COMPARISON OF LINEAR REGRESSIONS AND NEURAL NETWORKS FOR FORECASTING ELECTRICITY CONSUMPTION

Tyas Setiyorini[1*]; Frieyadie[2]

Teknik Informatika
STMIK Nusa Mandiri Jakarta
http://nusamandiri.ac.id
[1]tyas.setiyorini@gmail.com; [2]frieyadie@nusamandiri.ac.id
(*) Corresponding Author

Abstract—Electricity has a major role in humans that is very necessary for daily life. Forecasting of electricity consumption can guide the government's strategy for the use and development of energy in the future. But the complex and non-linear electricity consumption dataset is a challenge. Traditional time series models in such as linear regression are unable to solve nonlinear and complex data-related problems. While neural networks can overcome the problems of nonlinear and complex data relationships. This was proven in the experiments in this study. Experiments carried out with linear regressions and neural networks on the electricity consumption dataset A and the electricity consumption dataset B. Then the RMSE results are compared on the linear regressions and neural networks of the two datasets. On the electricity consumption dataset, A obtained by RMSE of 0.032 used the linear regression, and RMSE of 0.015 used the neural network. On the electricity consumption, dataset B obtained by RMSE of 0.488 used the linear regression, and RMSE of 0.466 used the neural network. The use of neural networks shows a smaller RMSE value compared to the use of linear regressions. This shows that the problem that is proven in the electricity consumption dataset A and the electricity consumption dataset B is that it has a nonlinear relationship, it can be overcome by the neural networks. So that neural networks can improve performance better than linear regressions.

Keywords: Linear Regressions, Neural Networks, Electricity Consumption

Abstrak— *Listrik memiliki peran utama bagi manusia yang sangat diperlukan dalam kehidupan sehari-hari. Peramalan konsumsi listrik dapat memandu strategi pemerintah untuk penggunaan dan pengembangan energi di masa depan. Tetapi dataset konsumsi listrik yang kompleks dan non-linear merupakan tantangan. Model deret waktu tradisional seperti linear regression tidak dapat menyelesaikan masalah hubungan data yang kompleks dan nonlinier. Sebaliknya, neural network dapat mengatasi masalah hubungan data yang kompleks dan nonlinier. Ini terbukti dalam percobaan dalam penelitian ini. Eksperimen telah dilakukan dengan linear regression dan neural network pada dataset konsumsi listrik A dan dataset konsumsi listrik B. Kemudian hasil RMSE dibandingkan dengan metode linear regression dan neural network pada kedua dataset. Pada dataset konsumsi listrik A diperoleh RMSE 0,032 menggunakan linear regression, dan RMSE 0,015 menggunakan neural network. Pada dataset konsumsi listrik B diperoleh RMSE sebesar 0,488 menggunakan linear regression, dan RMSE sebesar 0,466 menggunakan neural network. Penggunaan neural network menunjukkan nilai RMSE yang lebih kecil dibandingkan dengan penggunaan linear regression. Hal ini menunjukkan bahwa masalah yang terbukti pada dataset konsumsi listrik A dan dataset konsumsi listrik B yaitu memiliki hubungan nonlinear, dapat diatasi oleh neural network. Sehingga neural network dapat meningkatkan kinerja lebih baik dibanding linear regression.*

***Kata Kunci:** Linear Regression, Neural Network, Konsumsi Listrik*

INTRODUCTION

Electricity is a major human need nowadays. Electric power is an indispensable role in daily life because compared to traditional energy is more environmentally friendly (Nyberg & Saikkonen, 2014). At present, the rapidly increasing human population and technological developments are causing a sharp increase in power consumption (Zhao et al., 2017)(Kim & Cho, 2019). Therefore, accurate forecasting of electricity consumption can guide the government's strategy for the use and development of energy in the future (He et al., 2019).

Nonlinear problems, high frequency, and high volatility in the application features of electricity consumption in our society is a challenge in forecasting electricity consumption

(Islyaev & Date, 2015). Complex, non-linear, and non-stationary electricity price features negatively affect the generation of models. (Yang et al., 2020).

Electricity consumption prediction is a type of time series that depends on the pattern of time dependence (Zheng et al., 2019). Time series is a series of sequential data that is measured from time to time, like hourly, daily, or weekly peak loads (Dodamani et al., 2015). Time series is the most commonly used method for forecasting electricity consumption (Sadownik & Barbosa, 1999).

The most common approach to modeling of building energy consumption is by regression analysis (Satre-Meloy, 2019). Time series algorithms have generally been applied in electricity price forecasting or other time series forecasting (Yang et al., 2020). Regression models are included in the traditional time series methods (Yan et al., 2019) such as linear regression that is not suitable for predicting nonlinear and complex models (Satre-Meloy, 2019). The size and direction of the relationship between predictive factors and electricity consumption are determined by linear regression (Satre-Meloy, 2019). The linear regression model presents meager focuses, just as for the most part lower anticipated qualities particularly for the chiller, indicating low linearity among anticipated and genuine utilization (Pombeiro et al., 2017). This complexity nature makes it hard to comprehend the connections between data inputs and the reaction. In general, for short-term load forecasting, the use of the traditional method like the statistical model is the linear regression method which is a linear model, which suffers from nonlinearity and only provides reasonable accuracy (Lee & Ko, 2009).

Compared to traditional methods, Artificial Intelligence (AI) is developing fast, providing short-term forecasting solutions that are important for time series (Yan et al., 2019). The development of modern artificial intelligence methods is rapid, introducing artificial neural networks and population evolution algorithms for electric forecasting (Taylor, 2010). Neural networks have become popular in terms of nonlinearity in all fields of engineering including load forecasting, and are overcome functional dependence on forecasting models (Ferreira & Alves da Silva, 2007). Various variants of artificial neural networks were implemented to model complex and nonlinear relationships between features used for forecasting and achieved high accuracies (Agrawal et al., 2019). The non-linear relationship model in the complex structure of electricity demand allows being overcome by ANN (Nadtoka & Al-Zihery Balasim, 2015).

Based on previous studies stated the electricity consumption dataset has a nonlinear and complex relationship. Traditional time series models such as linear regression are unable to solve nonlinear and complex data-related problems. Otherwise, neural networks can overcome the problem of nonlinear and complex data relationships. Therefore, this study wants to prove that there is a nonlinear relationship in the electricity consumption dataset used in this study, and compare which performance is better between using linear regression and neural networks.

MATERIALS AND METHODS

Data

In this study used 2 datasets are electricity consumption. This dataset explains the use of electric power in daily life in kW or kWh units. Electricity consumption dataset A based on date and time and electricity consumption dataset B based on date. This dataset is obtained from the kaggle.com.

The electricity consumption dataset A shown in Table 1 consists of 5 attributes predictor are serial, kW, kVARh, kVAR, and time stamp, and 1 attribute class is kWh. The electricity consumption dataset B shown in Table 2 consists of 6 attributes predictor are reactive power, voltage, kitchen, laundry, HVAC, and attribute class is an active power.

Table 1 Attributes and Descriptions on the Electricity Consumption A

No	Attributes	Description
1	Serial	Id transaction
2	kWh	Working power in an hour
3	kW	Working power
4	kVARh	Reactive power
5	kVAR	Reactive power in an hour
6	Date	Date
7	Time	Time

Source: (Zahid, 2020)

Table 2 Attributes and Descriptions on the Electricity Consumption B

No	Attributes	Description
1	Date	Date
2	Active Power	Working power in an hour
3	Reactive Power	Reactive power
4	Voltage	Voltage
5	Kitchen	Kitchen
6	Laundry	Laundry
7	HVAC	HVAC

Source: (Ashfaque, 2020)

After the linearity test shows that in Table 3 the electricity consumption dataset A shows a significant value (sig.) on the KW, kVARh, kVAR, date is 0.000. This shows that 0,000 <0.05, so that the data has a nonlinear relationship. Likewise, the linearity test shows that in Table 4 that the electricity consumption dataset B shows a significant value (sig.) on the date, reactive power, voltage, kitchen, laundry, and HVAC is 0,000. This shows that 0,000 <0.05, so that the data has a nonlinear relationship. It can be concluded that the two electricity consumption datasets show data that have nonlinear relationships.

Table 3. Significant Value of Linearity Test on the Electricity Consumption Dataset A

Model	Sig.
(Constant)	.000
kW	.000
kVARh	.000
kVAR	.000
Date	.000
Time	.0564
Dependent variable: kWh	

Source: (Setiyorini, 2020)

Table 4. Significant Value of Linearity Test on the Electricity Consumption Dataset B

Model	Sig.
(Constant)	.000
Date	.000
Reactive Power	.000
Voltage	.000
Kitchen	.000
Dependent variable: Active Power	

Source: (Setiyorini, 2020)

Methodology

Figure 1 illustrates the method used in this study which compares the implementation of the linear regressions and neural networks. The initial step taken is preprocessing called normalization. The input variables need to be normalized because they have different dimensions (Tan et al., 2020). In the initial stage, preprocessing data is normalized. In all other data, all datasets are normalized to maximum and minimum values (Zagrebina et al., 2019). Data normalization is performed before the training process begins and each input space is normalized into the [-1, +1] range (Kaytez, 2020). So in this study normalization was conducted by transforming data into intervals of -1 to 1. Furthermore, the normalized dataset forms a new dataset that will be conducted training and testing with 10 fold cross-validation on the electricity consumption dataset using linear regression and neural

networks (Kaytez et al., 2015) to produce RMSE. The RMSE produced by linear regressions and neural networks, then compared to take the smallest RMSE.
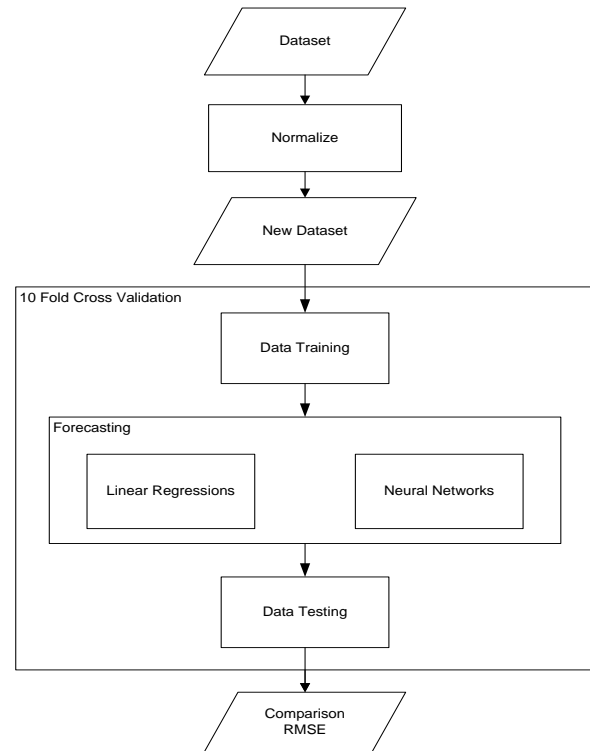
Source: (Setiyorini, 2020)
Figure 1. Comparison of Linear Regressions and Neural Networks

Linear Regression

The regression analysis model is the most famous modeling technique in forecasting electricity consumption (Abdel-Aal & Al-Garni, 1997). Linear regression is included in the statistical analysis method which is applied to characterize the impact of selected independent (predictors) variables on the dependent (response) variable (Fang & Lahdelma, 2016). Linear regression is used for numerical data analysis and modeling (Han et al., 2012). Linear regression still can't be related to nonlinear problems, so it must be studied to find out whether it can be applied to short-term predictions (Shao et al., 2020).

Multiple Linear Regression (MLR) is the generalization of the simple linear regression technique (Aiken et al., 2013)(Fumo & Rafe Biswas, 2015). MLR is an algorithm that describes the relationship between one dependent variable and several independent variables (Shao et al., 2020). The model in multiple linear regression consists of more than one predictor variable:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_P X_P + \varepsilon \quad \text{.................(1)}$$

where Y is the response variable, X_1; X_2; ... X_p is the predictor variables with p as the number of variables, β_0; β_1; ...β_p are the regression coefficients, and ε is an error to account for the discrepancy between predicted data and the observed data (Fumo & Rafe Biswas, 2015).

The linear regression model characterizes the behavior of the unknown quantity y in terms of known quantities x, parameters, and random noise ε (Fang & Lahdelma, 2016). Linear regression forecasting models are expressed in the following format:

$$Y_t = \beta X_t + \varepsilon_t \quad \text{... (2)}$$

where Y_t is the predicted value at time t; X_t = (1, X_{1t}, X_{2t}, X_{3t}, ... X_{nt}) is a vector of k explanatory variables at time t, β = (β_0, β_1, β_2, ..., β_k)^T is the vector of coefficients, and ε_t is a random error term at time t, t = 1, ..., N (Fang & Lahdelma, 2016).

Neural Network

The rapid development of artificial intelligence, neural networks are used for forecasting in various fields. (Yan et al., 2019). The neural network is the deep learning method developed at this time so that it can forecast energy consumption with very high predictive accuracy (Choi et al., 2018).

Neural networks are partial computational models for information processing beneficial for figuring out basic relationships among a set of patterns or variables in data. They are intelligent where the learning methods mimic the biological neural networks especially those in the human brain. The nonlinear and nonparametric nature of neural networks is more of a cable for modeling complex data problems in data mining (Brockmann et al., 2006).

RESULTS AND DISCUSSION

The electricity consumption dataset A and the electricity consumption dataset B experimented with linear regressions and neural networks. Then the RMSE results are compared on the linear regression and neural network of the two datasets. In Table 5 shows the experimental results in the electricity consumption dataset A obtained by RMSE of 0.032 used the linear regression, and RMSE of 0.015 used neural network. In Table 6 shows the experimental results on the electricity consumption dataset B obtained by RMSE of 0.488 used the linear regression, and RMSE of 0.466 used the neural network

Table 5. Comparison of RMSE Results with Linear Regression and Neural Network on the Electricity Dataset A
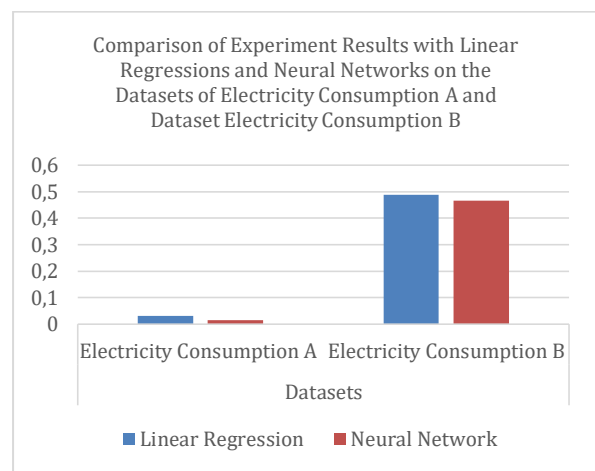
No	Method	RMSE
1	Linear Regression	0.032
2	Neural Network	0.015

Source: (Setiyorini, 2020)

Table 6. Comparison of RMSE Results with Linear Regression and Neural Network on the Electricity Dataset B

No	Method	RMSE
1	Linear Regression	0.488
2	Neural Network	0.466

Source: (Setiyorini, 2020)

Source: (Setiyorini, 2020)

Figure 2. Comparative Graph of Experiment Results of Linear Regressions and Neural Networks on the Dataset of Electricity Consumption A and Dataset Electricity Consumption B

The results of the comparison in Table 5 and Table 6 show significant differences in the value of RMSE between the use of linear regressions and neural networks. This is also shown in Figure 2. which shows the decrease in RMSE values in linear regressions and neural networks. The use of neural networks shows a smaller RMSE value compared to the use of linear regressions. This shows that neural networks have better performance than linear regressions.

As explained earlier, the electricity consumption dataset A and electricity consumption dataset B show nonlinear data relationships. Referring to previous research conducted by (Satre-Meloy, 2019) that linear regression was not suitable for nonlinear models, while neural networks were implemented to model nonlinear relationships to achieved high accuracy (Agrawal et al., 2019). This is proven in this study that neural networks can overcome nonlinear problems in the

electricity consumption dataset A and the electricity consumption dataset B. So that the neural networks can improve performance better than linear regressions.

CONCLUSION

Experiments conducted with linear regressions and neural networks on the electricity consumption dataset A and the electricity consumption dataset B. On the electricity consumption dataset A obtained by RMSE of 0.032 used the linear regression, and RMSE of 0.015 used the neural network. On the electricity consumption, dataset B obtained by RMSE of 0.488 used the linear regression, and RMSE of 0.466 used the neural network. The use of neural networks shows a smaller RMSE value compared to the use of linear regressions. This shows that the problem that is proven in the electricity consumption dataset A and the electricity consumption dataset B is that it has a nonlinear relationship, it can be overcome by the neural network. So that the neural networks can improve performance better than linear regressions.

REFERENCE

Abdel-Aal, R. E., & Al-Garni, A. Z. (1997). Forecasting monthly electric energy consumption in eastern Saudi Arabia using univariate time-series analysis. *Energy*, *22*(11), 1059–1069. https://doi.org/10.1016/S0360-5442(97)00032-7

Agrawal, R. K., Muchahary, F., & Tripathi, M. M. (2019). Ensemble of relevance vector machines and boosted trees for electricity price forecasting. *Applied Energy*, *250*(May), 540–548. https://doi.org/10.1016/j.apenergy.2019.05.062

Aiken, L. S., West, S. G., Pitts, S. C., Baraldi, A. N., & Wurpts, I. C. (2013). Multiple Linear Regression. In *Handbook of Psychology*. John Wiley & Sons, Inc. https://doi.org/10.1002/0471264385.wei0219

Ashfaque, J. M. (2020). *Electricity France*. https://www.kaggle.com/ukveteran/electricity-france/metadata

Brockmann, D., Hufnagel, L., & Geisel, T. (2006). Data Mining and Knowledge Discovery Handbook. In *Springer*. https://doi.org/10.1038/nature04292

Choi, H., Ryu, S., & Kim, H. (2018). Short-Term Load Forecasting based on ResNet and LSTM. *2018 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids, SmartGridComm 2018*. https://doi.org/10.1109/SmartGridComm.2018.8587554

Dodamani, S. N., Shetty, V. J., & Magadum, R. B. (2015). Short term load forecast based on time series analysis: A case study. *Proceedings of IEEE International Conference on Technological Advancements in Power and Energy, TAP Energy 2015*, 299–303. https://doi.org/10.1109/TAPENERGY.2015.7229635

Fang, T., & Lahdelma, R. (2016). Evaluation of a multiple linear regression model and SARIMA model in forecasting heat demand for district heating system. *Applied Energy*, *179*, 544–552. https://doi.org/10.1016/j.apenergy.2016.06.133

Ferreira, V. H., & Alves da Silva, A. P. (2007). Toward estimating autonomous neural network-based electric load forecasters. *IEEE Transactions on Power Systems*, *22*(4), 1554–1562. https://doi.org/10.1109/TPWRS.2007.908438

Fumo, N., & Rafe Biswas, M. A. (2015). Regression analysis for prediction of residential energy consumption. *Renewable and Sustainable Energy Reviews*, *47*, 332–343. https://doi.org/10.1016/j.rser.2015.03.035

Han, J., Kamber, M., & Pei, J. (2012). Data Mining Concepts and Techniques. In *Data Mining*. https://doi.org/10.1016/b978-0-12-381479-1.00001-0

He, Y., Qin, Y., Wang, S., Wang, X., & Wang, C. (2019). Electricity consumption probability density forecasting method based on LASSO-Quantile Regression Neural Network. *Applied Energy*, *233–234*(May 2018), 565–575. https://doi.org/10.1016/j.apenergy.2018.10.061

Islyaev, S., & Date, P. (2015). Electricity futures price models: Calibration and forecasting. *European Journal of Operational Research*, *247*(1), 144–154. https://doi.org/10.1016/j.ejor.2015.05.063

Kaytez, F. (2020). A hybrid approach based on autoregressive integrated moving average and least-square support vector machine for long-term forecasting of net electricity consumption. *Energy*, *197*. https://doi.org/10.1016/j.energy.2020.117200

Kaytez, F., Taplamacioglu, M. C., Cam, E., &

Hardalac, F. (2015). Forecasting electricity consumption: A comparison of regression analysis, neural networks and least squares support vector machines. *International Journal of Electrical Power and Energy Systems*, *67*, 431–438. https://doi.org/10.1016/j.ijepes.2014.12.036

Kim, T. Y., & Cho, S. B. (2019). Predicting residential energy consumption using CNN-LSTM neural networks. *Energy*, *182*, 72–81. https://doi.org/10.1016/j.energy.2019.05.230

Lee, C. M., & Ko, C. N. (2009). Time series prediction using RBF neural networks with a nonlinear time-varying evolution PSO algorithm. *Neurocomputing*, *73*(1–3), 449–460. https://doi.org/10.1016/j.neucom.2009.07.005

Nadtoka, I. I., & Al-Zihery Balasim, M. (2015). Mathematical modelling and short-term forecasting of electricity consumption of the power system, with due account of air temperature and natural illumination, based on support vector machine and particle swarm. *Procedia Engineering*, *129*, 657–663. https://doi.org/10.1016/j.proeng.2015.12.087

Nyberg, H., & Saikkonen, P. (2014). Forecasting with a noncausal VAR model. *Computational Statistics and Data Analysis*, *76*, 536–555. https://doi.org/10.1016/j.csda.2013.10.014

Pombeiro, H., Santos, R., Carreira, P., Silva, C., & Sousa, J. M. C. (2017). Comparative assessment of low-complexity models to predict electricity consumption in an institutional building: Linear regression vs. fuzzy modeling vs. neural networks. *Energy and Buildings*, *146*, 141–151. https://doi.org/10.1016/j.enbuild.2017.04.032

Sadownik, R., & Barbosa, E. P. (1999). Short-term forecasting of industrial electricity consumption in Brazil. *Journal of Forecasting*, *18*(3), 215–224. https://doi.org/10.1002/(sici)1099-131x(199905)18:3<215::aid-for719>3.3.co;2-2

Satre-Meloy, A. (2019). Investigating structural and occupant drivers of annual residential electricity consumption using regularization in regression models. *Energy*, *174*, 148–168. https://doi.org/10.1016/j.energy.2019.01.157

Setiyorini, T. (2020). *Laporan Akhir Penelitian Mandiri*.

Shao, M., Wang, X., Bu, Z., Chen, X., & Wang, Y. (2020). Prediction of energy consumption in hotel buildings via support vector machines. *Sustainable Cities and Society*, *57*(March), 102128. https://doi.org/10.1016/j.scs.2020.102128

Tan, Z., De, G., Li, M., Lin, H., Yang, S., Huang, L., & Tan, Q. (2020). Combined electricity-heat-cooling-gas load forecasting model for integrated energy system based on multi-task learning and least square support vector machine. *Journal of Cleaner Production*, *248*. https://doi.org/10.1016/j.jclepro.2019.119252

Taylor, J. W. (2010). Triple seasonal methods for short-term electricity demand forecasting. *European Journal of Operational Research*, *204*(1), 139–152. https://doi.org/10.1016/j.ejor.2009.10.003

Yan, K., Li, W., Ji, Z., Qi, M., & Du, Y. (2019). A Hybrid LSTM Neural Network for Energy Consumption Forecasting of Individual Households. *IEEE Access*, *7*, 157633–157642. https://doi.org/10.1109/ACCESS.2019.2949065

Yang, W., Wang, J., Niu, T., & Du, P. (2020). A novel system for multi-step electricity price forecasting for electricity market management. *Applied Soft Computing Journal*, *88*, 106029. https://doi.org/10.1016/j.asoc.2019.106029

Zagrebina, S. A., Mokhov, V. G., & Tsimbol, V. I. (2019). Electrical energy consumption prediction is based on the recurrent neural network. *Procedia Computer Science*, *150*, 340–346. https://doi.org/10.1016/j.procs.2019.02.061

Zahid, A. (2020). *Smart Meter Energy(kW) Demand Forecasting*. Kaggle.com. https://www.kaggle.com/asimzahid/smart-meter-energykw-demand-forecasting

Zhao, G. Y., Liu, Z. Y., He, Y., Cao, H. J., & Guo, Y. B. (2017). Energy consumption in machining: Classification, prediction, and reduction strategy. In *Energy* (Vol. 133, pp. 142–157). https://doi.org/10.1016/j.energy.2017.05.110

Zheng, Z., Chen, H., & Luo, X. (2019). Spatial granularity analysis on electricity consumption prediction using LSTM recurrent neural network. *Energy Procedia*, *158*, 2713–2718. https://doi.org/10.1016/j.egypro.2019.02.027