Proceedings World Geothermal Congress 2020+1
Reykjavik, Iceland, April - October 2021

A Robust Prediction Method Based on Artificial Neural Network for Power Output of Organic Rankine Cycle in Lahendong Geothermal Field

Satriyo Nurhanudin Wibowo[1], Prasetyo Aji[2], Cahyadi[2], Suyanto[2], M. Taufiq Fathaddin[1], Hari Karyadi Oetomo[1], Kris Pudyastuti[1], Yusraida Khairani Dalimunthe[1], M.N. Ali Akbar[3]

[1]Trisakti University, Jakarta Indonesia; [2]B2TKE-BPPT, Serpong, Indonesia; [3]Mol Group, Budapest, Hungary

satriyonwibowo@gmail.com

Keywords: Power Output Prediction, Artificial Intelligence, Real-Time Monitoring System, Remote Mobile Monitoring, Artificial Neural Network

ABSTRACT

The industry has adopted Artificial Intelligence (AI), and it is a significant opportunity for the industry; especially the application of Machine Learning (ML) to improve prediction results, which will affect decision making. The Lahendong Organic Rankine Cycle (ORC) power plant offers a unique cycle in power output optimization, with an additional cycle called the intermediate cycle using the hot water. As the most potential renewable energy in Indonesia, predicting performance power plant output is a crucial task, by implementing ML along with the Real-Time Monitoring System (RTMS).

The power prediction model follows the application of the Artificial Neural Network (ANN). A dataset with the Multilayer Perceptron (MLP) has been developed to find the most robust solution in prediction. It contains the historical measured variables from; brine temperature, brine pressure, hot water temperature, hot water pressure, and power output. The dataset is split into two parts; 70% for the data training and 30% for validation. Meanwhile, the accuracy between the power actual data and its prediction is evaluated by using normalized Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and R-squared ($R^{2)}$

Overall, the prediction results from ANN are excellent. The R^2 score around 0.9954 for the power data training, with a validation score of 2.9 % for MAE and 9.9 % for RMSE. The power validation data also shows an impressive result of 0.9955 R^2, with a validation score of 4.7 % for MAE and 9.9 % for RMSE. Both results were categorized as excellent, with R^2 score close to one and a validation score below 10 %. Furthermore, this study surely can improve fast and accurate decision-making in Lahendong ORC power plant, especially for power plant management.

1. INTRODUCTION

Indonesia is the largest geothermal energy source globally, with a potential up to 40% of the total geothermal sources in the world and an estimated amount of around 30 GWe (Wahjosoebidjoand Hasan, 2018). It consists of 252 geothermal source locations spread across the volcanic formation path, covering several islands, such as Sumatra, Java, Sulawesi, Maluku, and Nusa Tenggara (Suharmanto et al., 2013).

As the geothermal energy source location is spread along across islands, a Real-Time Monitoring System (RTMS) is crucially needed to monitoring power plants from remote locations. One of the power plants using RTMS is Lahendong Organic Rangkine Cycle (ORC). The Lahendong ORC power plant has been successfully commissioned in Lahendong, North Sulawesi, in March 2018 with an installed capacity of 500 kW (Wibowo et al., 2020). ORCs are the incomparable technical solution for generating electricity from low-medium temperature heat sources of limited capacity (Macchi, 2017). **Figure 1** shows the Lahendong ORC power plant.

Figure 1: Lahendong ORC Power Plant (Wibowo et al., 2020)

1.1 Lahendong ORC Power Plant

In general, the ORC power plant uses a single pressure cycle consisting of a brine cycle, working fluid cycle, and condensation cycle (Astolfi, 2017). The general ORC needs several components such as a Heat Exchanger (HE), turbine, pump, and condenser. **Figure 2** shows a general ORC flow diagram. The working fluid directly enters the HE and vaporize the working fluid. The working fluid in vapor condition with high pressure and a high temperature expands in the turbine, which produces the electrical power. The remaining heat from the expansion process is utilized in the recuperator to decrease the load in the condenser and preheater in the HE. The condenser acts as a cooler to transform the working fluid into the liquid condition. After that, the working fluid is pumped to increase the pressure. Next, the working fluid goes through the recuperator to utilize the rest of the heat and enters the preheater to increase the temperature. Later, the working fluid enters the evaporator to evaporate into vapor condition.

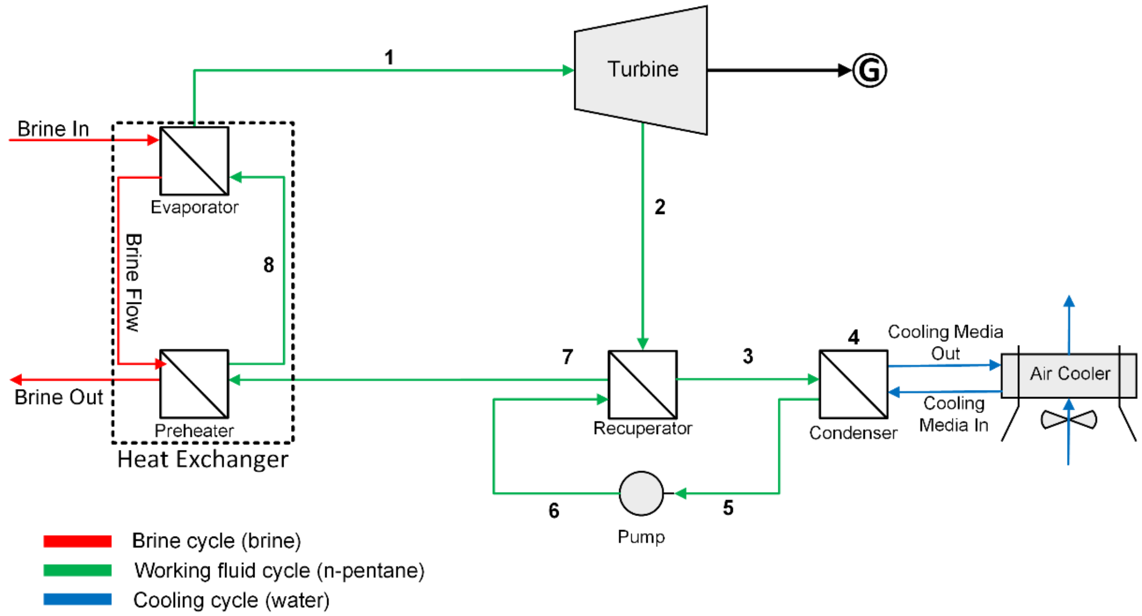

Figure 2: General ORC Flow Diagram (Wibowo et al., 2019)

The Lahendong ORC has an additional cycle called the Intermediate cycle, which makes it unique compared to the general ORC. Adding the intermediate cycle is to address the issue of high silica solubility in the ORC system (Wibowo, 2019). The brine does not evaporate the working fluid directly in the HE, as the working fluid is evaporated by the circulating hot water (not brine) from the Primary Heat Exchanger (PHE), preventing silica scaling problems in the ORC system. N-pentane is the working fluid used in the Lahendong ORC as the colorless liquid with a boiling point at 36.1 °C (Weitz and Loser, 2011). The uniqueness of the intermediate cycle is so far in Indonesia only applied in the Lahendong geothermal field. **Figure 3** shows the Lahendong ORC flow diagram.

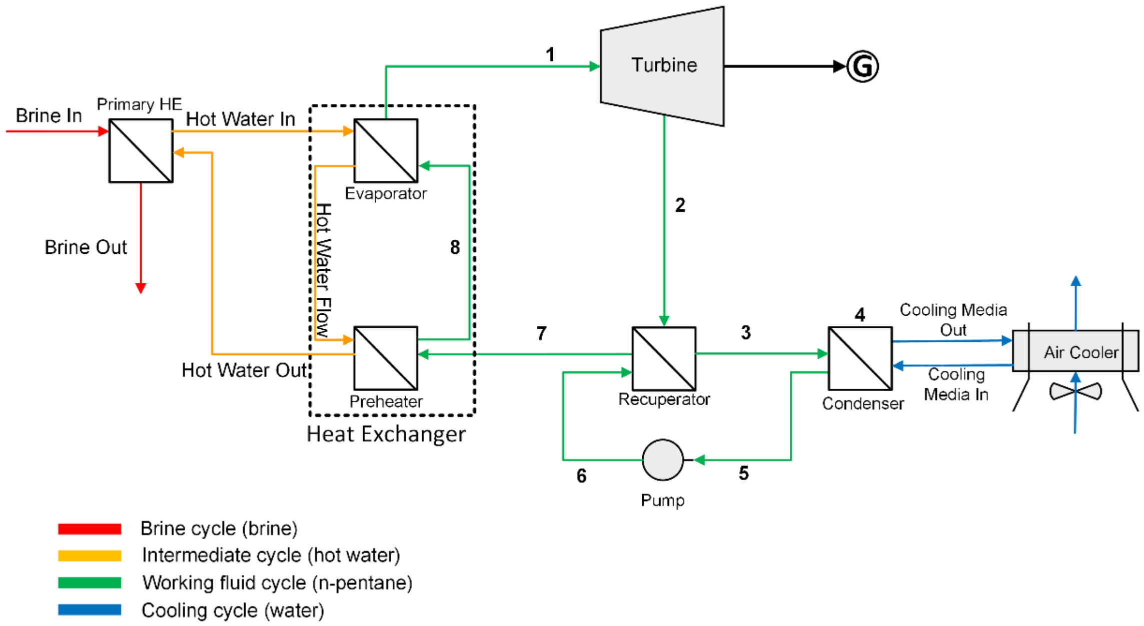

Figure 3: Lahendong ORC Flow Diagram (Wibowo et al., 2019)

1.2 Lahendong ORC Real-Time Monitoring System

The Lahendong ORC power plant applies the RTMS on its monitoring system in the remote area. The RTMS can be accessed through a Human-Machine Interface (HMI) via a web-based system, mirrored to the Smart-Micro-Grid (SMG). The SMG can be connected to several energy resources, such as a 10 kW photovoltaic (Nurdiana et al., 2018).

The RTMS provides data where the most robust variables were chosen to perform Machine Learning (ML) to predict the power plant power output. The data was from sensor readings, from the S7 and the A2000 devices. The S7 sensor reading was focused on the properties of the power plant, and A2000 focused on electrical. The sensors' data readings are transferred into a Programmable Logic Controller (PLC) as a small unit of computers that function as a local data collector from the sensors (Nugroho et al., 2011). All of the local devices connected to the communication hub act as the "heart" of communication in the power plant. The data from the hub are transferred to the server through Wireless LAN (WLAN) as the priority path and Universal Telecommunication Mobile System (UTMS) as the second priority path. The data transferred to the server with a Virtual Personal Network (VPN) and Secure Shell (SSH) for security reasons is from the connection hub and mirrored to the institution partner. After that, the server's data are transferred to the database, where data storage and visualization to HMI based on the Graphical User Interface (GUI) can be accessed using a web address. Lahendong RTMS was built based on industrial Supervisory Control and Data Acquisition (SCADA). **Figure 4** illustrates the Lahendong RTMS network architecture.

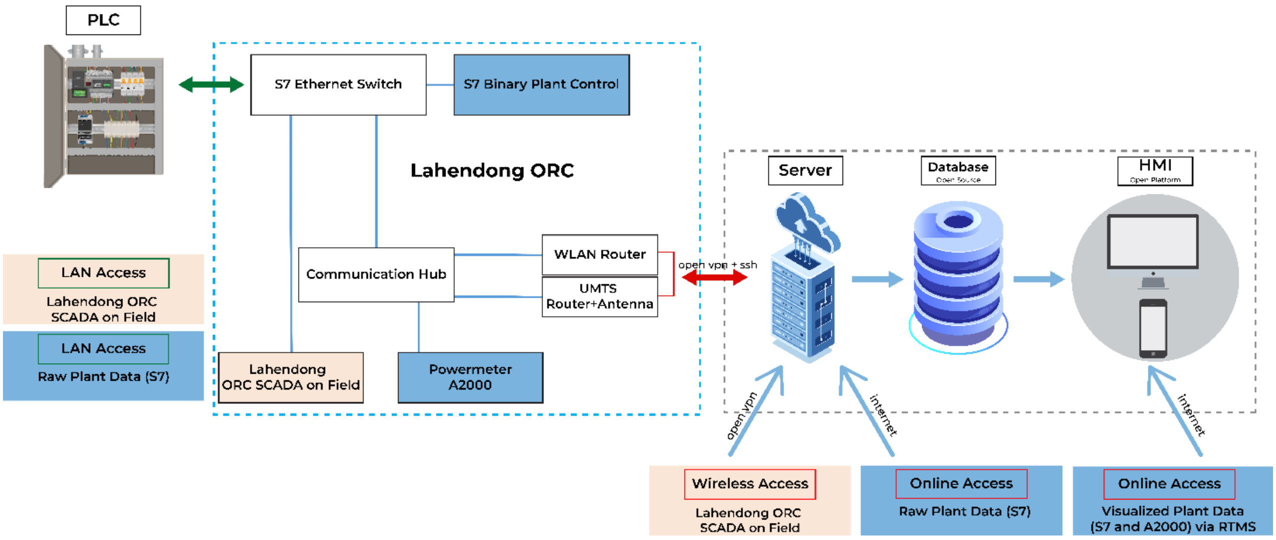

Figure 4: Lahendong RTMS Network Architecture (Wibowo et al., 2020)

The RTMS gives positive feedback to the user experience from the RTMS of the Lahendong ORC using the interactive visual data (Wibowo et al., 2020). This GUI was designed for management in the office, so the visualization is very easy to understand. **Figure 5** shows the GUI of Lahendong ORC RTMS, which contains information about the power plant. **Figure 5** provides information about brine properties input and output, hot water properties input and output, and electrical power (power output). In the figure, visualization of data in the GUI is from August until December 2019.

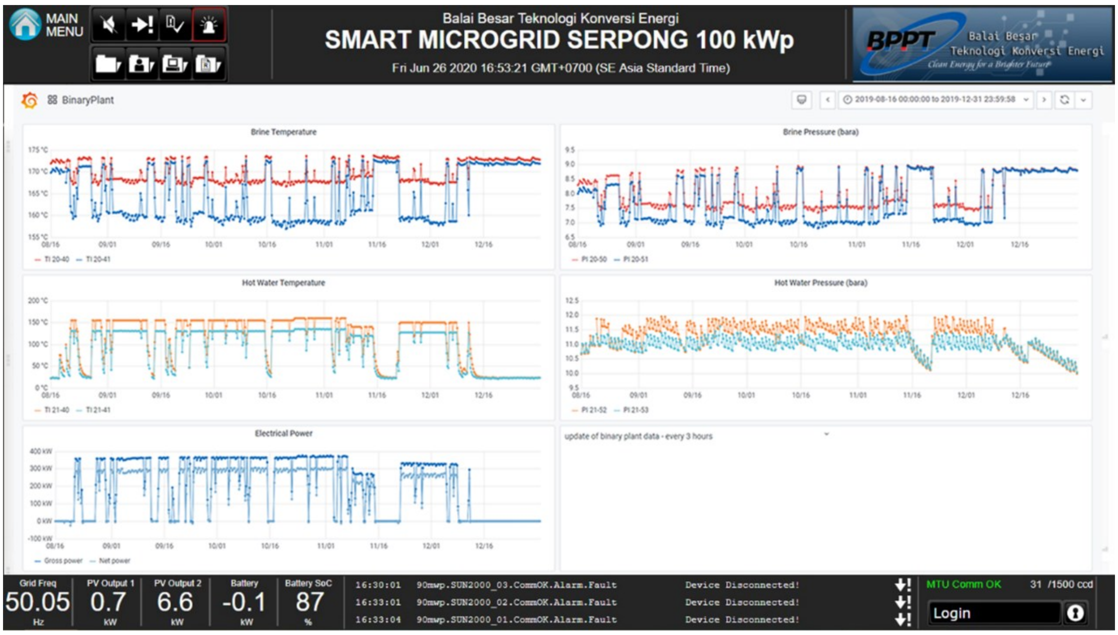

Figure 5: GUI of Lahendong ORC RTMS

Based on the sophisticated power plant and real-time access to the power plant data, ML has been developed to predict the power output of Lahendong ORC and can be used as a guide to the decision-making in the future. The prediction was created using an Artificial Neural Network (ANN), which can provide an excellent prediction result.

2. METHODOLOGY

The ANN model needs several steps to be modeled. First, the input dataset used in the model. Second, data preprocessing to handle flawed data. Next, splitting data into data training and data testing. Last, creating a ML model and validating using several methods. Illustration detail about creating the ML using the ANN model is shown in **Figure 6**.

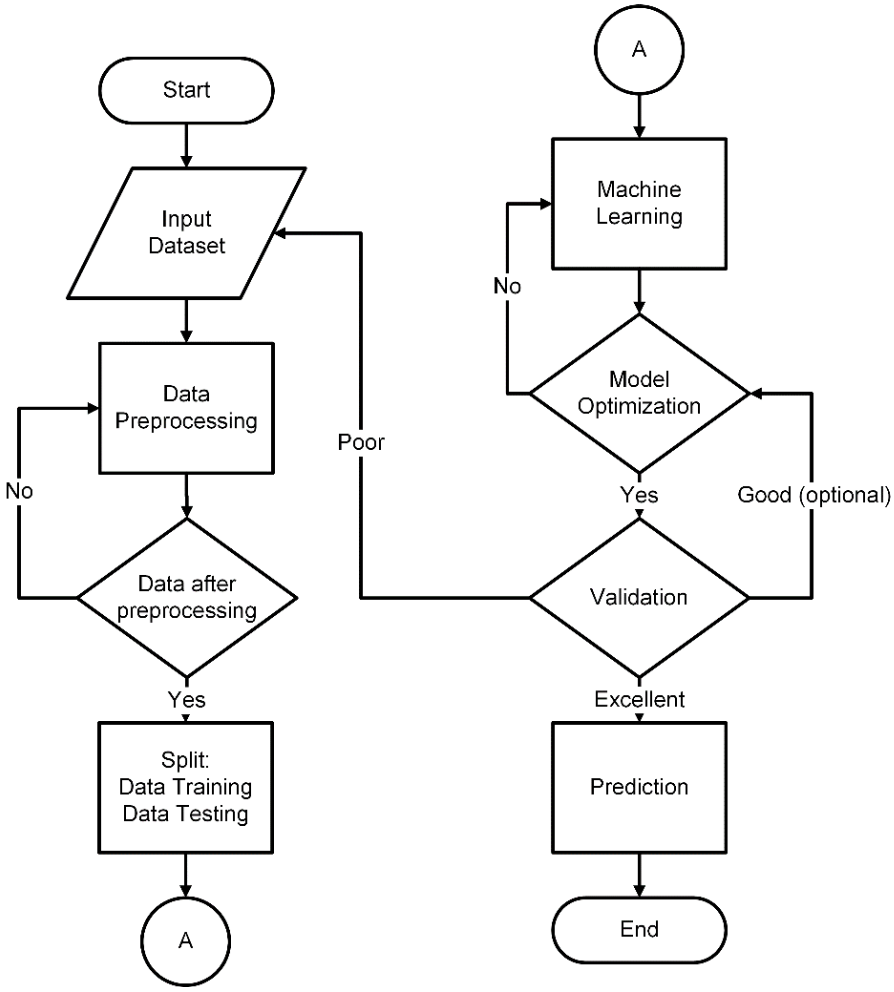

Figure 6: Flowchart to create ML

The ANN model used in this research is using Multilayer Perceptron Regressor (MLP) based on sckit-learn Python library. The multiple-perceptrons are connected in different ways and operate on unique activation functions to enable improved learning mechanisms. The training sample propagates forward through the network, and the output error is backpropagated to minimized error using a gradient-descent method, which will calculate a loss function for all the weights in the network (Swamynathan, 2019). The algorithm activation has been made using Rectified Linear Unit Function (RELU) combined with Broyden–Fletcher–Goldfarb–Shanno (LBFGS) or quasi-Newton method algorithm. By default, the MLP solver is using stochastic gradient-based optimizer (ADAM) offered by Kingma, Diederik, and Jimmy Ba. However, the reason for using RELU as a solver is because purposely for the small dataset can coverage faster and better than ADAM (Pedregosa et al., 2011).

ANNs commonly use one input layer, one output layer, and one or more hidden layers between the input and output layer. In this research, the hidden layers used in the ANN are comprised of four hidden layers with 100 neurons for the first until third hidden layers and the fourth hidden layer is using 50 neurons. The reason for using four hidden layers is because in this research the ANN improved to deep learning, where in deep learning there are usually more than one hidden layer between the input and output layers. Usually, deep learning consists of a minimum of three to four hidden layers (Sarkar et al., 2018). The input layer of ANN contains four units, which are brine temperature, brine pressure, hot water temperature, and hot water pressure. The output layer of the ANN contains one unit, which is the power output of the Lahendong ORC power plant (detail of variables in **Table 1**). **Figure 7** illustrates the work of the superficial layers of ANN.

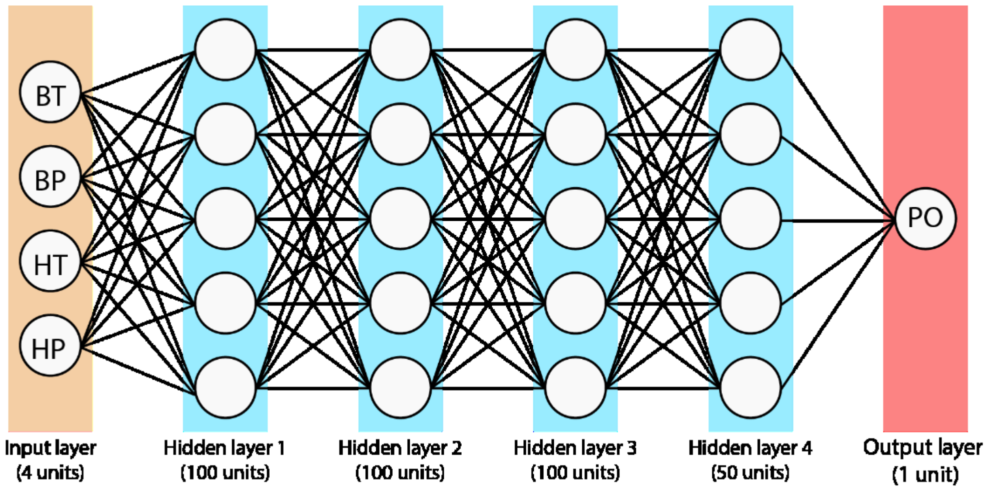

Figure 7: ANN Illustration

2.1 Input Dataset

The dataset input to the model came from the RTMS of the power plant. It was downloaded based on CSV format and input to the algorithm.

The data used was based on S7 sensor reading in Lahendong ORC RTMS in two month period from October to November 2019. Two months period has been chosen because of the minimal outliers. As the variables, the data were chosen because they are the most related to power output. The total data used in predicting ML is 1,440 data for each variable, so the total data used 7,200 data points for five variables. **Table 1** shows the data variables used in prediction. The hot water is used as a variable taking into consideration the unique cycle of the Lahendong ORC.

Table 1: Variables used in Prediction

Variables	Units	Sensor
Brine temperature inlet	°C	S7
Brine pressure inlet	bar	S7
Hot water temperature inlet	°C	S7
Hot water pressure inlet	bar	S7
Power output	kW	S7

2.2 Data Preprocessing

Commonly, preprocessing is needed before creating ML models. It is crucial to examine and preprocess a dataset before turning it into a learning algorithm (Raschka and Mirjalili, 2017). The data preprocessing has been done to deal with missing data and scaling the data to increase accuracy.

The Lahendong geothermal field location is near a heritage location and close to a rural city. In several periods, the RTMS did not perform to transfer data continuously because of some reason, such as an offline condition in the field. The offline condition commonly occurred because of lost signal, affecting the power plant's data transfer to the server. The offline condition ended with disconnected data transferred and created missing data. Handling missing data requires filling in the missing data with several actions such as mean, median, most frequent data, and replacing it with specific numbers. The average value of data was chosen to handle missing data because that shows the best correlation result between variables.

2.3 Splitting Data

When creating the ML, the dataset needs to be split into data training and data testing. The proportion of splitting was 70% for the data training and 30% for the data testing, which resulted in 1,008 for data training and 432 for data testing for each variable.

2.4 Validation

Validation compares the fit between the prediction and the actual. To make the comparison, it uses R-Squared (R^2) to check for the goodness of fit for the models (Sarkar et al., 2018). R-Squared is a measure that the actual with a prediction result using a line that indicates the variance of the result. R-Squared ranges from 0 to 1, meaning from no fit to perfect prediction (Liu, 2017).

In fitting between the prediction and the actual, several errors occurred in the result. Root Mean Squared Error (RMSE) is a correction error that indicates how close the predicted values are to the actual values (Swamynathan, 2019). One of the critical properties of RMSE is that the unit will be the same as the target variable. Lower RSME results imply excellent performance models. On the other hand, the error correction can also be done using Mean Absolute Error (MAE) which is correction based average absolute value of the errors (Swamynathan, 2019). MAE has similarities with Mean Squared Error (MSE), in that MSE uses the square of differences in actual and predictive values and MAE uses absolute differences in actual and predictive values. With two model validation methods, the result should be more accurate.

3. RESULT

The research results start with the correlations between variables using a scatter plot and the Pearson Correlation Coefficient (PCC), in which the determination variables are correctly paired with power output. This research will discuss the ML application in predicting the Lahendong ORC power plant's power output using ANN. R^2 results are close to one for a particular model with model validation error under 10% and are categorized as an excellent prediction.

3.1 Scatter Plot Analysis

A scatter plot has been done to find any correlation between variables based on the visualization, in which the linear pattern on the plot reveals that variables are related to others. **Figure 8** shows the scatter plot analysis of variables following with several abbreviations such as brine temperature (BT), brine pressure (BP), hot water temperature (HT), hot water pressure (HP), and power output (PO). The unit of temperature, pressure, and power is in Celcius (°C), bar, and kW, respectively.

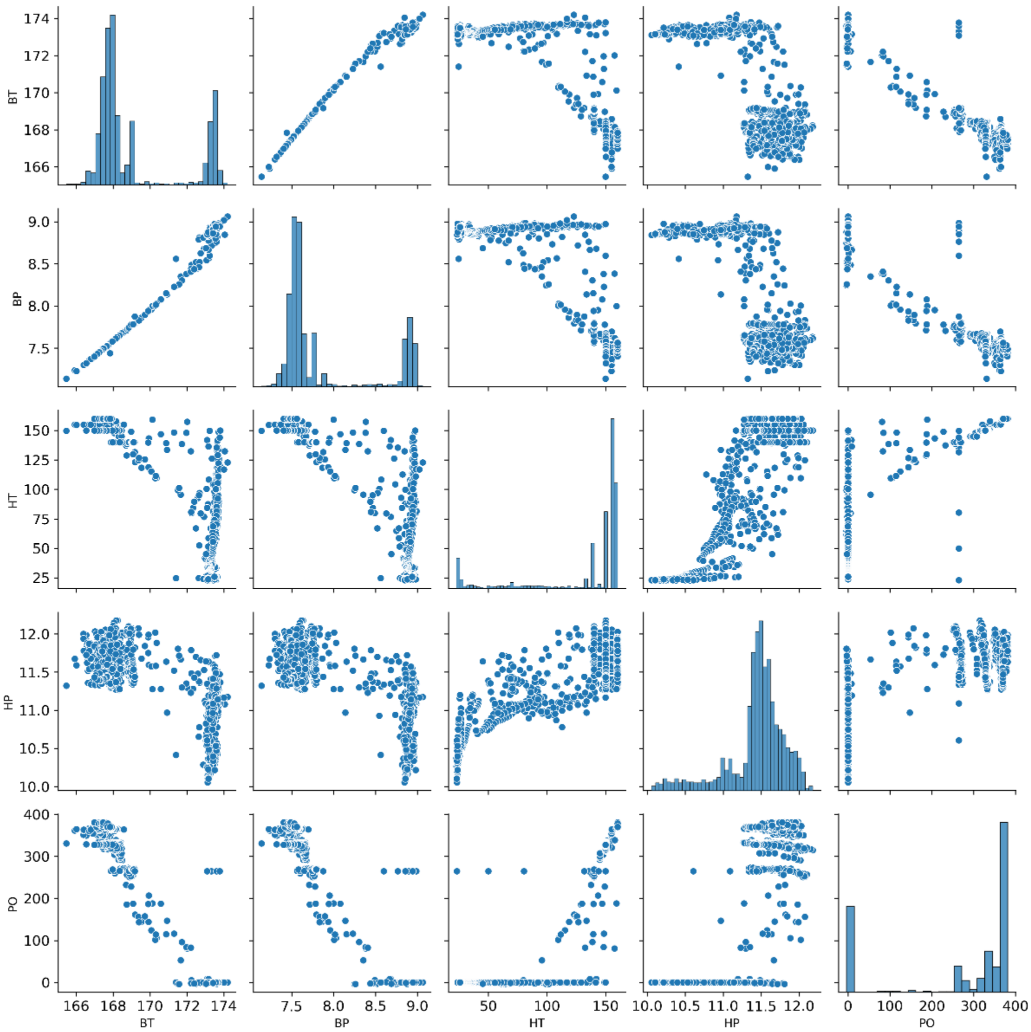

Figure 8: Scatter Plot Analysis

The plot in **Figure 8** mostly indicates a linear pattern when a variable is plotted with other variables. The linear plot pattern explains that the variable is related to others (Moore et al., 2013). It contains a constant pattern and progressive pattern. The constant pattern meas that the variable is capricious when the power plant is in operation. This is different from the progressive pattern, which shows significant changing. Looking at the power ouput PO versus PO, there is an indication of high outliers with a high zero score. The outliers can occur because of several factors, such as errors in sensor readings, lost data transfer, loss of power in the sensor, and offline condition of the system (Aji et al., 2020). Sometimes when the power plant was shut down, it still produces power output slightly in the range of 0.1 kW. Since power plant shut down still produces electricity, these points were not considered as outliers.

3.2 Pearson Correlation Coefficient

The dataset analysis continues with the PCC analysis, after the scatter plot analysis (see **Figure 8**). The analysis continues to the PCC to find the correlation of variables in the scatter plot. **Figure 9** shows the PCC analysis of variables.

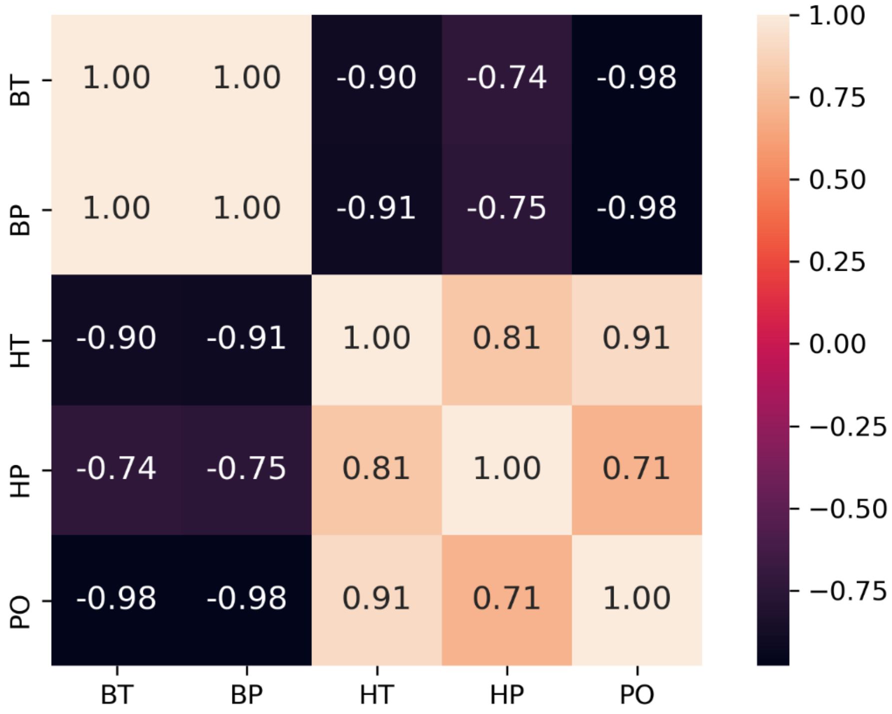

Figure 9: PPC Analysis Score

A value closer to zero means a weak correlation, and a value closer to one or minus one means a strong correlation (Moore et al., 2013). The most important point in the PCC analysis is that the brine variables have the highest correlation with power output with a score of -0.98, close to minus one. On the other side, hot water pressure has the lowest correlation with power output with a score of 0.71. The brine temperature and brine pressure have a perfect score of one occur only in the case of a perfect linear relationship when the points lie exactly along a straight line (Moore et al., 2013) (see **Figure 8**). It explains that brine temperature and pressure are related to each other due to the thermodynamics affecting the brine properties. The hot water temperature also has a high correlation value of 0.91.

Importantly, the graph shows both positive and minus correlation values. If the score is positive this means that the variable is related to others, so that if this variable increases, the others also increase. Opposite to the positive values, the negative values explain that others will decrease if the variable increases (Moore et al., 2013).

It should be noted that the hot water temperature is affected by brine properties during the convection process in the PHE (see **Figure 3**). Overall, the PCC analysis shows an excellent correlation to create ML models with an overall score higher than 0.7. If the score is under 0.3 the correlation is considered as none or very weak, if the score is between 0.3 to 0.5 the correlation is considered as weak or low effect, if the score is between 0.5 to 0.7 it considered as moderate effect, and if the score is higher than 0.7 it is considered as strong effect (Moore et al., 2013). As a note, **Figure 8** and **Figure 9** are using a dataset after preprocessing.

3.3 Machine Learning Result

The training sets contain 1008 data with four class labels for input and one class label for output. On the other side, the testing sets contain 432 data with four class labels for input and one class label for output. The output of this prediction is the power output of the Lahendong ORC power plant.

The result contains a comparison between predicted values and actual values based on statistical using R^2 score, which represents fitting between the actual and the prediction and validates the results using RMSE and MAE (Aji et al, 2018; Sarkar et al., 2018).

In the result, Support Vector Regression (SVR) and Ridge Regression (L2) also performed the ML prediction to compare with the ANN. **Table 2** shows the comparison results between the models.

Table 2: Comparison Result of ANN, SVR, and L2

Models	R^2		MAE (%)		RMSE (%)	
	Training	Testing	Training	Testing	Training	Testing
ANN_mlp	0.9954	0.9955	2.9891	4.7382	9.9104	9.9223
SVR_rbf	0.9897	0.9902	2.9433	5.9877	14.921	14.64
Ridge_L2	0.9665	0.9731	15.783	16.072	26.907	24.305

The ANN exhibits the best model result with an impressive R^2 score of 0.9954 for training and 0.9955 for testing, close to one that explains perfect prediction (Liu, 2017) and strong correlation between actual and prediction (Moore et al., 2013). The ANN validation score of the ANN can be categorized as an excellent prediction for MAE and RMSE with values under 10% (see **Table 2** for detail). Data testing has a small gap with data training of 2% in MAE validation and does not indicate overfitting or underfitting. In RMSE, the differences in data training and data testing under 1% do not indicate overfitting and underfitting.

Similarly, the SVR has R^2 score of 0.9897 for data training and 0.9902 for data testing and training, which is an excellent fit. The MAE validation also explains that the model was an excellent prediction with 2.9% for data training and 5.9% for data testing. Unfortunately, RMSE shows a result that is not as excellent result as MAE. The RMSE in the SVR model shows a result between 14%, which was a good prediction.

However, the ridge regression has not performed prediction as well as the two others with R^2 in 0.9665 and 0.9731 for data training dan testing. The validation results were categorical as a good prediction, with 15.7% and 16% for the MAE data training and data testing. On the other hand, the ridge regression RMSE validation result has the most inferior result between all of the model validations. It can only be satisfied with 26.9% and 24.3% for data training and testing, categorical as a fair prediction.

The validation result indicates that overfitting and underfitting when data training and testing have high contrast (Liu, 2017). If the data training has a better result in contrast with the data testing, it can be indicated as overfitting (low bias). Otherwise, if the data testing has a better result in contrast with data training, it can be indicated as underfitting (high bias) (Liu, 2017). Overall, the models do not show underfitting or overfitting when predicting the Lahendong ORC power output since the gap of R^2 score and validation error between training and testing is not extremely large, explaining the models were a good fit.

3.4 Model Coefficient of Determination Visualization

Since the prediction models show variable results, a coefficient of determination visualization was created to compare the model prediction quality. In the coefficient of determination, data testing, and prediction result was used to create a graph. Data testing was chosen since it was comparable to the prediction result based on several models. In this section, the ANN result only appears since the ANN was the main focus of this research. **Figure 10** shows the ANN coefficient of determination visualization between the actual value and prediction value of the power output of the Lahendong ORC. The graph shows that the outliers between actual and prediction still exist, but the outlier distance was close to the correlation line. The outliers were affecting the prediction quality by turning the correlation line. However, since the outliers are still in the range of graph values (bivariate outliers), the effect on the R2 score was not massive. The outliers not only affect the R^2 score but also the validation. Its fragmentation affects the validation with increased error but not by a massive number. As the best quality prediction, the validation scores are still under 10%.

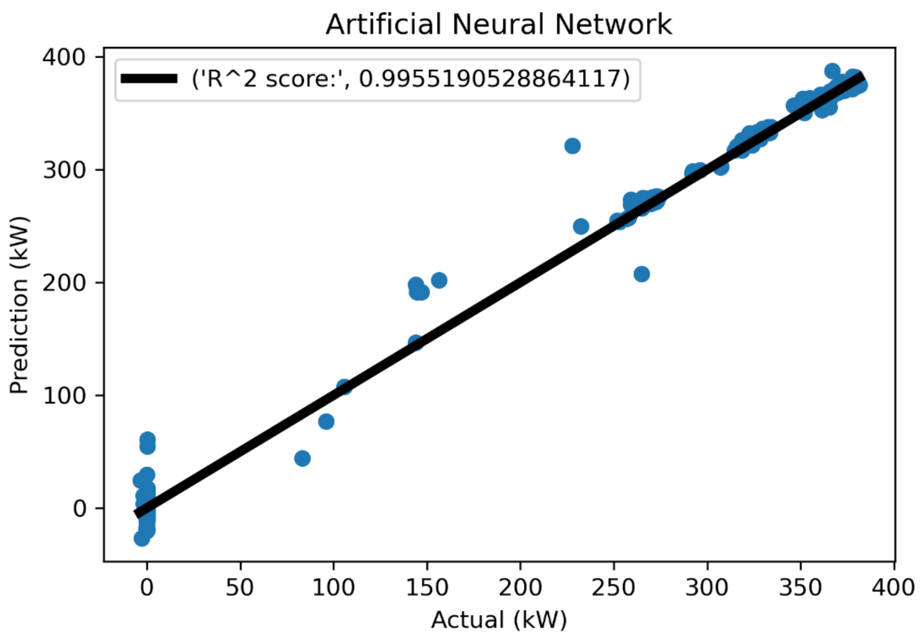

Figure 10: ANN Coefficient of Determination Visualization

3.5 Prediction Performance Evaluation

Generally, all model predictions are impressive. Nevertheless, evaluation is still needed by plotting the power output actual and prediction. This plot was created by the data testing and prediction with time-based with hour rigidity interval. **Figure 11** shows the graph of ANN performance evaluation. It contains three main plots of power output, such as actual, prediction, and residual. The residual means that were differences between actual and prediction as diagnosis regression models to detect nonlinearity and outliers (Raschka and Mirjalili, 2017). The actual data and prediction data seem to fit in **Figure 11**, which explains that prediction accuracy is excellent. It is seen that the cross symbols (prediction) and circle symbols (actual) show fit one to another. According to the ANN prediction, the power output is 262.8 ± 149.4 kW. 262.8 is the mean (μ) production of power output prediction and 149.4 is the standard deviation (σ) of power output prediction. It is explained that the prediction values have spread out over a wider range. The red square symbol (residual) values also tend to be constant as a line, explaining the slight difference in the actual and the prediction result. The graph still has a few residual outliers with a maximum value of residual outlier of 93.4 kW, which is the explanation of the error in RMSE and MAE. As the best result model, the ANN still contains a few outliers of the residual. A red square can be seen in the graph off from the residual linear pattern. However, the ANN is the model with lowest result in regard to residual fragmentation compared to the others.

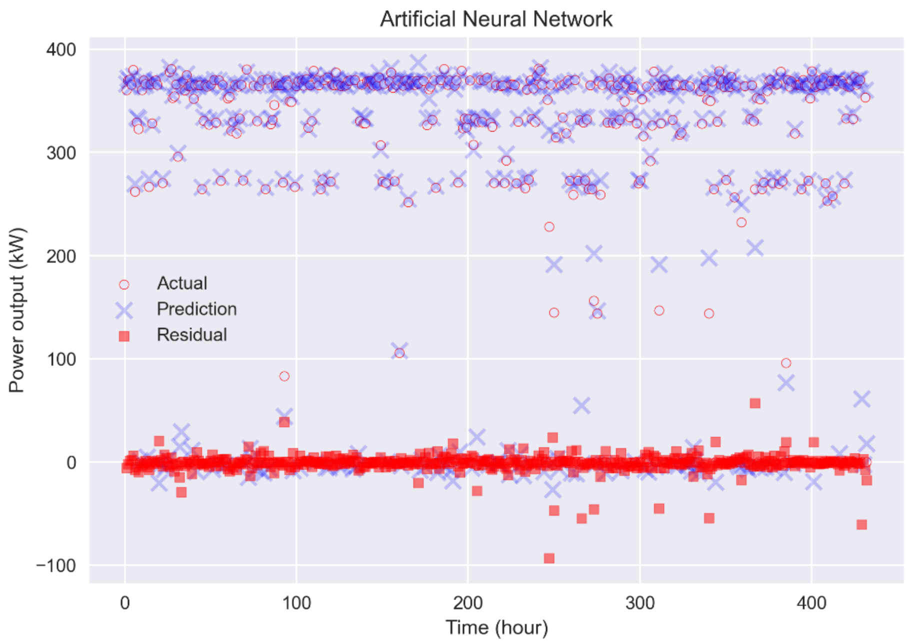

Figure 11: ANN Performance Evaluation

4. DISCUSSION

The outliers in the dataset occur because of the offline condition. It has happened because of the difficulties of the GSM signal to transfer the data and external factors.

The power output shows that data tends to zero several times. However, this is not the outliers but because the power plant was shutdown and still producing electricity between 0.1 kW to 1 kW.

The preprocessing is not only about handling missing data and scaling data., but also it creates the label class and partitioning a dataset into training and testing sets. Before handle missing data, the label class was created manually since the data was separately downloaded from the RTMS.

October and November 2019 were chosen as a prediction period because of the best data condition compared to other months. In the middle of December 2019, the power plant was completely shut down until early February 2020 because of the replacement of hot water cycle coupling. Since the coupling was hard to find, the power plant did not operate. After February 2020, the RTMS provideds poor data conditions because of GSM signal problems.

Several models have a better result without StandardScaler. However, StandardScaler is needed in preprocessing for ML algorithms such as L2 and SVR with RBF.

The parameters are used based on the stochastic method optimization, forum discussion, and GridSearchCV as parameter estimators. Since this research focused on ML application in the geothermal industry, the determination parameters are not discussed in detail.

PCC analysis has a different result after the data has preprocessing using SimpleImputer and before. The highest correlation before preprocessing was hot water temperature with a score of 0.91. Brine temperature only had a score of 0.13 and brine pressure had a score of -0.24. This revealed that the preprocessing to handle missing data was necessary.

SimpleImputer contains three main parameters to handle missing values consist of mean, median, and most frequent. The most frequent parameter has worked better in ANN that is not in others. Determination using mean or average value has a better result overall.

Without handling a missing value, the ANN score is 0.97, with an error above 10% for RMSE and MAE. It means that handling missing values is necessary to increase the prediction model.

SVR has a faster process in prediction compared to the ANN, as ANN more complex than SVR. SVR takes a time of only a few second while ANN takes a half minute.

Long-term future predictions actually will be made using an autoregressive integrated moving average (ARIMA). However, the long-term prediction still under review considering several aspects: the data consist of missing values, sensor performance decreases in the future and the prediction proper in the short-term.

5. CONCLUSION

The ANN prediction gives an excellent result with a score of 0.9955 very close to one with an excellent validation result under 10% for both RMSE and MAE following by SVR and L2. The correlation of variables shows a strong effect between variables especially the power output with a correlation score over 0.7. The brine properties correlation has the best correlation score with a perfect score of one since the points lie exactly along a straight line as the perfect linear and the highest score to the correlation with power output.

Hopefully, this research can inspire ML used in geothermal from the front to the end. The research can reference the expertise in the decision-making of the Lahendong ORC in the future. The RTMS server should be moved from Germany to Indonesia to create an online short-term prediction feature in the RTMS. The long-term prediction (forecasting) using time series also will be made after measuring sensor performance decreases for several years based on ARIMA. Increasing variables in the dataset could improve prediction accuracy. However, the study is still in progress and in the trial process for further research. Improvement in the data transfer started with fixing signal issues since most outliers occurred because of signal issues. Further study about data transfer should be taken because of the RTMS has been off for data visualization since June 2020.

ACKNOWLEDGEMENT

We gratitude to the Agency for the Assessment and Application of Technology (BPPT) in collaboration with Trisakti University for allowing us to present this paper. This paper is based on collaboration work between two institutions.

REFERENCES

Aji, P., K. Wakamori, and H. Mineo. 2018. "Highly Accurate Daily Solar Radiation Forecasting using SW-SVR for Hybrid Power Plant in Indonesia." IEEE.

Aji, P., K. Wakamori, and H. Mineo. 2020. "Short-Term Solar Power Forecasting Using SVR on Hybrid PV Power Plant in Indonesia." Chap. 23 in *Advances in Intelligent Networking and Collaborative Systems*, by L. Barolli et al, 235-246. Springer, Charm. doi:https://doi.org/10.1007/978-3-030-29035-1_23.

Astolfi, M. 2017. "Technical options for Organic Rankine Cycle systems." In *Organic Rankine Cycle (ORC) Power Systems Technologies and Applications*, by M. Astolfi, E. Macchi, C. Spadacini, L. G. Xodo and M. Quaia. Duxford: Elsevier Ltd.

Liu, Y. 2017. *Python Machine Learning By Example.* Birmingham: Packt Publishing Ltd.

Macchi, E. 2017. "Theoretical basis of the Organic Rankine Cycle." In *Organic Rankine Cycle (ORC) Power Systems Technologies and Applications*, by M. Astolfi, E. Macchi, C. Spadacini, L. G. Xodo and M. Quaia, edited by E. Macchi and M. Astolfi, 1-24. Duxford: Elsevier Ltd.

Moore, D. S., William I. Notz, and Michael Fligner Fligner. 2013. *The Basic Practice of Statistics.* 6th. New York: W. H. Freeman and Company.

Nugroho, T. H., Riza, N. A. Sasongko, and T. Lubis. 2011. "Sistem SCADA Untuk Pengoperasian PLTP Skala Kecil 3 MW." *The 11TH ANNUAL INDONESIAN GEOTHERMAL ASSOCIATION MEETING & CONFERENCE.* Bandar Lampung: INAGA.

Nurdiana, E., H. Hilal, Riza, N. A. Aryo, and A. Prastawa. 2018. "Sistem PLTS Rooftop 10 kWp Berbasis Smart Grid Untuk ImplementasiI Demand Response." *Simposium Nasional RAPI XVII.* Solo: FT UMS. 23-30.

Pedregosa, F., G. Varoquaux, A. Gramfort, V. Michel, O. Grisel, M. Blondel, P. Prettenhofer, et al. 2011. "Scikit-learn: Machine Learning in Python." *Journal of Machine Learning Research 12* 2825-2830.

Raschka, S., and V. Mirjalili. 2017. *Python Machine Learning.* 2nd. Birmingham: Packt Publishing Ltd.

Sarkar, D., R. Bali, and T. Sharma. 2018. *Practical Machine Learning with Python.* New York: Springer Science+Business Media New York.

Suharmanto, Puji , A. N. Fitria, and S. Ghaliyah. 2015. "Indonesian Geothermal Energy Potential as Source of Alternative." *Renewable Energy and Energy Conversion Conference and Exhibition.* KnowledgeE Publishing Services. 119-124. doi:http://dx.doi.org/10.18502/ken.v1i1.325.

Swamynathan, M. . 2019. *Mastering Machine Learning with Python in Six Steps.* 2nd. Bangalore: Apress.

Wahjosoebidjo, A. S., and M. Hasan. 2018. "Indonesia’s Geothermal Development: Where is it Going?" *43rd Workshop on Geothermal Reservoir Engineering.* Stanford: Stanford University.

Weitz, H. M., and E. Loser. 2011. "Isoprene." In *Ullmann's Encyclopedia of Industrial Chemistry*, by Wiley-VCH, Dieter Hönicke, Ringo Födisch, Peter Claus, Michael Olson, Eckhard Loser and Hans Martin Weitz, 13329-13348. Weinheim: Wiley-VCH.

Wibowo, S. N., K. Pudyastuti, Suyanto, L. Agustina, Harmen, and S. Patangke. 2019. "Review Application of ORC in Lahendong Geothermal Field." *Paper presented at the SPE/IATMI Asia Pacific Oil & Gas Conference and Exhibition.* Bali: SPE. doi:https://doi.org/10.2118/196278-MS.

Wibowo, S. N., P. Aji, Cahyadi, Suyanto, and K. Pudyastuti. 2020. "Development of Real-time Monitoring System in Organic Rankine Cycle Operation of Lahendong Geothermal Field." *SPE ANNUAL TECHNICAL CONFERENCE AND EXHIBITION 2020.* Denver: SPE.