at time t

F_t = forecasted value at time t

n = number of data points

R-squared, or the coefficient of determination, is a statistical measure indicating the influence exerted by the independent variable (X) on the dependent variable (Y). The equation (2) for the coefficient of determination is as follows [12].

$$R^2 = 1 - \frac{RSS}{\sum(y_i - \bar{y})^2} = 1 - \frac{\sum(y_i - f(x_i))^2}{\sum(y_i - \bar{y})^2} \quad (2)$$

Where

RSS = sum of squared residual

y_i = actual value at *i*-th observation

y = mean of actual values

$f(x_i)$ = predicted value at *i*-th observation

III. RESULTS AND DISCUSSIONS

The results of this research include the selection of significant independent variables and the development and validation of the model.

3.1 Significant Variables Selection

Hypothesis testing is conducted to obtain significant variables influencing CF, including the F-test and the t-test. The first hypothesis testing performed is the F-test. The F-test is a hypothesis test aimed at determining whether there is a relationship between the independent and dependent variables [8]. This test has two hypotheses: H0 (null hypothesis) and Ha (alternative hypothesis). H0 states that there is no relationship between the independent variables and the dependent variable, while Ha states that at least one independent variable affects the dependent variable[9].

The F-test is conducted by gathering all data of independent and dependent variables and forming a regression equation. The result of the F-test is a p-value with the following interpretations:

a) If the p-value < 0.05, H0 is rejected. It means that there is at least one independent variable that influences the dependent variable.
b) If the p-value ≥ 0.05, H0 is accepted. This means that the independent variables do not influence the dependent variable.

In this F-test, data from all variables suspected to influence CF are collected. These data include merit order, IFC, EFOR, SOF, Java-Bali DMN, Central Java DMN, Java Bali load, and Central Java load. The results of the F-test are obtained by running linear regression using the Python programming language, as shown in Figure 3.

OLS Regression Results

Dep. Variable:	CF_PLTU	R-squared:	0.908
Model:	OLS	Adj. R-squared:	0.900
Method:	Least Squares	F-statistic:	122.0
Date:	Sun, 17 Mar 2024	Prob (F-statistic):	9.33e-48
Time:	11:52:35	Log-Likelihood:	-341.15
No. Observations:	108	AIC:	700.3
Df Residuals:	99	BIC:	724.4
Df Model:	8		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	72.6242	17.772	4.087	0.000	37.362	107.887
Peringkat Merit	-0.7252	0.219	-3.310	0.001	-1.160	-0.290
Fuel_Cost	0.0460	0.024	1.880	0.063	-0.003	0.094
EFOR	-0.9067	0.070	-12.868	0.000	-1.046	-0.767
SOF	-0.8637	0.032	-26.713	0.000	-0.928	-0.800
Beban_Jamali	0.0018	0.001	1.208	0.230	-0.001	0.005
Beban_Jateng	0.0062	0.011	0.573	0.568	-0.015	0.028
DMN_Jamali	-0.0017	0.001	-1.592	0.115	-0.004	0.000
DMN_Jateng	-0.0001	0.001	-0.083	0.934	-0.003	0.003

Omnibus:	24.836	Durbin-Watson:	1.801
Prob(Omnibus):	0.000	Jarque-Bera (JB):	39.629
Skew:	-1.033	Prob(JB):	2.48e-09
Kurtosis:	5.130	Cond. No.	1.41e+06

Figure 3: Result of F-test obtained from the linear regression output

Based on Figure 3, the probability (F-statistic), the p-value, is obtained to be 9.33 x 10^{-48}. This value is significantly below the p-value threshold of 0.05, thus rejecting the null hypothesis (H_0). It means that at least one significant variable influences CF.

The second hypothesis test after the F-test is the t-test. The t-test aims to determine whether there is a partial (individual) effect of each independent variable on the dependent variable[13]. The t-test is combined with the backward elimination technique in selecting influential

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 8, Issue 4, pp 51-56, April-2024

https://doi.org/10.47001/IRJIET/2024.804006

independent variables. The t-test results yield the p-value for each independent variable. The independent variable with the highest p-value and its p-value is more than 0.05 is eliminated. Subsequently, another t-test is conducted on the remaining set of independent variables, and elimination is repeated for the independent variable with the highest p-value and more than 0.05. This process continues until all independent variables have p-values less than 0.05 [9]. The results of the t-test combined with the backward elimination technique are shown in Table 2.

Table 2: Results of the backward elimination process

Independent Variable	*p-value*			
	Elimination Stage 1	Elimination Stage 2	Elimination Stage3	Elimination Stage4
Merit Order	0,0013	0,0011	0,0011	0,0014
IFC	0,0631	0,0598	0.0821	
EFOR	$7{,}41x10^{-23}$	$4{,}32x10^{-23}$	$3{,}73x10^{-23}$	$9{,}43x10^{-23}$
SOF	$4{,}74x10^{-47}$	$1{,}65x10^{-47}$	$7{,}9x10^{-48}$	$5{,}58x10^{-48}$
Java Bali Load	0,2299	0,1667	$8{,}59x10^{-5}$	$1{,}02x10^{-6}$
Central Java Load	0,5679	0,4009		
Java Bali DMN	0,1146	0,0004	$6{,}45x10^{-5}$	$3{,}16x10^{-5}$
Central Java DMN	0,9337			

Based on Table 2, there were three rounds of elimination of independent variables until all variables had p-values below 0.05. In the first iteration, the Central Java DMN variable, which had a p-value of 0.9337, was eliminated. The Central Java load variable, which had a p-value of 0.4009, was eliminated in the second iteration. The IFC variable, which had a p-value of 0.0821, was eliminated in the third iteration. In the fourth iteration, no variable was eliminated, so the significant variables influencing the CF of the PLTU Rembang are merit order, EFOR, SOF, Java-Bali DMN, and Java-Bali load.

3.2 Model Development and Validation

In order to obtain the best CF prediction model, four regression models were developed: random forest regression, support vector regression, multiple polynomial regression, and multiple linear regression. Before running the models, the data is divided into two sets: 80% for training data and 20% for testing data. This composition of splitting is one of the commonly used compositions[14]. The data included in the model are the significant independent variables and dependent variables.

3.2.1 Regression model development

a) Random Forest Regression

Random forest is a machine learning algorithm that falls under supervised learning techniques, which combines the output of multiple decision trees to achieve a result. Random forest can be used for regression and classification tasks. Random forest regression consists of many trees that depend on random vectors, so predictor trees take numeric values instead of class labels [15]. In this research, the random forest regression model uses 100 estimators. The validation results are shown in Figure 4.

b) Support Vector Regression

Support Vector Regression (SVR) is an application of Support Vector Machines (SVM) in regression analysis[16]. The basic idea of SVM is to map the training data from the input space to a higher-dimensional feature space through a function and then construct a separating hyperplane with maximum margin in the feature space[17]. In this research, the SVR model is developed with a linear kernel. The validation results of the model can be seen in Figure 5.

c) Multiple Polynomial Regression

A regression can be called a polynomial regression if the relationship between the dependent and independent variables can be described by a curve [18]. Multiple polynomial regression is a polynomial regression with more than one independent variable. With more than one independent variable, the regression equation is a combination of predictor degrees and interactions between predictors [12]. In this research, multiple polynomial regression model is developed with a second order. The validation results of the model are shown in Figure 6.

d) Multiple Linear Regression

Multiple linear regression is essentially an extension of simple linear regression that involves more than one predictor variable [19]. The equation for multiple linear regression is essentially the same as the equation for simple linear regression, but with multiple independent variables[19]. The general equation for multiple linear regression is as equation (3)[12].

$$y=\beta_0+x_1+\beta_2x_2+\cdots+\beta_px_p \quad (3)$$

Where y is dependent variable. β_0 is *intercept*. β_1, β_2, β_p are independent variable coefficients. x_1, x_2, dan x_p are independent variables.

In this study, multiple linear regression model was developed using five independent variables. The validation results of the model are shown in Figure 7.

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 8, Issue 4, pp 51-56, April-2024

https://doi.org/10.47001/IRJIET/2024.804006

3.2.2 Model Validation

The validation of the models was conducted by calculating the MAPE and R-squared for each regression model. The comparison of the validation results of these four models can be seen in Figure 4 through Figure 7.

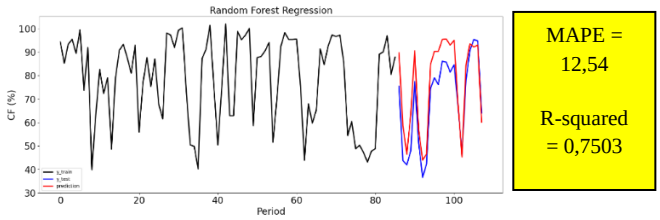

Figure 4: Random Forest Regression Model Validation

Figure 4 shows that the MAPE of the random forest regression model is 12.58, and the R-squared is 0.7488.

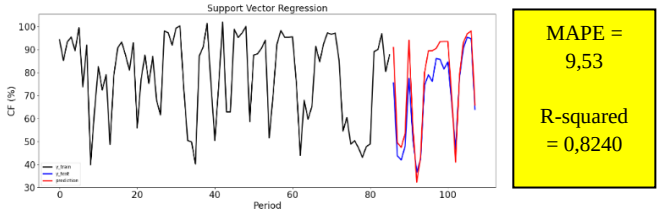

Figure 5: Support Vector Regression Model Validation

Figure 5 shows that the MAPE of the support vector regression model is 9.53, and the R-squared is 0.8240.

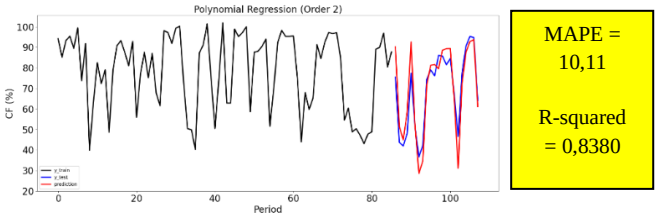

Figure 6: Multiple Polynomial Regression Model Validation

Figure 6 shows that the MAPE of the multiple polynomial regression model is 10.11, and the R-squared is 0.8380.

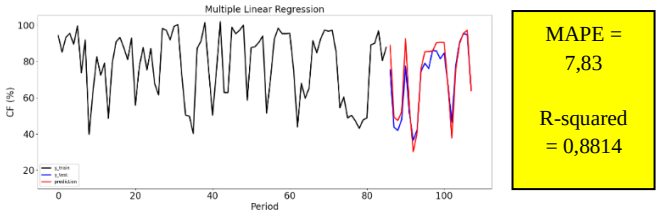

Figure 7: Multiple Linear Regression Model Validation

Figure 7 shows that the MAPE of the multiple linear regression model is 7.83, and the R-squared is 0.8824. A summary comparison of the MAPE and R-squared for all regression models is presented in Table 3.

Table 3: Comparison of MAPE dan R-squared from four regression models

Regression Model	MAPE	R-squared
Random Forest Regression	12,54	0,7503
Support Vector Regression	9,53	0,8240
Multiple Polynomial Regression	10,11	0,8380
Multiple Linear Regression	7,83	0,8814

Based on Table 3, the best regression model for predicting CF is the multiple linear regression model with a MAPE of 7.83 and an R-squared of 0.8814 because this model has the smallest MAPE and the largest R-squared among the four prediction models. The equation (4) for multiple linear regression that can be used to predict CF of PLTU Rembang is as follows.

$$y = 70{,}1714 - 0{,}2887x_1 - 0{,}9191x_2 - 0{,}8960\,x_3 + 0{,}0035x_4 - 0{,}0018x_5 \quad (4)$$

Where,

y = CF of PLTU Rembang
x_1 = Merit Order
x_2 = EFOR
x_3 = SOF
x_4 = Java-Bali Load
x_5 = Java-Bali DMN

This regression model is beneficial for accurate coal demand planning of PLTU Rembang. Accurate coal demand planning is expected to support the financial performance of the company.

IV. CONCLUSION

The prediction of CF is crucial in power plant operations as it helps estimate the fuel requirements of the power plant. In this research, four regression models were developed: random forest regression, support vector regression, multiple polynomial regression, and multiple linear regression, to predict the CF of PLTU Rembang. Significant variable selection was performed using hypothesis testing, including F-test and t-test, combined with the backward elimination method. By comparing the MAPE and R-squared of the four models, the multiple linear regression model emerged as the best model with the smallest MAPE of 7.83 and the largest R-squared of 0.8814. This regression model is beneficial for accurate coal demand planning of PLTU Rembang.

ACKNOWLEDGEMENT

The authors would like to acknowledge the management of PT PLN Nusantara Power UP Rembang for providing resources and facilities while conducting this research and PT PLN UIP2B Jawa-Bali for providing the data.

ISSN (online): 2581-3048
Volume 8, Issue 4, pp 51-56, April-2024
https://doi.org/10.47001/IRJIET/2024.804006

REFERENCES

[1] L. Rochman, *Feasibility Study PLTU-1 Jawa Tengah 2x (300-400) MW Rembang*. Jakarta: PT. Arkonin Engineering MP, 2017.

[2] PT PLN (Persero), "Financial Highlight AMC Jawa 2011-2021," Jakarta, 2022.

[3] Kementerian ESDM, "Permen ESDM No 10 Tahun 2017," 2017.

[4] PT PLN (Persero), "Protap Deklarasi Kondisi Pembangkit dan Indeks Kinerja Pembangkit." 2017.

[5] D. Mariani, Y. M. Safarudin, N. F. Aulia, A. H. Suudy, N. A. MS, and B. M. Hermawan, "Perhitungan Economic Dispatch Tiga Buah Pembangkit Menggunakan Metode Merit Order Dengan Mempertimbangkan Losses," *Eksergi J. Tek. Energi*, vol. 17, no. 3, pp. 221–232, 2021.

[6] Delima and Syafii, "Operasi Ekonomis dan Unit Commitment Pembangkit Thermal Pada Sistem Kelistrikan Jambi," *J. Nas. Tek. Elektro*, vol. 5, no. 3, 2016, doi: 10.20449/jnte.v5i3.331.

[7] M. F. Sanner, "Python : A Programming Language for Software Integration and Development," 1999.

[8] M. H. Kutner, C. J. Natchtseim, J. Neter, and W. Li, *Applied Linear Statistical Models*, 5th ed., vol. 29, no. 2. New York: McGraw-Hill, 2004. doi: 10.1080/00224065.1997.11979760.

[9] G. James, D. Witten, T. Hastie, R. Tibshirani, and J. Taylor, "An Introduction Statistical Machine Learning With Applications in Python," *Springer Texts Stat.*, pp. 425–472, 2023, [Online]. Available: https://www.statlearning.com/

[10] D. Polzer, "7 of the Most Used Regression Algorithms and How to Choose the Right One," *https://towardsdatascience.com/*, 2021. https://towardsdatascience.com/7-of-the-most-commonly-used-regression-algorithms-and-how-to-choose-the-right-one-fc3c8890f9e3 (accessed Oct. 24, 2022).

[11] B. Putro, M. T. Furqon, and S. H. Wijoyo, "Prediksi Jumlah Kebutuhan Pemakaian Air Menggunakan Metode Exponential Smoothing," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 2, no. 11, pp. 4679–4686, 2018.

[12] S. Weisberg, *Applied Linier Regression*, Fourth. Wiley, 2014.

[13] N. Sudjana, "Metode Statistika Edisi keenam," *Bandung PT. Tarsito*, 2005.

[14] S. Raschka, "Model Evaluation, Model Selection, and Algorithm Selection in Machine Learning," 2018.

[15] Z. Jin, J. Shang, Q. Zhu, C. Ling, W. Xie, and B. Qiang, "RFRSF: Employee Turnover Prediction Based on Random Forests and Survival Analysis," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 12343 LNCS, pp. 503–515, 2020, doi: 10.1007/978-3-030-62008-0_35.

[16] L. Bi, O. Tsimhoni, and Y. Liu, "Using the support vector regression approach to model human performance," *IEEE Trans. Syst. Man, Cybern. Part ASystems Humans*, vol. 41, no. 3, pp. 410–417, 2011, doi: 10.1109/TSMCA.2010.2078501.

[17] C. H. Wu, J. M. Ho, and D. T. Lee, "Travel-time prediction with support vector regression," *IEEE Trans. Intell. Transp. Syst.*, vol. 5, no. 4, pp. 276–281, 2004, doi: 10.1109/TITS.2004.837813.

[18] Y. G. Akhlaghi, X. Zhao, S. Shittu, J. Li, C. Science, and T. Engineering, "A statistical model for dew point air cooler based on the multiple polynomial regression approach," 2019.

[19] K. A. Marill, "Advanced Statistics: Linear Regression, Part II: Multiple Linear Regression," *Acad. Emerg. Med.*, vol. 11, no. 1, pp. 94–102, 2004, doi: 10.1197/j.aem.2003.09.006.

Citation of this Article:

Ery Perdana, Sulardjaka, Budi Warsito, "Modeling of Capacity Factor in Rembang Coal-Fired Steam Power Plant Using Regression Modeling" Published in *International Research Journal of Innovations in Engineering and Technology - IRJIET*, Volume 8, Issue 4, pp 51-56, April 2024. Article DOI https://doi.org/10.47001/IRJIET/2024.804006

****

