Prediction Modeling of Capacity Factor of Rembang Coal-Fired Steam Power Plant Based on Machine Learning to Improve the Accuracy of Primary Energy Planning

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

[a] Master of Energy, School of Postgraduate, Diponegoro University, Jl. Imam Barjo, SH No 5, Semarang and 50275, Indonesia
[b] Department of Mechanical Engineering, Faculty of Engineering, Diponegoro University, Jl. Imam Barjo, SH No 5, Semarang and 50275, Indonesia
[c] Department of Statistics, Faculty of Science and Mathematics, Diponegoro University, Jl. Imam Barjo, SH No 5, Semarang and 50275, Indonesia
[1] ery.perdana@plnnusantarapower.co.id*; [2] sulardjaka@lecturer.undip.ac.id; [3] budiwarsito@lecturer.undip.ac.id
* corresponding author

ARTICLE INFO

Article history
Received
Revised
Accepted

Keywords
Capacity Factor
Regression
Machine Learning
Multiple Linear Regression
MAPE

ABSTRACT

The Rembang Coal-Fired Power Plant (PLTU Rembang), with a capacity of 2 x 315 MW, is a key power plant in Central Java, where fuel expenses represent the largest cost component. Accurate fuel procurement planning, which relies on projecting electricity sales, is essential to reduce these costs. This study develops and compares four machine learning-based Capacity Factor (CF) prediction models: random forest regression, support vector regression, multiple polynomial regression, and multiple linear regression. The independent variables are selected from internal and external sources using F-tests and t-tests. Among the four models, the multiple linear regression model demonstrated the smallest Mean Absolute Percentage Error (MAPE) of 7.83%. Using this model, the annual CF for PLTU Rembang in 2024-2026 is predicted to be between 82% and 84%, while the CF for February-June 2024 is expected to range from 87% to 91%. With a monthly CF prediction accuracy classified as very good (MAPE of 2.35%), these predictions are valuable for optimizing monthly fuel purchase allocations, considering initial fuel stock and target inventory age (17-30 Days of Plant Operation).

1. Introduction

Rembang Coal Fired Steam Power Plant (PLTU Rembang), with an installed capacity of 2 x 315 MW, plays a crucial role in the electricity system of Central Java. This plant maintains voltage stability and provides active power source in eastern Central Java, which previously faced limitations in electricity availability. Before the operation of PLTU Rembang, the northern coastal areas of Central Java, from Kudus to Cepu, heavily relied on the Tambak Lorok Power Plant and the transmission network along the Pantura route [1].

However, one of the biggest challenges in operating PLTU Rembang is the high fuel cost, as illustrated in Fig. 1.

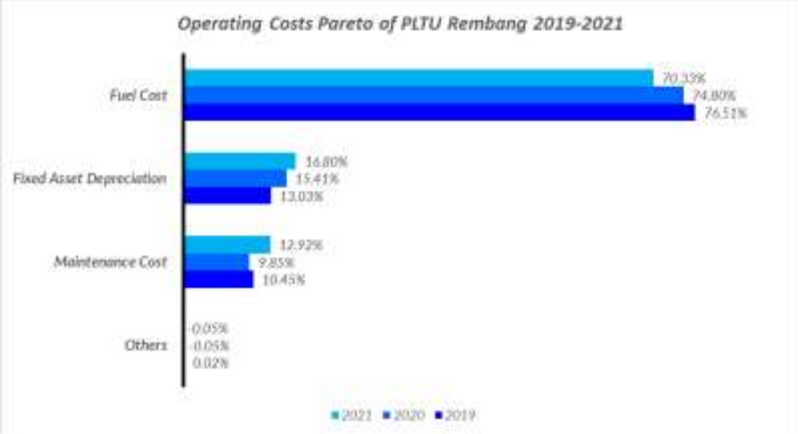


Fig. 1. PLTU Rembang Operational Costs 2019–2021

During the 2019–2021 period, fuel costs accounted for approximately 73.88% of the total operational costs, with a value of 70.33% in 2021 [2] . This indicates that fuel expenditure is the largest cost component that needs to be controlled. Even a small reduction in primary energy consumption can significantly reduce operational costs. Therefore, accurate fuel procurement planning becomes essential to minimizing the plant's operational expenses.

To optimize operational cost efficiency, accurate projections of the Capacity Factor (CF) are necessary. CF is the ratio of the actual energy produced by the plant to the total energy that could be generated if the generator were operated at full capacity without interruptions [3]. Accurate CF projections help determine fuel requirements, influencing fuel procurement planning. Operational data from PLTU Rembang from 2016 to 2021 show significant discrepancies between the planned CF in the Annual Operational Plan (ROT) and the actual CF. In 2017 and 2019, the differences between planned and actual CF reached -18.22% and -34.48%, respectively, resulting in higher unbudgeted fuel costs.

To address this issue, this study developed a CF prediction model based on machine learning to reduce these CF gaps. Four models were developed: random forest regression, support vector regression, multiple polynomial regression, and multiple linear regression. These models were chosen for their proven effectiveness in various applications. Independent variables were selected using the backward elimination method based on hypothesis testing (F-test and t-test).

Previous studies on CF prediction for power plants have several limitations, including the exclusion of external factors, a focus on short-term forecasts, and the use of limited modeling approaches. Babatunde et al. [4] implemented K-Fold Cross Validation but did not incorporate external variables, leading to reduced model robustness for long-term forecasting. Gufron et al. [5] utilized Support Vector Machine (SVM) for solar power plants but neglected crucial external factors such as weather conditions. Rashid et al. [6] employed random forest for wind turbine forecasting but only focused on isolated, non-interconnected systems. Tüfekci [7] applied regression models for gas power plants but restricted the analysis to base load operations with minimal load variations. Purwanto et al. [8] developed multiple regression models for PLTU Paiton; however, their study was limited to short-term predictions without considering interconnection effects.

This study bridges these gaps by introducing a more comprehensive CF prediction model that integrates both internal and external factors, capturing broader operational influences. Unlike previous studies, this research includes external grid conditions, such as Java-Bali and Central Java loads, and dynamic variables like Merit Order and Incremental Fuel Cost (IFC), which significantly impact CF variations in interconnected systems. The modeling process is conducted using Python on the Jupyter Anaconda platform, ensuring flexibility and scalability. The model selection is based on Mean Absolute Percentage Error (MAPE) and R-squared, with the optimal model demonstrating the highest accuracy and lowest prediction error for CF forecasting at PLTU Rembang. By addressing these methodological limitations, this study provides a more reliable and data-driven approach to CF prediction, which is essential for optimizing primary energy planning and operational efficiency in coal-fired power plants.

The predictions generated by the selected model are expected to improve the accuracy of primary energy planning and optimize the operational cost efficiency of PLTU Rembang.

2. Method

This study utilized primary data from the PLTU Rembang Operation Report, which includes key operational parameters such as Capacity Factor (CF), Equivalent Forced Outage Rate (EFOR), and Scheduled Outage Factor (SOF). EFOR and SOF were selected due to their significant impact on CF, as forced and scheduled outages directly reduce the available generation capacity and influence the plant's overall performance. By incorporating these parameters, the study ensures a comprehensive evaluation of the factors affecting CF at PLTU Rembang.

In addition to primary data, secondary data were obtained from PLN P2B, covering system-wide variables that affect power plant dispatching and performance. These secondary data include Java-Bali Load, Central Java Load, and Net Generating Capacity (DMN) for both regions, as well as Merit Order and Incremental Fuel Cost (IFC). These variables play a crucial role in determining the operational efficiency of PLTU Rembang, given that the plant is integrated into the Java-Bali interconnection system and the 150 kV Central Java-DIY subsystem. In this interconnected network, system-wide load conditions and dispatch priorities—driven by Merit Order and IFC—affect the plant's actual CF, making these external variables essential for accurate prediction modeling.

Unlike previous studies that rely on expert opinions or qualitative assessments, this research is grounded entirely in historical operational data and rigorous statistical modeling. The use of actual recorded data ensures that the study's findings are objective, reliable, and applicable to real-world operational decision-making. The dataset used in this study is sourced exclusively from official operational reports and real-time monitoring systems, minimizing the risk of biases or inaccuracies. By leveraging verified historical data, this study provides a robust and data-driven approach to predicting CF, offering valuable insights for power system optimization.

The types of data and their respective sources used in this study are summarized in Table 1. This table presents both primary and secondary data, outlining their relevance to CF prediction and their respective sources. The inclusion of these datasets enables a holistic analysis of CF by incorporating both plant-specific and grid-wide factors.

Table 1. Type and Source of Data

Data Type	Data Name	Data Source
Primary Data	CF PLTU Rembang SOF PLTU Rembang EFOR PLTU Rembang CF PLTU Unit 10 SOF PLTU Unit 10 EFOR PLTU Unit 10 CF PLTU Unit 20 SOF PLTU Unit 20 EFOR PLTU Unit 20	Production Report of PLTU Rembang
Secondary Data	Java-Bali Load Central Java Load Java-Bali DMN Central Java DMN	PLN P2B Load Monitoring Website
	Merit order *Incremental Fuel Cost (IFC)*	Monthly Release of PLN P2B in RAE

Furthermore, the methodology employed in this study follows a structured approach, including data preprocessing, selection of independent variables, determination of the best regression model, and application of CF predictions for primary energy allocation. The predictive modeling process was implemented using Python, chosen for its extensive libraries, ease of syntax, and high reliability in handling large datasets [9]. This structured workflow enhances the accuracy of CF predictions and facilitates their application in operational decision-making. The research flowchart outlining the key methodological steps is presented in Fig 2.

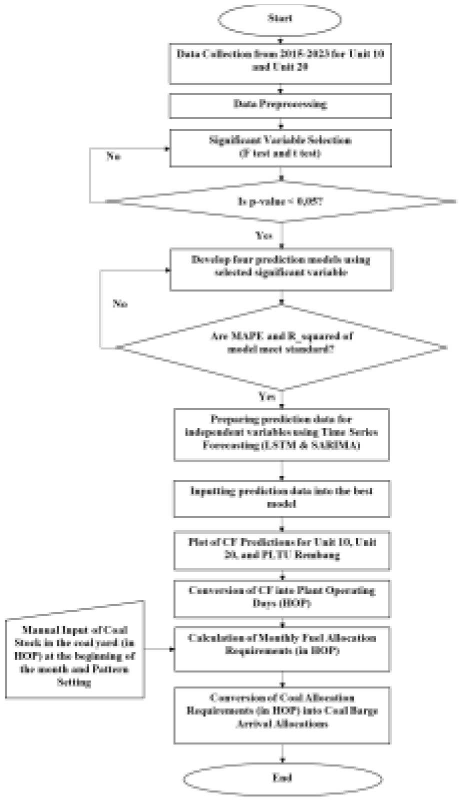


Fig. 2 Research Flowchart

The study began with data collection from 2015 to 2023. The data is presented in Table 2.

Table 2. CF Prediction Modeling Data

Month	Merit Order	EFOR (%)	SOF (%)	Java Bali Load (MW)	Central Java Load (MW)	Java Bali DMN (MW)	Central Java DMN (MW)	CF (%)
Jan 2015	2	2.72	0	22,375	3.608	31.189	5.193	94,24
Feb 2015	2	4.83	0	21,755	3.570	31.189	5.193	85,2
Mar 2015	3	0.51	0	22,356	3.730	31.189	5.193	93,31
Apr 2015	1	1.56	0	22,953	3.730	31.166	5.193	95,41
…	...	…	…	…	…	…	…	…
Sep 2023	20	0.76	0	29,975	4.996	43.349	11.094	90,58
Oct 2023	13	1.08	0	31,082	5.203	43.351	11.096	95,31
Nov 2023	11	1.14	0	31,515	5.248	43.489	11.096	94,75
Dec 2023	17	0.23	32,03	30,905	5.086	44.369	11.096	64,01

After data collection, preprocessing was conducted to detect missing data using linear interpolation. Significant variables were selected through F-tests and t-tests, where variables with p-values < 0.05 were deemed significant [10]. The prediction models developed included random forest regression, support vector regression, multiple polynomial regression, and multiple linear regression, which are common regression methods [11].

Model validation was carried out using MAPE (Mean Absolute Percentage Error) and R-squared. The best model had the smallest MAPE and largest R-squared. MAPE is the average absolute difference between predicted and actual values expressed as a percentage of the actual values, calculated using Equation (1) [12].

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

R-squared, or the coefficient of determination, measures the influence of independent variables (X) on the dependent variable (Y), as shown in Equation (2) [13].

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

where RSS is the residual sum of squares, y_i is the i - th actual value, $\bar{y}$ is the mean of the actual values, and $f(x_i)$ is the i - th predicted value.

The next step is to predict the independent variables. The predicted values of the independent variables are then used as input for the regression model. The techniques used to predict the dependent variable are shown in Table 3.

Table 3 Techniques for Predicting Independent Variables

No	Independent Variable	Prediction Technique
1	SOF	Using outage planning data
2	EFOR	Time series forecasting with RNN and LSTM
3	Jawa-Bali Load, Central Java Load	Time series forecasting with ARIMA and SARIMA
4	Jawa-Bali DMN, Central Java DMN	Using the latest RUPTL data
5	*Merit order, incremental fuel cost*	Time series forecasting with RNN and LSTM

The prediction of EFOR, merit order, and IFC uses LSTM and RNN algorithms because the variables are random. Meanwhile, the Java-Bali and Central Java loads are predicted using ARIMA and SARIMA because the variables exhibit a trend. After all independent variable predictions are obtained, the prediction of the dependent variable, which is CF, can be easily performed. The next step is the processing of CF predictions into monthly coal barge allocation requirements, as shown in Figure 3.

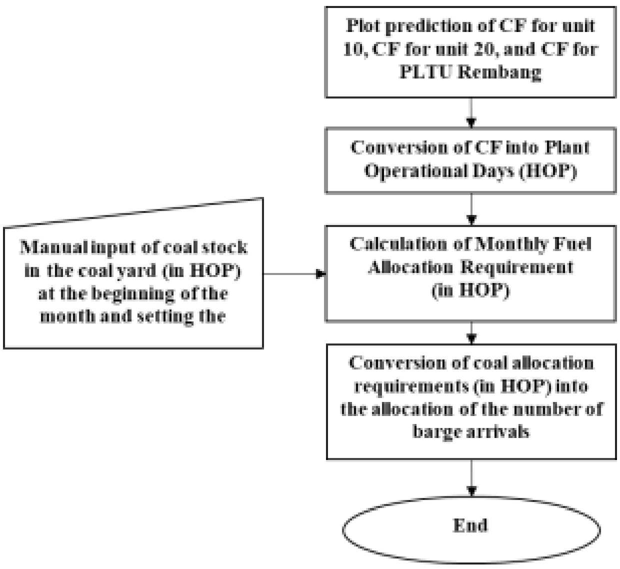


Fig. 3 Process of converting CF predictions into monthly coal barge allocation requirements

The process of converting CF predictions into monthly barge allocation requirements begins with converting CF to Power Plant Operational Days (HOP).

$$CF\ Prediction\ x\ DMN\ x\ Period\ Hours = Nett\ kWh \quad (3)$$

$$(Nett\ kWh + Auxiliary\ Power\ Assumption) x\ SFC\ Assumption = Coal\ Demand\ (Tonne) \quad (4)$$

$$CF\ Conversion\ (in\ HOP) = \frac{Coal\ Demand\ (Tonne)}{8523} \quad (5)$$

The result of converting CF is then compared with the initial coal stock at the coal yard for the beginning of the month in HOP. The difference becomes the HOP requirement. This HOP requirement is added to the monthly pattern setting data in HOP. The result is the coal allocation in HOP. Subsequently, the coal allocation in HOP is converted into monthly barge requirement allocation. This process is formulated in equations (6) to (8).

$$Coal\ Demand\ (in\ HOP) = CF\ Conversion\ (in\ HOP) - Initial\ Stock\ (in\ HOP) \quad (6)$$

$$Coal\ Allocation\ (in\ HOP) = Coal\ Demand\ (in\ HOP) + Pattern\ Setting\ (in\ HOP) \quad (7)$$

$$Coal\ Barge\ Allocation = \frac{Caol\ Allocation\ (in\ HOP) x\ 8523}{8000} \quad (8)$$

The value of the barge requirement allocation is then presented at the monthly coal coordination meeting.

3. Results and Discussion

The results of this study include selecting significant independent variables, developing and validating the model, predicting CF using the best model, and utilizing CF predictions.

3.1. Selection of Significant Variables

To identify significant variables affecting CF, hypothesis tests were conducted, namely F-test and t-test. The F-test, which assesses whether there is a relationship between independent and dependent variables [14], showed a p-value of 9.33×10^{-48}, far below 0.05, indicating the rejection of H_0. This confirms at least one variable significantly influences CF. Subsequently, the t-test combined with backward elimination was performed to identify significant independent variables. Table 4 shows the results.

Table 4. Backward Elimination Results

Independent Variable	*p-value*			
	Elimination Stage 1	**Elimination Stage 2**	**Elimination Stage 3**	**Elimination Stage 4**
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	$4{,}74x10^{-47}$	$1{,}65x10^{-47}$	$7{,}9x10^{-48}$	$5{,}58x10^{-48}$
Jamali Load	0,2299	0,1667	$8{,}59x10^{-5}$	$1{,}02x10^{-6}$
Central Java Load	0,5679	0,4009		
Jamali DMN	0,1146	0,0004	$6{,}45x10^{-5}$	$3{,}16x10^{-5}$
Central Java DMN	0,9337			

The eliminated variables are Beban Jateng, DMN Jateng, and IFC. Consequently, the significant variables influencing the CF of PLTU Rembang are Merit Order, EFOR, SOF, Beban Jamali, and DMN Jamali. Beban Jateng and DMN Jateng are considered insignificant due to the significant increase in Central Java's reserve capacity (from 1585 MW to 6010 MW between 2015 and 2023), which reduced the impact of Beban Jateng on the CF of PLTU Rembang.

3.2. Model Development and Validation

To obtain the best CF prediction model, four models were developed: random forest regression, support vector regression, multiple polynomial regression, and multiple linear regression. The data was split into 80% training data and 20% test data, which is a commonly used composition [15].

3.2.1 Model Development

a. Random Forest Regression

Random forest is a supervised machine learning algorithm that combines the outputs of multiple decision trees. Random forest regression consists of multiple trees that rely on random vectors, enabling predictor trees to take numerical values rather than class labels [16]. In this study, the random forest regression model was developed using 100 estimators.

b. Support Vector Regression

Support vector regression (SVR) is an application of Support Vector Machines (SVM) for regression analysis [17]. The basic idea of SVM is to map the data into a higher-dimensional feature space and construct a separating hyperplane with maximum margin [18]. In this study, the SVR model was developed with a linear kernel.

c. Multiple Polynomial Regression

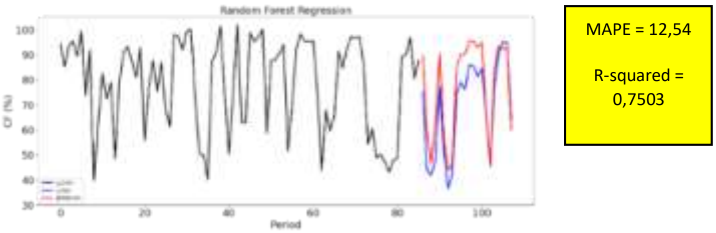
A regression is referred to as polynomial regression if the relationship between the dependent variable and independent variables can be represented by a curve [19]. In this study, the multiple polynomial regression model was developed with a second order.

d. Multiple Linear Regression

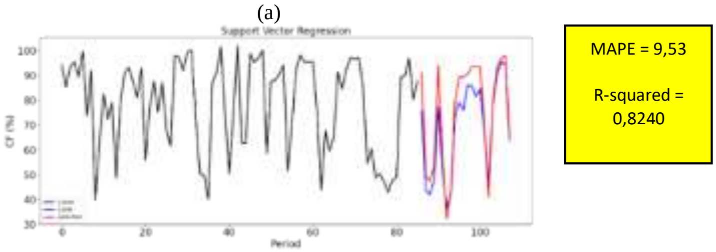
Multiple linear regression is essentially an extension of simple linear regression with more than one predictor variable [20]. In this study, the multiple linear regression model was developed using five independent variables.

3.2.1 Model Validation

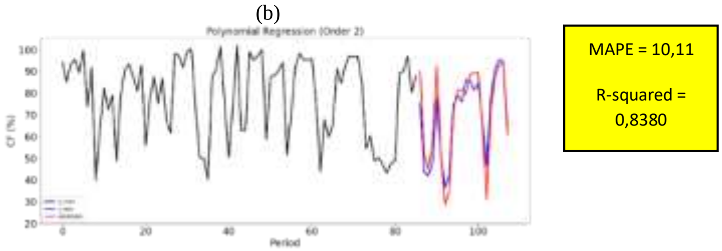
Model validation was conducted by calculating the MAPE and R-squared values for each prediction model. A comparison of the validation results for the four models is shown in Figure 4.



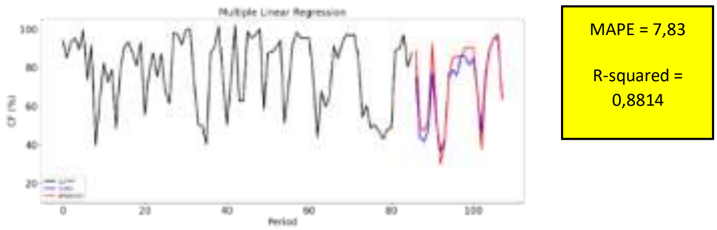
(a)



(b)



(c)



(d)

Fig. 4. Model Validation: (a) Random Forest Regression, (b) Support Vector Regression, (c) Multiple Polynomial Regression, (d) Multiple Linear Regression

The validation results indicate that multiple linear regression is the best model for predicting CF, as it has the smallest MAPE of 7.83% and the largest R-squared value of 0.8814. This model demonstrates a linear trend in the data, which can be observed from its highest R-squared coefficient.

3.3. CF Prediction with the Best Model

The CF prediction for PLTU Rembang over the next three years is made using the selected model, multiple linear regression. The multiple linear regression equation for predicting CF at PLTU Rembang is shown in Equation (9):

$$y = 70{,}1714 - 0{,}2887x_1 - 0{,}9191x_2 - 0{,}8960x_3 + 0{,}0035x_4 - 0{,}0018x_5 \tag{9}$$

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

Before predicting the CF of PLTU Rembang using the selected model, it is necessary to predict the values of the five independent variables for the next three years.

3.3.1 Merit Order Prediction

Merit Order is a significant variable affecting CF, but its values are random without a specific trend. To predict future values, a time series forecasting method based on artificial neural networks, namely RNN and LSTM, was employed. A comparison of validation for historical data prediction shows that the RNN model has a smaller MAPE and RMSE, at 20.17% and 4.4131, respectively. Thus, the RNN model was used to predict the merit order. The merit order predictions for the next three years are shown in Figure 5.

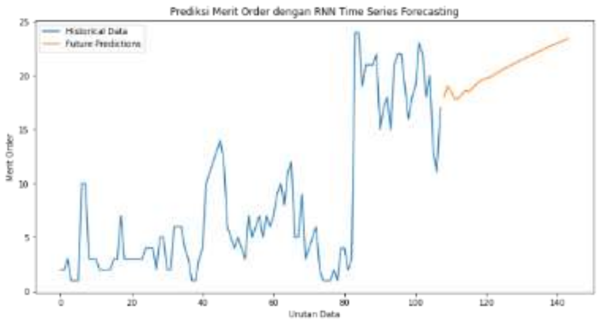


Fig. 5 Merit Order Prediction Using RNN

With a sequence length of 10, merit order prediction for PLTU Rembang shows an upward trend over the next three years. By 2026, merit order of PLTU Rembang is predicted to reach rank 23.

3.3.2 EFOR Prediction

EFOR pattern is similar to the merit order, as it is random. Therefore, EFOR value predictions were conducted using time series forecasting based on artificial neural networks, namely LSTM and RNN. A comparison of historical data prediction validation shows that the RNN model has a smaller

RMSE of 4.37. Thus, the RNN model was used to predict EFOR. The EFOR prediction results are shown in Figure 6.

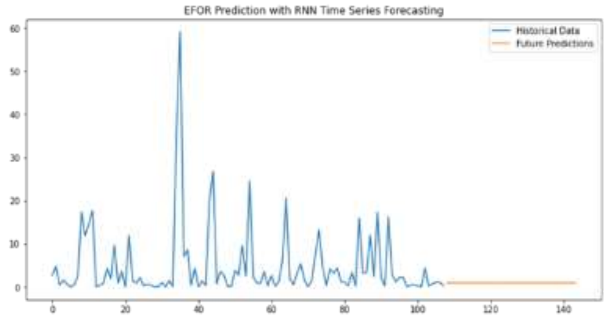


Fig. 6 EFOR Prediction Using RNN

With a sequence length of 10, the EFOR prediction for PLTU Rembang tends to remain constant over the next three years. The average EFOR value over this period is predicted to be below 1%.

3.3.3 SOF Prediction

The SOF value is a planned value. The Outage Management department of PLTU Rembang has prepared an outage schedule for PLTU Rembang. The duration of this planned outage is then converted into the planned SOF values. The SOF predictions for the next three years are shown in Table 5.

Table 5. SOF Prediction January 2024 – December 2026

Month	SOF #10	SOF #20	SOF PLTU
Jan-2024	0	9,68	4,84
Feb-2024	0	0	0
Mar-2024	0	0	0
…			
Oct-2026	0	50,45	27,7
Nov-2026	0	35,64	43,42
Dec-2026	0	0	0

3.3.4 Jamali Load Prediction

The Jamali load data pattern shows an upward trend over time. For Jamali load data with a specific trend, forecasting is conducted using classical time series methods, namely ARIMA and SARIMA. A comparison of historical data prediction validations shows that SARIMA has smaller MAPE and RMSE values, at 2.999% and 1019.83, respectively. The SARIMA model is used to predict Jamali load. The prediction results are shown in Figure 7.

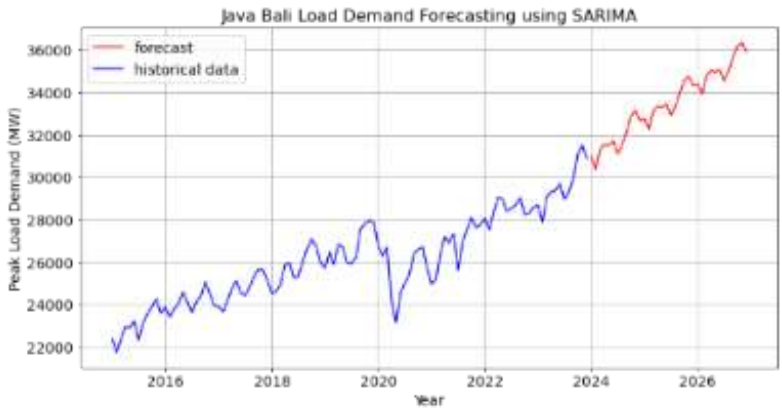


Fig. 7. Jamali Load Prediction with SARIMA

Using the SARIMA (1,1,1)(1,1,1,12) model, the Jamali load prediction shows an upward trend over the next three years. The highest Jamali load is predicted to reach 36,336 MW in 2026.

3.3.5 Jamali DMN Prediction

The DMN Jamali value can be predicted by referring to the latest RUPTL (Rencana Umum Penyediaan Tenaga Listrik). DMN Jamali predictions for the next three years are shown in Table 6.

Table 6. DMN Jamali Prediction 2024-2026

Bulan	DMN Jamali (MW)
Jan-2024	44.369
Feb-2024	44.369
Mar-2024	44.577
…	…
Oct-2026	51.627
Nov-2026	51.627
Dec-2026	51.737

3.3.6 Prediction of CF for PLTU Rembang

CF for PLTU Rembang serves as the dependent variable in the regression model. By using the multiple linear regression equation in equation (9) and the predicted values of the independent variables, the dependent variable, CF for PLTU Rembang, can be obtained. The prediction results for CF are shown in Figure 8 and Table 7.

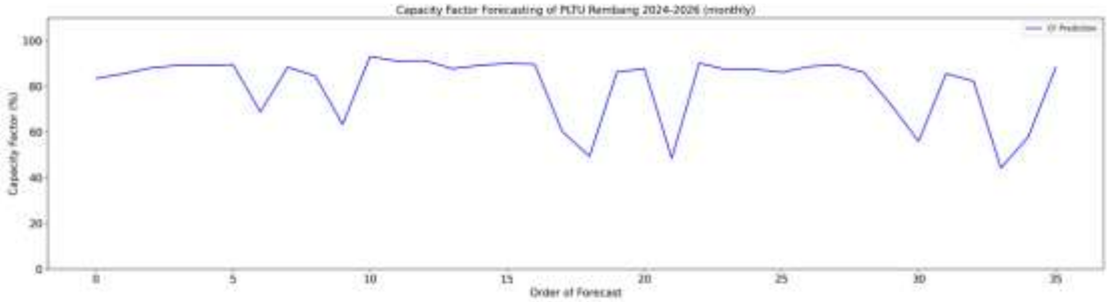


Fig. 8 CF Prediction for PLTU Rembang 2024-2026

Table 7 Monthly CF Predictions for PLTU Rembang (2024 – 2026)

Month	Merit Order	EFOR	SOF	Jamali Load	Jamali DMN	CF
Jan-2024	18	0,96	4,84	30.954	44.369	87,50
Feb-2024	19	0,9	0	30.379	44.369	89,60
Mar-2024	19	1	0	31.300	44.577	92,34
…	…	…	…	…	…	…
Oct-2026	23	0,94	50,45	36.186	51.627	50,34
Nov-2026	23	0,94	35,64	36.336	51.627	64,13
Dec-2026	23	0,94	0	35.929	51.737	94,45

Table 7 provides the monthly CF predictions from January 2024 to December 2026. The CF for PLTU Rembang is predicted to fluctuate, with the lowest value of 50.16% in September 2024 due to the overhaul period and the highest value in December 2024. These monthly predictions are crucial for the annual fuel allocation. The annual CF predictions are obtained by aggregating the monthly CF predictions, as shown in Table 8.

Table 8 Annual CF Predictions for PLTU Rembang (2024-2026)

Years	CF Prediction
2024	84,13
2025	84,83
2026	82,84

Table 8 shows that the annual CF predictions for PLTU Rembang range from 82% to 84%. These predictions are highly beneficial for primary energy planning.

3.4 Accuracy Evaluation of CF Predictions

The accuracy of CF predictions needs to be evaluated by comparing predicted CF values with the actual monthly CF realizations for PLTU Rembang during the February–June 2024 period. Monthly CF predictions were updated by adding new training data each month to enhance the predictive model's learning capability. A comparison of predicted CF values from four regression models and the actual CF realizations for February–June 2024 is shown in Table 9.

Table 9 Comparison of Predicted and Actual CF for PLTU Rembang (February–June 2024)

Prediction Month	CF Realizationn	CF RFR Prediction	CF MPR Prediction	CF SVR Prediction	CF MLR Prediction
February 2024	93,03%	87,72%	68,90%	91,73%	87,66%
March 2024	88,79%	87,47%	69,84%	92,96%	89,97%
April 2024	93,87%	87,48%	70,70%	93,77%	91,13%
May 2024	90,76%	86,98%	79,74%	93,57%	90,76%
June 2024	89,62%	87,43%	68,77%	92,03%	91,16%
MAPE		**4,12%**	**21,47%**	**2,40%**	**2,35%**

Based on Table 9, the Multiple Linear Regression (MLR) model consistently achieved the smallest MAPE of 2.35%, demonstrating its superior predictive capability compared to other models. The MAPE value below 10% confirms that the model provides highly accurate CF predictions [21], making it reliable for medium-term forecasting.

The analysis reveals that Merit Order, EFOR, and SOF are the most influential factors affecting CF, with EFOR showing a stronger negative impact than initially anticipated. Unlike previous studies that overlooked external factors, this research demonstrates that Java-Bali load variations significantly contribute to CF fluctuations, emphasizing the necessity of integrating grid-wide operational data into forecasting models.

To validate the model's reliability, the predicted CF values were compared with actual CF data, confirming that the MLR model effectively captures CF dynamics with an R^2 value of 0.8814. This indicates a strong correlation between the predicted and actual values, reinforcing the accuracy of the model.

The model was further tested through multiple training data splits and monthly retraining with updated data to assess its adaptability. The results consistently showed accurate predictions across different timeframes, confirming the model's robustness for medium-term forecasting. Additionally, scenario testing with different fuel cost assumptions demonstrated that the model remains stable under varying economic conditions, ensuring its practical applicability in real-world energy planning.

3.5 Utilization of CF Predictions

The CF prediction results are integrated into a web-based application called FOR CITY, which visualizes primary energy requirements.

Fig. 9 FOR CITY Application ("Forecasting of Capacity Factor")

Several assumptions are used in this process, including an SFC of 0.6333 kg/kWh and PS set at 5.5% of generator output. Table 10 provides guidance on period hours and monthly pattern settings used to estimate coal stock.

Table 10 Period Hours and Monthly Pattern Settings

Month	Period Hours (hours)	Pattern Setting (HOP)
January	744	31
February	672	25
March	744	25
April	720	25
May	744	23
June	720	23
July	744	23
August	744	25
September	720	27
October	744	30
November	720	39
December	744	40

The primary output of FOR CITY is the monthly coal allocation. For May 2024, the coal allocation is estimated at 376,546 tons. This data will be presented in the monthly Primary Energy Coordination Meeting. According to the meeting, the end-of-May 2024 HOP is expected to be approximately 24.8 HOP. Proposed HOP, confirmed HOP from the coordination meeting, and effective monthly HOP trends are shown in Figure 10.

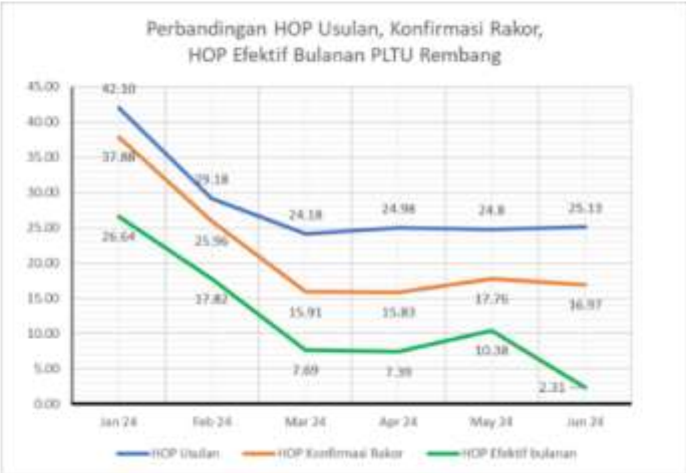


Fig. 10 Monthly Coal Stock Trends at PLTU Rembang for 2024

The trend indicates a decline in coal stock levels, with proposed stock consistently higher than confirmed stock. Cumulatively, total HOP at the end of June 2024 remains above 17 HOP, at 19.48 HOP. Accurate CF predictions help maintain cumulative total HOP above 17 HOP. For comparison, HOP for Semester 1 of 2023 was 25.27 HOP, while Semester 2 of 2024 is projected at 19.48 HOP.

In addition to monthly coal allocation, CF predictions are also utilized for annual allocation planning. The annual CF predictions presented in Table 8 are converted into annual coal requirements. The conversion results are shown in Table 11.

Table 11 Projected Coal Requirements (2024–2026) Based on CF Predictions

Year	Predicted CF (%)	Coal Requirement (tons)	Coal Requirement Projection (Rp)
2024	84,13%	2.773.374	2.462.756.386.839
2025	84,83%	2.788.809	2.476.462.814.440
2026	82,84%	2.723.388	2.418.368.260.618

Based on Table 11, the annual coal requirement projection for PLTU Rembang ranges between 2.7 million tons and 2.8 million tons. This projection is valuable for planning corporate coal procurement contracts.

4. Conclusion

This study successfully demonstrates that the Multiple Linear Regression (MLR) model is the most accurate approach for predicting the Capacity Factor (CF) of PLTU Rembang, outperforming other regression models. The model achieved an exceptionally low Mean Absolute Percentage Error (MAPE) of 2.35% and an R^2 value of 0.8814, confirming its high predictive accuracy. The validation process, conducted by comparing predicted CF values with actual realizations from February to June 2024, further reinforces the reliability of the model in real-world applications.

The findings highlight that Merit Order, EFOR, SOF, Java-Bali Load, and Java-Bali DMN are the dominant factors influencing CF, with EFOR exerting the strongest negative impact. The methodological process—incorporating data preprocessing, feature selection using F-tests and t-tests, and time-series forecasting of independent variables—has played a crucial role in improving prediction accuracy. Additionally, the iterative training approach, where the model is updated monthly with real-time data, has proven its adaptability and robustness across different timeframes.

Based on the CF predictions for 2024–2026, PLTU Rembang is expected to operate at an average CF of 82–84%, resulting in an estimated annual coal requirement of 2.7–2.8 million tons. The projected HOP levels remain above the critical threshold of 17 HOP, ensuring a stable and reliable power plant operation. These predictions are critical for optimizing fuel procurement strategies, maintaining sufficient coal stock levels, and supporting long-term energy planning.

By bridging the gaps left by previous studies, this research sets a new benchmark in CF prediction, demonstrating that integrating external grid-related factors significantly enhances forecasting accuracy in interconnected power systems. The validated methodology is highly scalable and applicable to other coal-fired power plants, providing a data-driven solution for strategic energy planning and coal procurement optimization.

With these breakthroughs, this study not only contributes to advancing machine learning applications in power generation but also provides a validated, real-world-tested framework for improving primary energy planning and decision-making processes.

Acknowledgment

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

Declarations

Author contribution. Ery Perdana: Writing-original draft, Developing the code, Methodology, Investigation and Analysis. Sulardjaka: Review, editing and supervision. Budi Warsito: Review, editing and Supervision
Funding statement. N/A
Conflict of interest. The authors declare no conflict of interest.

Data and Software Availability Statements

Data and Software will be made based on request.

References

[1] L. Rochman, “Feasibility Study PLTU - 1 Jawa Tengah 2 x (300-400) MW Rembang,” 2007, *PT Arkonin Engineering MP, Jakarta*.

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

[3] E. Grubert, “Same-plant trends in capacity factor and heat rate for US power plants, 2001–2018,” *IOP SciNotes*, vol. 1, no. 2, p. 024007, 2020, doi: 10.1088/2633-1357/abb9f1.

[4] D. E. Babatunde, A. N. Anozie, J. A. Omoleye, O. Oyebode, O. M. Babatunde, and A. O. Ayeni, “Modelling the power output from a steam power plant in Nigeria,” *Energy Reports*, vol. 6, pp. 822–828, 2020, doi: 10.1016/j.egyr.2020.11.126.

[5] H. Gufron, D. Rusirawan, and L. Widyawati, “Forecasting Produksi Energi PLTS 1 kWp Menggunakan Mesin Pembelajaran Dengan Algoritma Support Vector Machine,” *J. Tekno Insentif*, vol. 16, no. 2, pp. 79–91, 2022, doi: 10.36787/jti.v16i2.843.

[6] H. Rashid, W. Haider, and C. Batunlu, “Forecasting of Wind Turbine Output Power Using Machine learning,” *Proc. - Int. Conf. Adv. Comput. Inf. Technol. ACIT*, pp. 396–399, 2020, doi: 10.1109/ACIT49673.2020.9208852.

[7] P. Tüfekci, “Prediction of full load electrical power output of a base load operated combined cycle power plant using machine learning methods,” *Int. J. Electr. Power Energy Syst.*, vol. 60, pp. 126–140, 2014, doi: 10.1016/j.ijepes.2014.02.027.

[8] P. A. Purwanto, T. Ivantoro, and M. K. Jayadi, “Predictive Analytics of Electrical Power Output of Coal-Fired Power Plant Using Machine Learning,” *ICT-PEP 2021 - Int. Conf. Technol. Policy Energy Electr. Power Emerg. Energy Sustain. Smart Grid, Microgrid Technol. Futur. Power Syst. Proc.*, pp. 224–229, 2021, doi: 10.1109/ICT-PEP53949.2021.9601136.

[9] M. F. Sanner, “Python: a programming language for software integration and development,” *J Mol Graph Model*, vol. 17, no. 1, pp. 57–61, 1999.

[10] 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/

[11] D. Polzer, “7 of the Most Used Regression Algorithm : How to Choose the Right Ones.” Accessed: Oct. 24, 2022. [Online]. Available: https://towardsdatascience.com/7-of-the-most-commonly-used-regression-algorithms-and-how-to-choose-the-right-one-fc3c8890f9e3

[12] B. Putro, M. Tanzil Furqon, and S. H. Wijoyo, “Prediksi Jumlah Kebutuhan Pemakaian Air Menggunakan Metode Exponential Smoothing (Studi Kasus : PDAM Kota Malang),” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 2, no. 11, pp. 4679–4686, 2018, [Online]. Available: http://j-ptiik.ub.ac.id

[13] S. Weisberg, *Applied Linear Regression Fourth Edition*. New Jersey: John Wiley & Sons, Inc., 2014.

[14] M. H. Kutner, C. J. Natchtsheim, J. Neter, and Wi. Li, *Applied Linear Statistical Models 4th Edition*. New York: Mc Graw-Hill, 2004. doi: 10.1080/00224065.1997.11979760.

[15] S. Raschka, “Model Evaluation, Model Selection, and Algorithm Selection in Machine Learning,” 2018, [Online]. Available: http://arxiv.org/abs/1811.12808

[16] 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.

[17] 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.

[18] 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.

[19] Y. G. Akhlaghi, X. Ma, X. Zhao, S. Shittu, and J. Li, “A statistical model for dew point air cooler based on the multiple polynomial regression approach,” *Energy*, vol. 181, pp. 868–881, 2019.

[20] 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.

[21] P. C. Chang, Y. W. Wang, and C. H. Liu, “The development of a weighted evolving fuzzy neural network for PCB sales forecasting,” *Expert Syst. Appl.*, vol. 32, no. 1, pp. 86–96, 2007, doi: 10.1016/j.eswa.2005.11.021.