Advance Sustainable Science, Engineering and Technology (ASSET)
Vol. 8, No.4, October 2026, pp. 02604013-01 ~ 02604013-013
ISSN: 2715-4211 DOI: https://doi.org/10.26877/asset.v8i4.2378

Data-Driven Machine Learning Models for Steam Turbine Efficiency Prediction

Prima Arifa Alfariza, Hasna Fauziyah, Nikita Tjandra, Muhammad Asrol*

Industrial Engineering Department, BINUS Graduate Program - Master of Industrial Engineering, Bina Nusantara University, Jakarta 11480, Indonesia

*muhammad.asrol@binus.ac.id

Abstract. Increased electricity consumption in Indonesia requires coal-fired power plants to operate at high efficiency. This study proposes a data-driven approach to predict the efficiency of coal-fired steam turbines by utilizing 844 historical operational data points from the XYZ coal-fired power plant. Input variables include main steam pressure and temperature, main steam flow, final feedwater temperature, and condenser vacuum pressure. Four machine learning algorithms, namely Random Forest, Extra Trees, XGBoost, and Support Vector Regression, were evaluated using R^2, RMSE, and MAE. Based on the results of the study, it can be concluded that predictive models using Random Forest, Extra Trees, and XGBoost can predict steam turbine efficiency with high accuracy, especially after hyperparameter adjustments that increase the R^2 value and reduce the RMSE and MAE values in all models. The Extra Trees model proved to be the best model with an R^2 value of 0.8655 and an MAE of 0.5612, demonstrating its ability to capture the complex relationship between operational variables and turbine efficiency. This approach has practical implications for continuous improvement strategies in coal-fired power plant operations.

Keywords: Data-based predictions, importance of features, machine learning, steam turbine efficiency, thermal power plants.

(Received 2025-07-16, Revised 2025-09-18, Accepted 2026-06-24, Available Online by 2026-09-03)

1. Introduction

Electricity demand in Indonesia continues to increase, with per capita electricity consumption reaching 1,337 kWh/capita in 2023, up 13.98% from 1,173 kWh/capita in the previous year [1],[2]. Steam power plants play an important role as a primary energy source and must operate reliably to meet increasing electricity demand [3]. The efficiency of turbines, which convert thermal energy from steam into mechanical energy to turn generators, greatly affects the reliability of steam power plants [4]. Turbines are the second largest source of exergy loss after boilers, accounting for 6,403 kW (7.6%) of the total 84,193 kW [5], making improving turbine efficiency a crucial factor. Low efficiency not only increases fuel consumption, but also increases operating costs and reduces the competitiveness of power plants [6].

The method of monitoring steam turbine efficiency has thus far relied on physical approaches and manual inspections that are time-consuming, inflexible, and less than optimal in handling the complexity of operational variables. The availability of real-time operational data opens up opportunities for the application of machine learning (ML) as an alternative solution. ML algorithms are capable of handling non-linear relationships between variables, processing large amounts of data, and generating faster and more accurate predictions. This approach can not only detect efficiency declines earlier, but also provide recommendations for improvements and information about the operational variables that most affect turbine performance [7].

The application of ML in the energy sector has been proven to improve efficiency and predictive capabilities, for example through DNN and LS-SVM to predict SOx-NOx emissions in coal-fired power plants and the service life of transformers in nuclear power plants [7],[8]. Other studies show the role of ML in asset management, energy diagnostics, gas turbine prediction, power plant health monitoring, and steam turbine efficiency analysis based on operational data [9]-[12]. RF and SVR algorithms are effective in predicting power plant output [13], while Extra Trees excel in boiler data classification [14].

Furthermore, data approaches are used in steam turbine optimization to address load, design, and environmental variations [15], while the integration of IoT and ML has also been proven to improve energy prediction [16]. ML is increasingly being used for predictive maintenance, RUL prediction, hydro energy estimation, and factory performance monitoring, which have been proven effective in preventing early failure of critical components [17]-[20]. As the demand for electrical energy increases, the application of ML in monitoring and optimizing the efficiency of steam turbines in coal-fired power plants is becoming increasingly relevant [21].

However, there is still a significant research gap, where most studies focus more on power prediction, emissions, fault detection, load forecasting, and predictive maintenance, while studies targeting steam turbine efficiency in actual power plant operating conditions are still limited. In-depth analysis of the causal relationship between operational variables such as main steam pressure, main steam temperature, and condenser vacuum with turbine efficiency is rarely explained explicitly, as is the use of feature importance analysis in ML for operational improvement strategies.

The uniqueness of this research lies in the development of an ML-based steam turbine efficiency prediction model that emphasizes prediction accuracy and interpretability through feature importance. Using historical data from the DCS of the XYZ Company power plant, this study not only produced an accurate prediction model but also identified the most influential operational variables, thereby providing added value compared to previous studies that focused more on model evaluation metrics (MAE, RMSE, R^2).

The objective of this study is to develop a prediction model for the efficiency of coal-fired steam turbines using machine learning algorithms, while identifying the operational variables that most influence efficiency. In terms of managerial implications, this study is expected to support technical decision-making in coal-fired power plants through data-based optimization strategies. The practical benefits that can be achieved include increased turbine thermal efficiency, reduced fuel consumption, lower operating costs, and increased competitiveness of power plants. Thus, this research makes a real contribution to the sustainability of the national power generation system through the application of intelligent technologies that are adaptive to the complexity of power plant operations.

2. Methods

This study uses a data-based quantitative approach by applying machine learning algorithms (Random Forest, XGBoost, Extra Trees Regression, and Support Vector Machine) to predict steam turbine efficiency. The input variables used include Main Steam Pressure (MPa), Main Steam Temperature (°C), Main Steam Flow (tons/hour), Final Fuel Water Temperature (°C), and Condenser Pressure (kPa), while the target variable is Turbine Efficiency (%).

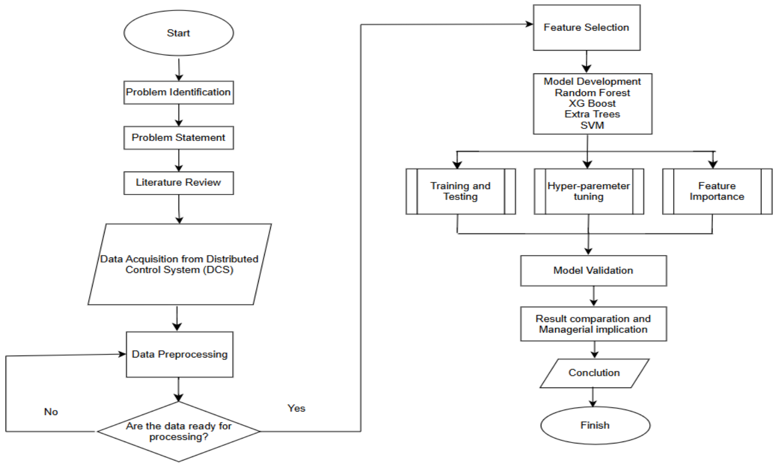

Figure 1. Research Flowchart

2.1 Pre-Processing and outlier removal

Operational turbine data was obtained from the Centralized Control System (DCS) of the XYZ Power Plant with a resolution of one hour, producing 844 data points. Preprocessing included the removal of irrelevant columns ('No', 'TIME'), checking for duplicates, and detecting and removing outliers using the Interquartile Range (IQR) method. To maintain model stability, normalization and standardization were performed with StandardScaler, especially on SVR, which is sensitive to scale. Correlation analysis with heatmaps was used to evaluate the relationship between variables, and features with very low correlations were eliminated.

2.2 Training and testing data method, feature selection and feature importance

Feature selection was performed using the embedded method through feature importance analysis on Random Forest, Extra Trees, and XGBoost. The data was divided into 80% training and 20% testing using train_test_split with a fixed random_state to ensure reproducibility. The 80/20 ratio was chosen because the dataset size was limited but still provided enough data for validation.

2.3 Model Development

The modeling stage is carried out using Random Forest, Extra Trees, XGBoost, and SVR algorithms in a structured pipeline to maintain the flow of pre-processing and training consistently on each model tested. The model utilizes five input variables: Main Steam Pressure (x_1, in MPa), Main Steam Temperature (x_2, in °C), Main Steam Flow (x_3, in ton/hour), Final Feedwater Temperature (x_4, in °C), and Condenser Pressure (x_5, in kPa), with Turbine Efficiency (y, in %) serving as the target output variable.

Random Forest Regression Mathematical Model: where each Decision tree is trained from a random subset of data and a random subset of features and the average of all trees will be the final prediction. (Eq. 1)

$$\hat{y}_{RF}(x) = \frac{1}{N}\sum_{i=1}^{N} T_i(x_{1,}\, x_{2,}\, x_{3,}\, x_4, x_{5,}) \tag{1}$$

T_i is the i-th Decision tree

EachT_i generates efficiency predictions based on a tiered division of features up to x_1 x_5

Random Extra Tree Mathematical Model where there is no boostrap and all data is used (Eq. 2)

$$\hat{y}_{ET}(x) = \frac{1}{N}\sum_{i=1}^{N} \tilde{T}_i(x_{1,}\, x_{2,}\, x_{3,}\, x_4, x_5)$$

$$(2)$$

$\tilde{T}_i$ is the I-th tree with random splits
XGBoost Mathematical Model (Eq. 3)

$$\hat{y}_{XGB}(x) = \sum_{m=1}^{M} f_m(x_{1,} x_{2,} x_{3,} x_4, x_{5,}) \quad f_m \in \mathcal{F} \tag{3}$$

With total loss function (Eq. 4)

$$\mathcal{L}^{(t)} = \sum_{i=1}^{n} [l(y_1, \hat{y}_i^{(t-1)} + f_t(x_i))] + \Omega(f_t) \tag{4}$$

Model regularization function (Eq. 5)

$$\Omega(f) = \gamma T + \frac{1}{2}\lambda \sum_{J=1}^{T} w \frac{2}{j} \tag{5}$$

To improve model performance, hyperparameter tuning was performed using GridSearchCV with 10-fold cross-validation for each algorithm. Hyperparameter tuning was performed on Random Forest, Extra Trees, XGBoost, and SVR to improve prediction accuracy and reduce error. In Random Forest and Extra Trees, important parameters that were set included "the number of trees (n_estimators), maximum depth (max_depth), minimum number of samples at a node (min_samples_split, min_samples_leaf), and the number of features considered (max_features)". In XGBoost, the settings include "n_estimators, max_depth, learning_rate, subsample, and colsample_bytree" to balance model capacity and the risk of overfitting.

For SVR, adjustments are made through a pipeline with StandardScaler, including kernel selection (linear, rbf, or poly), regulation parameters (C), error tolerance (epsilon), and kernel-specific parameters such as gamma, degree, and coef0.

In conclusion, the hyperparameter tuning process through grid exploration with cross-validation successfully produced an optimal configuration that improved R^2 and reduced error values (RMSE and MAE), enabling the model to better capture data characteristics without overfitting.

An analysis of feature importance in tree-based models is also presented to identify the most influential variables in efficiency prediction, which can be used as a basis for decision-making in coal-fired power plant operational optimization.

2.4 Model validation (RMSE, R2 and MAE)

Model performance evaluation was conducted using Root Mean Squared Error (RMSE) [22](Eq. 8), Mean Absolute Error (MAE) (Eq. 7), and R-squared (R^2) [23] (Eq. 6)metrics, which were selected because they provide a comprehensive overview of the model's ability to predict steam turbine efficiency compared to actual values.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2} \tag{6}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{7}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

(8)

R² (Eq. 6) is used to measure the variation in data explained by the model, MAE (Eq. 7) for mean absolute error, and RMSE (Eq. 8) to detect large errors. Scatter plot and time series graph visualizations are used to compare predicted values with actual values, while also verifying trend patterns and detecting deviations.

2.5 Reproducibility and Ethics Notes

All analyses were performed in Python (NumPy, Pandas, Matplotlib, Seaborn, Scikit-learn, XGBoost) with a consistent random_state. The dataset is limited and cannot be made public, but the processing procedures are described in detail so that they can be replicated on other PLTU data.

3. Results and Discussion

3.1 Data Pre-processing

Data preprocessing began with correlation map analysis (Figure 2) to examine the relationship between input features and turbine efficiency. The results show that Main Steam Flow (0.77) and Main Steam Pressure (0.72) have a strong correlation with efficiency, while Main Steam Temperature (0.20) has only a weak correlation, and Final Fuel Water Temperature shows a negative correlation. These findings form the basis for feature selection in the modeling stage

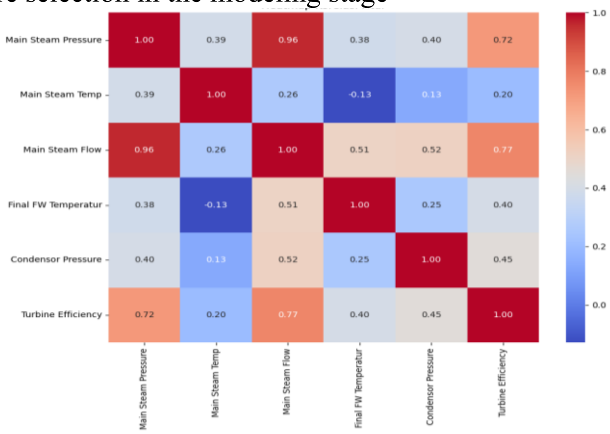

Figure 2. Heat Map Correlation Features for Initial Data

Initial examination of the dataset removed irrelevant columns such as No and TIME, and detected abnormal distributions and outliers through histograms. Outliers were then cleaned using the IQR method, so that the data distribution became more symmetrical and consistent. The results (in Figure 3) show that correlation analysis indicates an increase in the strength of the relationship, particularly between Final Inlet Water Temperature and turbine efficiency (from 0.40 to 0.79), as well as a small increase in Main Steam Pressure (0.73) and Main Steam Flow (0.77).

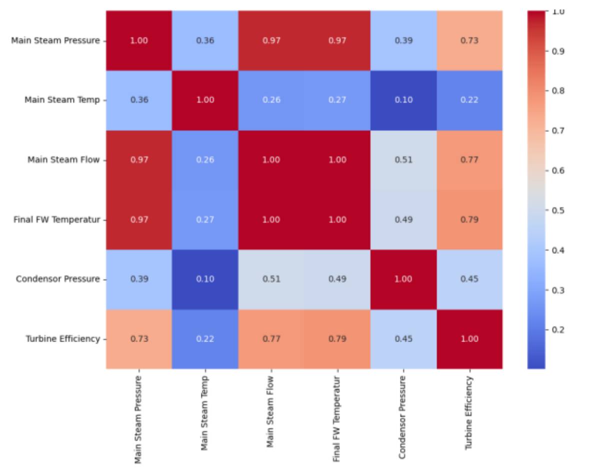

Figure 3. Heat Map Correlation Feature for the Cleared Dataset

Table 1. Descriptive Statistics of Features

Feature	count	mean	std	min	25%	50%	75%	max
Main Steam Pressure	844	6.510142	1.085548	4.79	5.52	6.11	7.72	8.17
Main Steam Temp	844	520.9736	5.068686	510.04	518.035	521.38	524.2925	533.47
Main Steam Flow	844	274.7638	62.15693	201.89	214.4575	248.655	345.3775	366.09
Final FW Temperature	844	215.6453	10.1794	202.13	205.475	212.325	227.4425	229.4
Condenser Pressure	844	-88.4719	1.710292	-91.72	-90.3125	-87.91	-87.31	-83.99
Turbine Efficiency	844	35.04387	2.242806	31.10668	33.06052	34.38769	37.47813	40.44481

Based on Table 1, the data shows stable characteristics and is suitable for modeling. Main Steam Pressure has an average of 6.51 MPa with a fairly symmetrical distribution (4.79–8.17 MPa), while Main Steam Temperature is relatively stable at 520.97°C with small variations (510.04–533.47°C). The Main Steam Flow Rate shows a wide range with an average of 274.76 tons/hour, reflecting variations in operating load. The Feed Water Final Temperature is stable (215.65°C, range 202.13–229.4°C), and the Condenser Pressure is consistent at -88.47 kPa with small variations (-91.72 to -83.99 kPa). The average Turbine Efficiency Target is 35.04%, with considerable variation (31.11–40.44%), indicating significant differences between operating conditions. Overall, the data does not show extreme values and can be used directly as a basis for predictive modeling without the need for additional cleaning.

3.2 Feature Selection

Feature selection was performed using an embedded method by calculating the importance level of features from Random Forest, Extra Trees, and XGBoost. Only features that contributed significantly to the target (Turbine Efficiency) were retained, while features with low contributions were not prioritized. The main features used include Main Steam Pressure (MPa), Main Steam Temperature (°C), Main Steam Flow Rate (tons/hour), Feed Water Final Temperature (°C), and Condenser Pressure (kPa). These five variables were selected because they are directly related to energy conversion efficiency, thermodynamic performance, power output, thermal cycle, and turbine discharge conditions.

All features were then processed through normalization or standardization to ensure consistency in the data scale. This is particularly important for scale-sensitive algorithms, such as Support Vector Regression (SVR), to improve the accuracy of model predictions.

3.3 Hyper Parameter Tuning Strategy

Hyperparameters were adjusted on four main models: Random Forest, XGBoost, Extra Trees, and SVR. The process was carried out with grid search and 10-fold cross-validation using the R^2 score as the main metric.

3.3.1 Random Forest Regressor

For the Random Forest model, tuning was performed on five important parameters:

a) **n_estimators**: [100, 200, 300, 400, 500] : Indicates the number of trees (*decision trees*) in the forest.
b) **max_depth**: [10, 20, 30, 40, None] : Specifies the maximum depth of the tree. If None is set, the tree is allowed to grow until all leaves are pure.
c) **min_samples_split**: [2, 5, 10] : The minimum number of samples to split a node.
d) **min_samples_leaf**: [1, 2, 4] : The minimum number of samples a tree leaf should have.
e) **max_features**: ['sqrt', 'log2', None] : The strategy for selecting the number of features to consider when forming the best split.

With these combinations, the total possible parameter combinations for Random Forest are $5 \times 5 \times 3 \times 3 \times 3 = 675$ combinations, which were evaluated through cross-validation.

3.3.2 XGBoost Regressor

The tuning parameters for the XGBoost model include:

a) **n_estimators**: [100, 200, 300, 400]
b) **max_depth**: [3, 5, 7, 9, 11]
c) **learning_rate**: [0.01, 0.05, 0.1, 0.2]
d) **subsample**: [0.6, 0.8, 1.0] : Fraction of training data used to build each tree.
e) **colsample_bytree**: [0.6, 0.8, 1.0] : Fraction of features used when building the tree.

Total parameter combinations for XGBoost:
$4 \times 5 \times 4 \times 3 \times 3 = 720$ combinations.

3.3.3 Extra Trees Regressor

The customized parameters are almost identical to Random Forest, including:

a) **n_estimators**: [100, 200, 300, 400]
b) **max_depth**: [10, 20, 30, 40, None]
c) **min_samples_split**: [2, 5, 10]
d) **min_samples_leaf**: [1, 2, 4]
e) **max_features**: ['sqrt', 'log2', None]

Total combinations:
$4 \times 5 \times 3 \times 3 \times 3 = 540$ combinations.

3.3.4 Support Vector Regressor (SVR)

The SVR model was used in the *pipeline* with *StandardScaler*, and tuning was performed on the parameters:

a) **kernel**: ['rbf', 'linear', 'poly']
b) **C**: [0.1, 1, 10, 100] : Regularization parameter.
c) **epsilon**: [0.01, 0.1, 0.2]
d) **gamma**: ['scale', 'auto']
e) **degree**: [2, 3, 4] (only relevant if kernel = 'poly')
f) **coef0**: [0.0, 0.1, 0.5]

Since degree and coef0 only matter for the 'poly' kernel, not all combinations are fully used for all kernels. But roughly, the estimated total number of combinations is close:
3 (kernels) $\times 4 \times 3 \times 2 \times 3 \times 3 = 648$ combinations (maximum).

3.4 Model Development Result

This section presents the results of steam turbine efficiency predictions using the XGBoost, SVR, Random Forest, and Extra Trees algorithms. The evaluation was conducted using scatter and line diagrams to compare the predicted values with the actual values, while also assessing the accuracy and consistency of each model in capturing turbine efficiency trends.

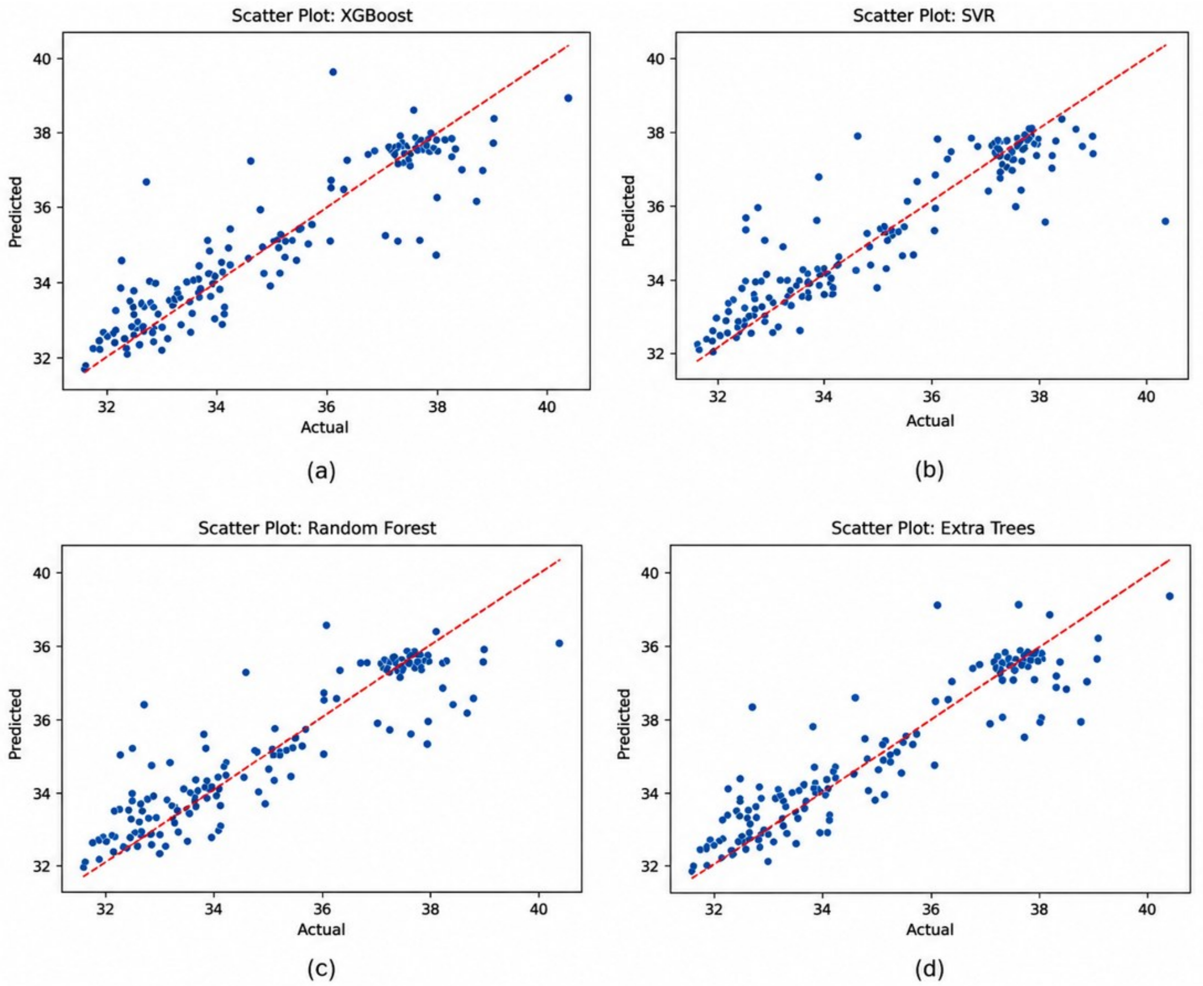

Figure 4. Scatter Plot prediction results for (a) XG Boost, (b) SVR, (c) Random Forest, (d) Extra Trees

The scatter plot between the predicted and actual values shows that Random Forest and Extra Trees produce predictions that are very close to the ideal line (y = x), indicating good predictive power. XGBoost shows larger deviations, especially at high efficiency values, indicating possible overfitting or sensitivity to non-linearity.

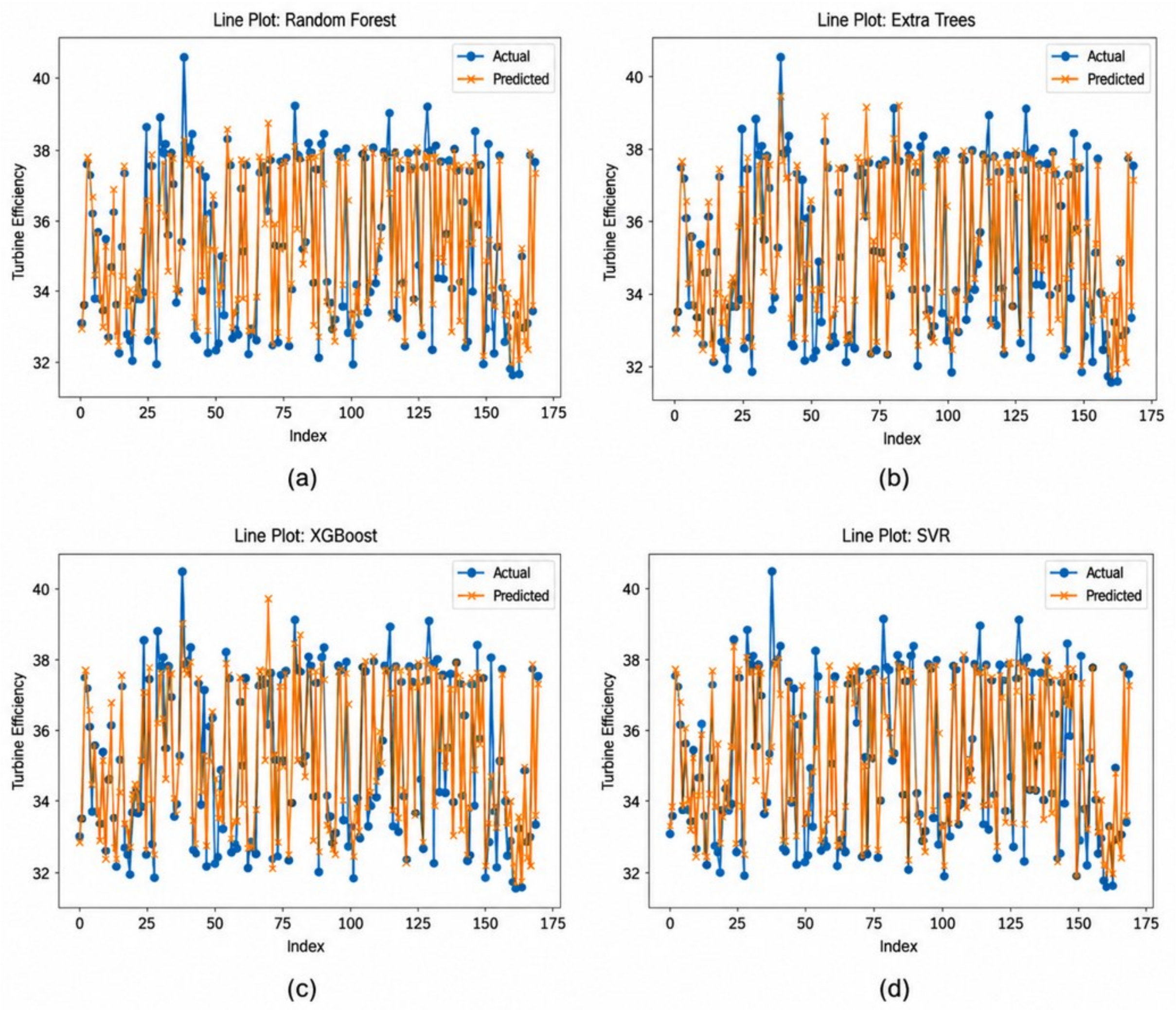

Figure 5. Line Plots of prediction results for (a) Random Forest, (b) Extra Tree, (c) XG Boost, (d) SVR

The visualization results (Figure 5) show that Extra Trees and Random Forest produce the most stable and accurate turbine efficiency predictions compared to other models, because their prediction trends are more consistent and closer to actual values, making them potentially implementable in real-time turbine efficiency monitoring to support power plant operation optimization.

3.5 Model Validation Result

Four ML algorithms-Random Forest, Extra Trees, XGBoost, and SVM-were used to develop the regression model. The data was split 80% for training and 20% for testing. Model evaluation was performed using R^2, RMSE, and MAE metrics.

Table2 . Evaluation Metrics

Model	R^2 basic model	R^2 hyperparameter tuning	RMSE basic model	RMSE hyperparameter tuning	MAE basic model	MAE hyperparameter tuning
Random Forest	0.8381	0.8405	0.9118	0.9052	0.6034	0.5957
XGBoost	0.8006	0.8476	1.012	0.8848	0.6284	0.5813
Extra Trees	0.8579	0.8655	0.8544	0.8311	0.5732	0.5612
SVR (RBF)	0.8349	0.8377	0.9208	0.913	0.5748	0.5738

Table 2 shows that all models experienced improved performance after hyperparameter optimization, as indicated by an increase in R^2 values and a decrease in RMSE and MAE values. XGBoost recorded the most significant improvement (R^2 from 0.8006 to 0.8476), while Extra Trees remained the best model with the highest R^2 (0.8655) and lowest MAE (0.5612). Random Forest also showed improvement, albeit relatively small, while SVR only experienced a marginal improvement. Overall, Extra Trees is recommended for implementation due to its accuracy and consistency, with XGBoost as a competitive alternative.

3.6 Feature Importance Analysis

The feature importance analysis (Figure 6) shows that Main Steam Flow is the most dominant variable in predicting turbine efficiency in all models. Random Forest and Extra Trees also highlight the significant role of Final Inlet Water Temperature and Condenser Pressure, while XGBoost tends to emphasize only Main Steam Flow. Overall, Extra Trees is rated as the best model because it provides the highest prediction accuracy and stability, making it a worthy primary reference in data-based turbine efficiency optimization.

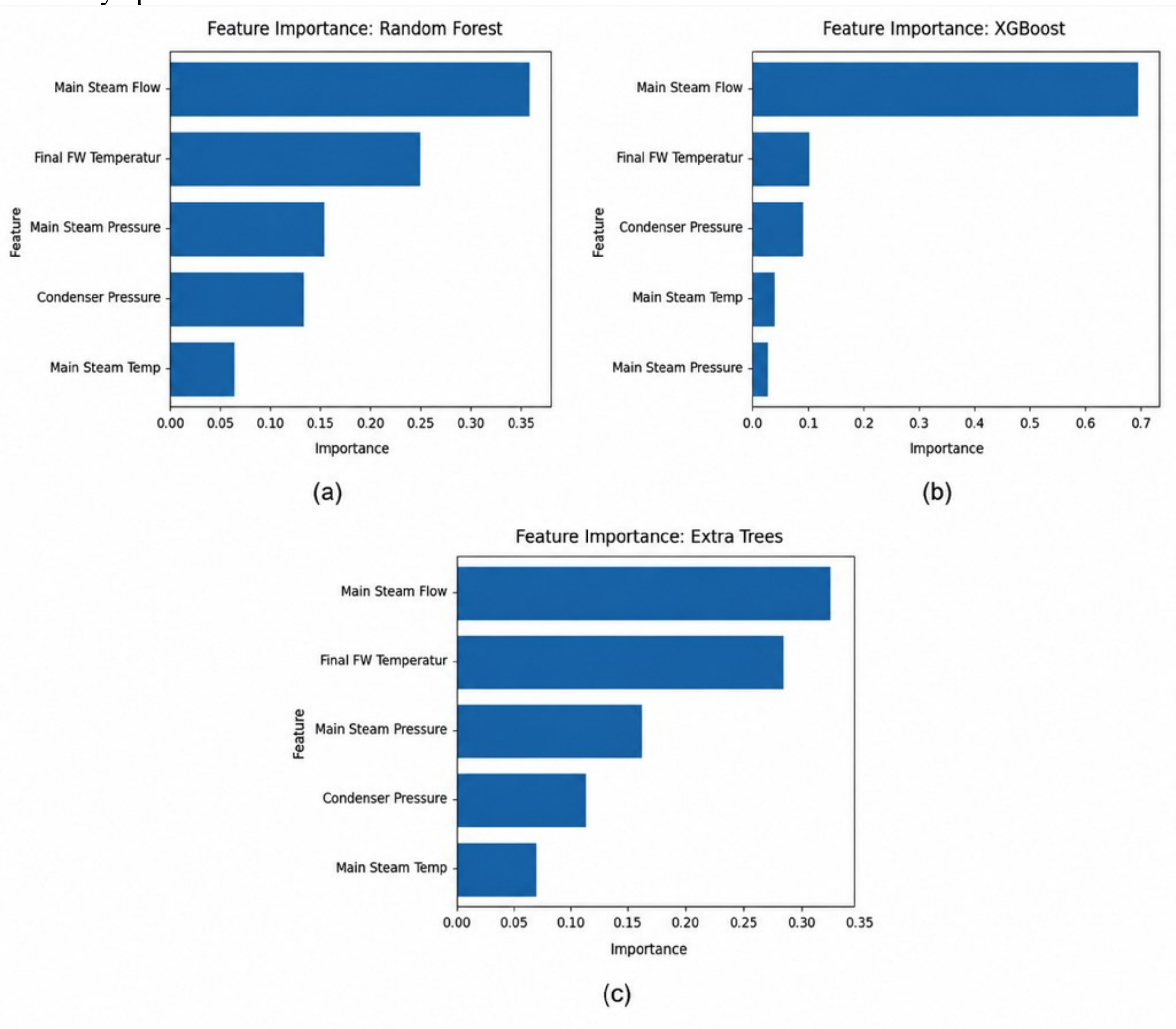

Figure 6. Feature Importance for (a) Random Forest, (b) Extra Tree, (c) XG Boost

Discussion

The application of a machine learning (ML)-based steam turbine efficiency prediction model makes an important contribution to managerial strategies at coal-fired power plants. With its ability to generate real-time predictions, this model enables more accurate and continuous monitoring of turbine performance. This is in line with research by Chatterjee et al. [24], which emphasizes that the application of ML in power generation systems can support early detection of performance decline, enabling management to take corrective measures before significant operational losses occur.

The results of the feature importance analysis in this study confirm that the variables of Main Steam Flow, Final Fuel Water Temperature, and Condenser Pressure are the dominant factors affecting turbine efficiency. These findings are consistent with the study by Han et al. [25], which shows that steam flow conditions and condensation system performance are the main determinants of power plant thermal efficiency. By understanding these key variables, power plant management can focus maintenance efforts on the most critical components, making optimization strategies more effective.

In addition, the use of this predictive model opens up opportunities for condition-based maintenance, which is more cost-effective than conventional maintenance schedules. Research by Al-Humairi et al. [26], shows that this type of predictive approach can reduce unnecessary cleaning operations by around 30%, thereby reducing mechanical wear and maintenance costs. Furthermore, by minimizing delays in necessary cleaning, the system is able to increase annual energy production by 3–5% under high contamination conditions. Thus, the application of this predictive model not only improves the operational reliability of turbines, but also supports cost efficiency and operational sustainability of coal-fired power plants.

From a policy perspective, these research results support the integration of ML-based monitoring dashboards into energy management systems at coal-fired power plants. Thus, the integration of IoT and ML in power plant monitoring can increase operational transparency and accelerate decision-making. By establishing Key Performance Indicators (KPIs) based on efficiency predictions, management can more quickly identify deviations and make continuous improvements. Ultimately, this strategy not only contributes to technical efficiency but also supports carbon emission reduction targets through fuel consumption optimization.

4. Conclusion

This study proves that machine learning algorithms particularly Extra Trees, are capable of predicting steam turbine efficiency with high accuracy ($R^2 = 0.8655$; MAE = 0.5612) using historical operational data from coal-fired power plants. The main contribution of this study is to demonstrate the superiority of a data-driven approach in capturing complex relationships between variables and identifying key factors, namely main steam flow, final feedwater temperature, and condenser pressure, which most influence turbine efficiency. The uniqueness of this research lies in its application to a real coal-fired power plant, providing relevant technical insights for sustainable efficiency improvement strategies. However, this study has limitations because it only uses a dataset of 844 points from one PLTU unit, so the generalization of the results is still limited. Therefore, further research needs to be conducted involving multi-PLTU data, a wider variety of operational conditions, and the development of a real-time prediction system so that it can be directly implemented in operational control. The direction of future research is expected to strengthen model validation while making a tangible contribution to improving performance and energy efficiency in Indonesia's electricity sector.

Declaration of AI and AI assisted technologies in the writing process
During the preparation of this work the author(s) used ChatGPT (OpenAI) to improve the language, grammar, clarity, and readability of the manuscript. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the publication

Declaration of Competing Interest
The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper

Acknowledgements
The authors gratefully acknowledge the institution that provided access to the research data and facilitated the data collection process.

References

[1] D. F. Silalahi, A. Blakers, and C. Cheng, “100% renewable electricity in Indonesia,” *Energies*, vol. 17, no. 1, p. 3, 2023, doi: https://doi.org/10.3390/en17010003.
[2] R. B. Wibowo and S. M. Damayanti, “Financial Feasibility of Geothermal Project J-3 and J-4 in Indonesia,” *Cognizance Journal of Multidisciplinary Studies*, vol. 5, no. 5, pp. 279–287, May 2025, doi: https://doi.org/10.47760/cognizance.2025.v05i05.018.
[3] A. E. Putra, “Literature Review: Study on the Utilization of Solar Power Plants (PLTS) as an Effort to Increase Electricity Access in Rural and Remote Areas,” *International Journal of Research in Vocational Studies (IJRVOCAS)*, vol. 5, no. 1, pp. 31–43, 2025, doi: https://doi.org/10.53893/ijrvocas.v5i1.352.
[4] A. S. Karakurt and Ü. Güneş, “Performance analysis of a steam turbine power plant at part load conditions,” *Journal of Thermal Engineering*, vol. 3, no. 2, pp. 1121–1128, 2017, doi: https://doi.org/10.18186/thermal.298611.
[5] O. J. Khaleel, F. Basim Ismail, T. Khalil Ibrahim, and S. H. bin Abu Hassan, “Energy and exergy analysis of the steam power plants: A comprehensive review on the Classification, Development, Improvements, and configurations,” *Ain Shams Engineering Journal*, vol. 13, no. 3, p. 101640, 2022, doi: https://doi.org/10.1016/j.asej.2021.11.009.
[6] F. A. Hatem, M. K. A. Alhumairi, M. A. Al-Obaidi, A. T. Mohammad, and A. S. K. Darwish, “Upgrading Gas Turbine Efficiency for Sustainable Power Generation: An Energy and Exergy Analyses,” *Results in Engineering*, vol. 105489, 2025, doi: https://doi.org/10.1016/j.rineng.2025.105489.
[7] D. Adams, D. H. Oh, D. W. Kim, C. H. Lee, and M. Oh, “Prediction of SOx–NOx emission from a coal-fired CFB power plant with machine learning: Plant data learned by deep neural network and least square support vector machine,” *Journal of Cleaner Production*, vol. 270, p. 122310, 2020, doi: https://doi.org/10.1016/j.jclepro.2020.122310.
[8] J. I. Aizpurua, S. D. J. McArthur, B. G. Stewart, B. Lambert, J. G. Cross, and V. M. Catterson, “Adaptive Power Transformer Lifetime Predictions Through Machine Learning and Uncertainty Modeling in Nuclear Power Plants,” *IEEE Transactions on Industrial Electronics*, vol. 66, no. 6, pp. 4726–4737, 2019, doi: https://doi.org/10.1109/TIE.2018.2860532.
[9] F. Aminifar, M. Abedini, T. Amraee, P. Jafarian, M. H. Samimi, and M. Shahidehpour, *A review of power system protection and asset management with machine learning techniques*, vol. 13, no. 4. Springer Berlin Heidelberg, 2022. doi: https://doi.org/10.1007/s12667-021-00448-6.
[10] M. M. Forootan, I. Larki, R. Zahedi, and A. Ahmadi, “Machine Learning and Deep Learning in Energy Systems: A Review,” *Sustainability (Switzerland)*, vol. 14, no. 8, 2022, doi: https://doi.org/10.3390/su14084832.
[11] Z. Liu and I. A. Karimi, “Gas turbine performance prediction via machine learning,” *Energy*, vol. 192, 2020, doi: https://doi.org/10.1016/j.energy.2019.116627.
[12] P. Hundi and R. Shahsavari, “Comparative studies among machine learning models for performance estimation and health monitoring of thermal power plants,” *Applied Energy*, vol. 265, no. February, p. 114775, 2020, doi: https://doi.org/10.1016/j.apenergy.2020.114775.
[13] S. Alketbi, A. B. Nassif, M. A. Eddin, I. Shahin, and A. Elnagar, “Predicting the power of a combined cycle power plant using machine learning methods,” *Proceedings of the 2020 IEEE International Conference on Communications, Computing, Cybersecurity, and Informatics, CCCI 2020*, 2020, doi: https://doi.org/10.1109/CCCI49893.2020.9256742.
[14] J. Dhanalakshmi, N. Ayyanathan, and N. S. Pandian, “Energy analytics and comparative performance analysis of machine learning classifiers on power boiler dataset,” *Proceedings of*

the 11th International Conference on Advanced Computing, ICoAC 2019, pp. 72–78, 2019, doi: https://doi.org/10.1109/ICoAC48765.2019.246819.
[15] L. Zhao and F. You, “A data-driven approach for industrial utility systems optimization under uncertainty,” *Energy*, vol. 182, pp. 559–569, 2019, doi: https://doi.org/10.1016/j.energy.2019.06.086.
[16] C. Emexidis and P. Gkonis, “The Integration of Internet of Things and Machine Learning for Energy Prediction of Wind Turbines,” *Applied Sciences (Switzerland)*, vol. 14, no. 22, 2024, doi: https://doi.org/10.3390/app142210276.
[17] W. Zhang, D. Yang, and H. Wang, “Data-Driven Methods for Predictive Maintenance of Industrial Equipment: A Survey,” *IEEE Systems Journal*, vol. 13, no. 3, pp. 2213–2227, 2019, doi: https://doi.org/10.1109/JSYST.2019.2905565.
[18] J. Jung, H. Han, K. Kim, and H. S. Kim, “Machine learning-based small hydropower potential prediction under climate change,” *Energies*, vol. 14, no. 12, 2021, doi: https://doi.org/10.3390/en14123643.
[19] G. Prokhorskii, M. Preißinger, S. Rudra, and E. Eder, “An interpretable and adaptable data-driven model for performance prediction in thermal plants,” *Energy Conversion and Management: X*, vol. 26, no. November 2024, p. 100950, 2025, doi: https://doi.org/10.1016/j.ecmx.2025.100950.
[20] J. F. Olesen and H. R. Shaker, “Predictive maintenance for pump systems and thermal power plants: State-of-the-art review, trends and challenges,” *Sensors (Switzerland)*, vol. 20, no. 8, 2020, doi: https://doi.org/10.3390/s20082425.
[21] C. Bisset, P. V. Z. Venter, and R. Coetzer, “A systematic literature review on machine learning applications at coal-fired thermal power plants for improved energy efficiency,” *International Journal of Sustainable Energy*, vol. 42, no. 1, pp. 845–872, 2023, doi: https://doi.org/10.1080/14786451.2023.2244618.
[22] G. Prokhorskii, M. Preißinger, S. Rudra, and E. Eder, “An interpretable and adaptable data-driven model for performance prediction in thermal plants,” *Energy Conversion and Management: X*, vol. 26, no. November 2024, p. 100950, 2025, doi: https://doi.org/10.1016/j.ecmx.2025.100950.
[23] R. S. Kumar, A. R. Singh, P. L. Narayana, V. S. Chandrika, M. Bajaj, and I. Zaitsev, “Hybrid machine learning framework for predictive maintenance and anomaly detection in lithium-ion batteries using enhanced random forest,” *Scientific Reports*, vol. 15, no. 1, pp. 1–20, 2025, doi: https://doi.org/10.1038/s41598-025-90810-w.
[24] S. Chatterjee, P. W. Khan, and Y. C. Byun, “Recent advances and applications of machine learning in the variable renewable energy sector,” *Energy Reports*, vol. 12, pp. 5044–5065, 2024, doi: https://doi.org/10.1016/j.egyr.2024.09.073.
[25] X. Han, Q. Zhu, J. Guan, and Z. Han, “Research on wet steam condensation flow characteristics of steam turbine last stage under zero output condition,” *International Journal of Thermal Sciences*, vol. 179, p. 107691, 2022, doi: https://doi.org/10.1016/j.ijthermalsci.2022.107691.
[26] A. Al-Humairi, E. Khalis, Z. A. Al Hemyari, and P. Jung, “Machine learning-based predictive maintenance for photovoltaic systems,” *AI*, vol. 6, no. 7, p. 133, 2025, doi: https://doi.org/10.3390/ai6070133.