Prediction of Electricity Production from PLTMH Using Linear Regression and LSTM

Case Study on at PLTMH Sangir Hulu

Sandhi Ading Wasana - 23522308
Program Studi Magister Informatika
Sekolah Teknik Elektro dan Informatika
Institut Teknologi Bandung, Jalan Ganesha 10 Bandung
E-mail (gmail): adingism@gmail.com

***Abstract*— The Government of Indonesia has set a target for hydropower plant capacity addition of 14.3 GW in the RUPTL 2016-2025[1]. However, the absorption of energy from Micro-Hydro Power Plants (PLTMH) faces many challenges in predicting electricity production due to numerous influencing factors. The fluctuating electricity production poses a new problem in planning the optimal absorption of renewable energy.**

Several theories can be employed to address this issue, including time series models such as Long Short Term Memory (LSTM), linear regression models, and other algorithms like Support Vector Machines (SVM) or Random Forest. Research by Chi 2022[2] using LSTM to predict electricity consumption achieved very high accuracy. Unlike consumption, electricity production data from PLTMH has its own challenges due to its significant fluctuations. Studies by Prakash et al., 2022 [3] and Adiwarman et al., 2023 [4] have developed models for predicting electricity production from water using various methods, including LSTM and linear regression.

The proposed methods to address this issue are the use of LSTM and linear regression models. These methods can utilize historical datasets of electricity production from the Sangir Hulu PLTMH to make accurate predictions. This research is titled "Prediction of Electricity Production from PLTMH Using Linear Regression and LSTM."

Keywords: PLTMH, LSTM, Linear Regression, Renewable Energy

I. INTRODUCTION

Currently, PT PLN (Persero) is challenged to address global demands by creating a power plant utilizing renewable energy (RE) to mitigate global warming. As the demand for electrical energy increases, PT PLN (Persero) is obligated to meet this demand using various more environmentally friendly alternative energy sources. However, the use of RE as an energy source faces numerous challenges. Several factors hinder the absorption of electrical energy from RE, including natural factors such as weather. For instance, hydroelectric power plants heavily depend on the water flow rate into rivers or lakes that serve as the intake for such plants. Various efforts are made to improve efficiency in absorbing energy from water. One way to enhance the absorption of clean energy from water is by predicting the electricity production that will be generated. [3].

In the study by Adiwarman et al. (2023) [4], electricity production predictions were developed using data collected through IoT technology installed in the area around the Microhydro Power Plant (MHP) in Pagentan, Banjarnegara Regency, Central Java. The prediction model was built using Linear Regression, utilizing parameters recorded by the installed sensors. The recorded parameters included rainfall data, humidity, temperature, and water level. The model's accuracy, tested previously, resulted in an R^2 score of 0.75076. This accuracy is significantly better than that achieved in the study by Prakash et al. (2022) [3].

This study focuses on developing a prediction model using LSTM and Linear Regression with a case study at the Sangir Hulu MHP. In the previous discussion, the LSTM model showed excellent accuracy scores; however, for electricity production prediction, the Linear Regression model demonstrated better accuracy. This research aims to develop an electricity production prediction model with better accuracy scores than previous studies.

II. RELATED WORK

A. LSTM

LSTM can handle common issues encountered when processing time series data with long data sequences during training. The structure of the LSTM neural network is shown in Figure II.1. In the LSTM structure, some information is discarded if it is not needed. LSTM consists of a Sigmoid layer, where the forget gate determines which information will be discarded. This way, LSTM can retain important information for a long time, and the memory can be dynamically adjusted based on the input data. [5].

B. Linear Regression

Linear regression is a widely used technique for predicting trends. In the case study of electricity production prediction, multiple linear regression was employed. Linear regression uses lagged data, leveraging several previous time steps to predict one step ahead.

In the equation above, Y is the dependent variable, and a is the constant. The variables b1, b2, and bn are weights or regression coefficients that determine how much the dependent variable changes concerning the independent variables X1, X2, and Xn. X1, X2, and Xn are the independent variables that serve as the model's input. [4]

C. Evaluation Metric

In evaluating the performance of a predictive model, evaluation indices can be used to measure the accuracy of the predictions. The accuracy scores can be compared with the performance of other models to identify which model produces the highest accuracy. This study employs several evaluation procedures to assess accuracy as follows.

Root Mean Squared Error (RMSE) is a metric that measures the deviation between predicted values and actual values. The formula for RMSE is the average of the squared differences between predicted and actual values. The smaller the RMSE value, the better the predictions produced by the model. [6]

The Coefficient of Determination, or R^2, is a metric that measures how much the independent variables influence the dependent variable. This formula results in a value ranging from 0 to 1. The closer the value is to 1, the better the predictions. [2]

III. Methodology

In this research, the methodology used includes data collection, data preprocessing, modeling, model training, and evaluation. The following is an outline of the methodological steps taken.

A. Data Collection

The data used in this research is the electricity production dataset from a Microhydro Power Plant (PLTMH) in West Sumatra for the years 2019-2023. This dataset records production history every 30 minutes. The recorded values are in MegaWatts (MW).

The electricity production dataset from the PLTMH is recorded daily, resulting in 1,878 days of data from 2019 to 2023. Therefore, 1,878 datasets need to be combined and preprocessed into a single dataset..

B. Data Preparation

In the preprocessing stage, several steps are performed, Format Correction: This step involves correcting the data format so that it can be automatically combined into a single file.

Data Cleansing: Removing noise from the data that could cause inaccuracies in model training, as well as eliminating NULL values and other data anomalies.

The data is then split into training and testing sets with a ratio of 70% for training and 30% for testing. From the 70% training data, a validation set is further split to be used as part of the model training process.

C. Model Building

In this stage, a predictive model for Microhydro Power Plant (PLTMH) electricity production will be built. The algorithms to be used are LSTM and Linear Regression. Several parameters will be adjusted to achieve the target accuracy.

D. Data Training

In this process, the preprocessed dataset will be trained using the same training data and timeframe. This ensures balanced evaluation metrics that can be compared effectively

E. Model Evaluation

After the model is built, evaluation is necessary to test the prediction quality of the model. The evaluation scores will be compared with the predictive models described in the previous chapter. Several evaluation metrics that will be used are RMSE and R^2.

IV. Discussion And Result

After the text edit has been completed, the paper is ready for the template. Duplicate the template file by using the Save As command, and use the naming convention prescribed by your conference for the name of your paper. In this newly created file, highlight all of the contents and import your prepared text file. You are now ready to style your paper; use the scroll down window on the left of the MS Word Formatting toolbar.

A. Discussion

In this research, the preprocessed dataset is processed into 2 models that have been prepared, namely the LSTM model and Linear Regression. The dataset is separated into two types, namely datasets with 30-minute intervals and datasets with 60-minute intervals. Parameters in LSTM use the Relu activation function and data lag of 12. The tunning parameters used are epoch 25 and 50. Then LSTM units are 64 and 128. In Linear Regression, the tunning parameter used is fit intercept which is TRUE and FALSE. Then all models are measured using the RMSE and R2 evaluation metrics.

B. Resiult

Tests that have been carried out using LSTM and Linear Regression with hyperparameter tunning that has been described in the previous chapter produce the following accuracy values.

No	Interval Data	Epoch	LSTM Units	Activation	WinSize	RMSE	R2
1	30 Menit	25	64	Relu	12	0.6	0.776
2	30 Menit	50	64	Relu	12	0.59	0.787
3	30 Menit	25	128	Relu	12	0.6	0.777
4	30 Menit	50	128	Relu	12	0.6	0.778
5	60 Menit	25	64	Relu	12	0.68	0.667
6	60 Menit	50	64	Relu	12	0.66	0.68
7	60 Menit	25	128	Relu	12	0.67	0.677
8	60 Menit	50	128	Relu	12	0.66	0.689

a)

No	Interval Data	fit_intercept	WinSize	RMSE	R2
1	30 Menit	TRUE	12	0.618	0.731
2	30 Menit	FALSE	12	0.627	0.723
3	60 Menit	TRUE	12	0.751	0.541
4	60 Menit	FALSE	12	0.762	0.527

b)

Table IV.1

From the test results it is known that the model that produces the best value on data with a range of 30 minutes is LSTM by using LSTM units of 64 and epoch as much as 50. In the 60 minute range the best model produced is also LSTM with LSTM units of 128 and epoch 50. Overall the best model produced is a dataset with a range of every 30 minutes with LSTM units of 64 and epoch 50. This model produces an RMSE value of 0.59 and R2 of 0.787. The results of the R2 evaluation metric show better results than research conducted by Prakash et al., (2022) [3] and Adiwarman et al., (2023). [4]

No	Penelitian	Model Prediksi	R2
1	Model LSTM yang dibangun	LSTM	0.787
2	Penelitian Prakash, 2022	LSTM	0.566
3	Penelitian Adiwarman, 2023	Regresi Linear	0.751

Table IV.2

The prediction model produced in this study shows better results than previous studies. However, the above comparison uses different datasets, and different modeling

V. Conclusion

The higher level of air pollution, the higher global warming, the alternative utilization of electricity generation from more environmentally friendly energy needs to be done immediately. The development of electricity production prediction models needs to be continuously developed so as to produce better prediction models. Prediction models with high accuracy can answer the challenges of absorbing new renewable energy which is fluctuating in increase and decrease in production.

In this study, higher results were obtained than the two previous studies related to the prediction of power plants that utilize energy from water. This study shows the accuracy value measured using RMSE of 0.59 and R2 value of 0.787.

Acknowledgment

Firstly, the author is an employee of PT PLN (Persero), in the UP2D Sumatera Barat. Currently, the author is pursuing a master's degree in the Faculty of Electrical Engineering and Informatics, Bandung Institute of Technology. PT PLN (Persero) has always supported this research.

References

[1] L. N. Rahayu and J. Windarta, “Tinjauan Potensi dan Kebijakan Pengembangan PLTA dan PLTMH di Indonesia,” *Jurnal Energi Baru dan Terbarukan*, vol. 3, no. 2, pp. 88–98, Jun. 2022, doi: 10.14710/jebt.2022.13327.

[2] D. Chi, “Research on electricity consumption forecasting model based on wavelet transform and multi-layer LSTM model,” *Energy Reports*, vol. 8, pp. 220–228, Jul. 2022, doi: 10.1016/j.egyr.2022.01.169.

[3] S. A. Prakash, D. Shah, K. Jayavel, and K. Mtonga, “Hydropower Energy Generation Prediction Model: A Machine Learning Approch,” in *2022 International Conference on Computer Communication and Informatics, ICCCI 2022*, Institute of Electrical and Electronics Engineers Inc., 2022. doi: 10.1109/ICCCI54379.2022.9740915.

[4] A. S. Adiwarman, R. R. Pahlevi, and H. H. Nuha, “Microhydro Power Plants with IoT-based Monitoring and Prediction using Linear Regression,” in *2023 International Conference on Artificial Intelligence, Blockchain, Cloud Computing, and Data Analytics, ICoABCD 2023*, Institute of Electrical and Electronics Engineers Inc., 2023, pp. 7–12. doi: 10.1109/ICoABCD59879.2023.10390935.

[5] R. Cao, H. Ye, Z. Fang, L. Chen, and S. Zhang, “A Regional Monthly Electricity Consumption Forecast Method Based on HP Filter and Seq2Seq,” in *2023 IEEE 6th International Conference on Automation, Electronics and Electrical Engineering, AUTEEE 2023*, Institute of Electrical and Electronics Engineers Inc., 2023, pp. 515–521. doi: 10.1109/AUTEEE60196.2023.10407947.

[6] J. Gaboitaolelwe, A. M. Zungeru, A. Yahya, and C. K. Lebekwe, “Electricity Load Prediction Using Machine Learning,” in *2022 International Conference on Smart Applications, Communications and Networking, SmartNets 2022*, Institute of Electrical and Electronics Engineers Inc., 2022. doi: 10.1109/SmartNets55823.2022.9993990.

PERNYATAAN

Dengan ini saya menyatakan bahwa makalah yang saya tulis ini adalah tulisan saya sendiri, bukan saduran, atau terjemahan dari makalah orang lain, dan bukan plagiasi.

Bandung, 12 Juni 2024

Ttd

Sandhi Ading Wasana dan 23522308