Forecasting annual energy consumption using machine learnings: Case of Indonesia

Robi Kurniawan*)1, 2 and Shunsuke Managi3

1 Graduate School of Environmental Studies, Tohoku University, Japan
2 Ministry of Energy and Mineral Resources, Indonesia
3 Department of Urban and Environmental Engineering, Kyushu University, Japan
*) Corresponding author email: kurniawan.tohoku@gmail.com

Abstract. To understand the future trajectory of energy consumption, we propose to utilize two different machine learning algorithm, artificial neural networks (ANN) and a model tree. Taking Indonesia as a case, the annual gross energy consumption was estimated by modelling a function of urbanization, real GDP per capita proxy for affluence (economic growth), and real capital use per capita. Utilizing the time period of 1971–2014, we train and test the model. Utilizing the root mean square error and the mean absolute error for model selection, we found the tree-based model has a better performance rather than the ANN. Having more superior performance, the tree-based model was then used to forecast the annual energy consumption for the future years. Using specific scenario, the energy consumption is predicted will increase from 883 kg per capita in 2014 to become 1243 kg per capita in 2040. Providing better accuracy, the approach applied in this study can easily be replicated for other countries. Furthermore, it also can be considered in simulating energy demand and environmental consequence in the future.

1. Introduction

Considered as the world’s 10th largest economy, Indonesia has been enjoying high rates of economic development. As the fourth most populous country with fast urbanization growth, the natural capital extraction, including energy, has increased significantly in the few decades [1]. It is highly linkage to Indonesia's sustainable development trajectory. Increasing economic and urbanization has led to increasing energy demand in the country. Having accurate forecasting has many benefits for fulfilling future energy demand and environmental consequence to the country.

Several works found the strong contribution of some socioeconomic variables such as urbanization, economic development, and capital use to energy consumption. Urbanization which takes place in most of developing countries potentially will generate more energy consumption. Urbanization is one of the main determinants of energy consumption in some ASEAN countries including Indonesia [2]. Urbanization, a shifting from agriculture sector toward industries and services sector, is also able to structure out the economic patterns of resource use [3]. Several studies also confirmed the linkage between energy consumption and economic growth. Using Saudi Arabia as a case, Alshehry and Belloumi [4] found a long-run relationship between energy consumption and economic growth in the country. Modifying the Stochastic Impacts by Regression on Population, Affluence and Technology, it is found that the capital use per capita has a significant contribution to energy consumption [5].

On the other hand, many different estimation methods have been developed to forecast the future energy demand. For instance, Bianco, Manca [6] employ parametric approach, multiple and single regression models, to predict electricity consumption in Italia using the socioeconomic data. To
estimate the future trajectory of China’s primary energy consumption, Yuan, Liu [7] developing two univariate models. Combining multiple linear regression and artificial neural networks, Gü nay [8] conduct forecasting the annual electricity demand in Turkey case. Even though after mentioned forecasting approach are developed, selection of the most appropriate technique is of a vital issue. The parametric approach such as linear regression has several drawbacks such as developed from strong assumptions of the data and the model’s specification has to be specified by the user.

For forecasting purposes, instead of using the parametric models, in this study we utilize two different non-parametric machine learning methods known as model tree and ANN. Providing the technical literature on forecasting energy consumptions is the main contribution of this study. Furthermore, the understanding of the future trajectory of energy consumption is crucial especially for Indonesia, the world’s 10th largest economy and the fourth most populous country. The future energy of the country also highly linkage to the environment since primary energy will be fulfilled by fossil fuel such as oil, coal and natural gas [9].

This study is organized as follows. Section 2 provides a methodology and characteristic of the datasets. The results and discussion are presented in section 3. Conclusions and policy implications are highlighted in section 4.

2. Datasets and methods

2.1. Datasets

In this study, annual data of GDP per capita, urbanization, and capital use per capita, are treated as an explanatory variable of the energy consumption per capita in Indonesia. The data for GDP per capita and capital use per capita have been converted in US$ 2010 constant prices. Summary of the descriptive statistics utilized in this study is presented in table 1.

| Variable       | energy consumption | GDP per capita | urbanization | capital use |
|----------------|--------------------|----------------|--------------|-------------|
| Mean           | 578                | 1941           | 34           | 753         |
| Maximum        | 884                | 3693           | 53           | 1296        |
| Minimum        | 297                | 804            | 17           | 201         |
| Std. Dev.      | 201                | 797            | 11           | 291         |
| Unit           | KgOE               | US$ 2010       | %            | US$ 2010    |
| Observations   | 44                 | 44             | 44           | 44          |

Historical data for GDP per capita, urbanization, capital use per capita, and the energy demand per capita has been taken from the World Banks world development indicators. The data period of this work is 1971-2014 as shown in figure 1.
Figure 1. Data used in this study: (a) energy consumption per capita, (b) GDP per capita, (c) urbanization, (d) capital use per capita

2.2. Methods
We employ two different non-parametric machine learning approach, ANN and model tree. Capturing some unknown information which hidden in the data is considered as the primary advantage of ANN approach [10]. ANN capture the linkage between dependents and independent variable using a model derived from the biological brain responds to the stimuli from the input sensor. While the brain employs a neuron, an interconnected cell within the brain, the ANN use the artificial neuron to learn and solve the problem. For estimating the ANN model, we employ the neuralnet package (R environment package) developed by Günther and Fritsch [11].
Figure 2. M5P-model tree algorithm schematic (a) splitting datasets (b) tree development (c) new input data for prediction.

In the other side, we employ a model tree which replaces the single value of leaf nodes with a linear regression model in order to improve the model performance of the regression tree [12]. Doing this approach instead of improving the accuracy, it enables us to maintain models simplicity [13]. To conduct the approach, we utilize the M5P algorithm (see Hornik, Buchta [14]) which provided in the RWeka package of R environment. Figure 1 presents the schematic of the model.
To get higher accuracy, we select the model using the RMSE and the MAE which are regularly employed in prediction model evaluation studies [15]. Without considering the direction of the errors, MAE measures the average magnitude of them in the set of predictions. On the other hand, RMSE is calculated by a quadratic scoring rule that also measures the average magnitude of the error. The indicator is a function of the square root of the average of squared differences between prediction and actual measurement.

3. Result and discussion
This section consists of three main parts. First, we split the datasets randomly into training and testing. Training data, which consist of 70% of the data, is used to build the model. Then, we assess the model accuracy using the remaining 30% of the data. For modelling Indonesia’s energy consumption over the three decades, we apply an ANN and tree model. ANN and tree model were developed to simulate the predicted values of these variables to forecast the consumption using testing model. The results were compared with the real energy consumption in training datasets. This way, the success and the reliability of the forecasting approach applied in this work was validated.

Table 2. Predictive performance of the model

| Indicator/Model              | ANN   | M5P   |
|------------------------------|-------|-------|
| Root mean squared error      | 0.066565 | 0.0222 |
| Mean absolute error         | 0.056363 | 0.022967 |

Table 2 presents the performance of ANN and M5P model. We utilized MAE and RMSE which capture average model prediction error in units of the variable to evaluate the models. The indicators are considered as negatively-oriented scores. According to this approach, lower values of the score are better. From the table, we can derive that the M5P model tree is more superior compared to the ANN approach.

Second, after selecting the appropriate model, we utilize the approach for the out of sample forecasting (2015-2040). Concerning population, we consider the world population prospect of the United Nations for Indonesia projected population. We adapted the socioeconomic projection of shared socioeconomic pathways 1 (SSP1). Considering the scenario, Indonesia’s annual urbanization grows at 1.38 % on average. Furthermore, the economy grows at a constant rate of 4.4 % per annum, we assume the capital also grow at the same rate. Figure 3 presents the historical and forecasting of energy consumption in Indonesia using after mentioned assumption.

Figure 3. Actual and forecasting of the energy consumption using model tree
According to historical datasets, the average annual growth rate of Indonesia’s energy consumption per capita reaches 2.63%. Using the scenario of economic and urbanization projection, the growth rate of consumption per capita will decrease by 1.3%. Translating the result into kg per capita, energy consumption in the country will increase from 883 kg per capita in 2014 to become 1243 kg per capita in 2040.

Third, we discuss the result of the study with another study which applies machine learning method. ANN is considered as the most widely used approach among the machine learning-based model, which has been applied in the field of energy. For instance, Kankal et al. utilized the model to forecast Turkey’s energy consumption [16]. Kialashaki and Reisel also employed the method for modelling the energy demand in the industrial sector of the United States [17]. Comparing the ANN and model tree, our study revealed the model tree outperformed the ANN approach in the energy sector.

On the other hand, focusing on the causality of energy and CO2 emission in Indonesia, Hwang, & Yoo found the bi-directional causality existence among the variables [18]. According to their finding, an increase in energy consumption in the country directly linkage to CO2 emissions, while CO2 emissions also stimulate more energy consumption. It is related to the high share of coal in electricity generation which led to increasing indirect emissions in Indonesia [19]. Some socioeconomic variable such as urbanization economic growth and coal consumption are cointegrated over long-term in Indonesia [20]. The trend of Indonesia’s energy consumption in the next decades should be considered in Indonesia’s CO2 emission mitigation.

4. Conclusion

This study aims to provide the alternative to the technical literature on forecasting energy consumptions. Covering the time period of 1971–2014 datasets, we build two different machine learning approach, artificial neural networks and model tree. We utilized historical data of urbanization, real GDP per capita, and real capital use per capita to capture energy consumption. Employing the RMSE and MAE for model selection, we found the tree model outperformed the ANN which widely utilized for forecasting energy management.

Using a model tree, we used to forecast the annual energy consumption for the future years using the specific scenario of the independent variable. Accordingly, the energy demand will increase from 883 kg per capita in 2014 to become 1243 kg per capita in 2040.

The result brings several implications. First, having strong accuracy, the approach applied in this study can easily be replicated for other countries to make better forecasting of energy consumption for the future. Second forecasting result of the study can be considered to the stakeholder in simulating energy demand and environmental consequence in the future.

5. References

[1] Kurniawan, R. and S. Managi, Economic Growth and Sustainable Development in Indonesia: An Assessment. Bulletin of Indonesian Economic Studies, 2018(just-accepted): p. 1-31.
[2] Azam, M., et al., Factors determining energy consumption: Evidence from Indonesia, Malaysia and Thailand. Renewable and Sustainable Energy Reviews, 2015. 42: p. 1123-1131.
[3] Parikh, J. and V. Shukla, Urbanization, energy use and greenhouse effects in economic development: Results from a cross-national study of developing countries. Global Environmental Change, 1995. 5(2): p. 87-103.
[4] Alshehry, A.S. and M. Belloumi, Energy consumption, carbon dioxide emissions and economic growth: The case of Saudi Arabia. Renewable and Sustainable Energy Reviews, 2015. 41: p. 237-247.
[5] Shahbaz, M., et al., The effect of urbanization, affluence and trade openness on energy consumption: A time series analysis in Malaysia. Renewable and Sustainable Energy Reviews, 2015. 47: p. 683-693.
[6] Bianco, V., O. Manca, and S. Nardini, *Linear regression models to forecast electricity consumption in Italy*. Energy Sources, Part B: Economics, Planning, and Policy, 2013. 8(1): p. 86-93.

[7] Yuan, C., S. Liu, and Z. Fang, Comparison of China's primary energy consumption forecasting by using ARIMA (the autoregressive integrated moving average) model and GM (1, 1) model. Energy, 2016. 100: p. 384-390.

[8] Güney, M.E., Forecasting annual gross electricity demand by artificial neural networks using predicted values of socio-economic indicators and climatic conditions: Case of Turkey. Energy Policy, 2016. 90: p. 92-101.

[9] Hasan, M.H., et al., A review on the pattern of electricity generation and emission in Indonesia from 1987 to 2009. Renewable and Sustainable Energy Reviews, 2012. 16(5): p. 3206-3219.

[10] Zhang, Z., Artificial neural network, in *Multivariate Time Series Analysis in Climate and Environmental Research*. 2018, Springer. p. 1-35.

[11] Günther, F. and S. Fritsch, *neuralnet: Training of neural networks*. The R journal, 2010. 2(1): p. 30-38.

[12] Wang, Y. and I.H. Witten, Induction of model trees for predicting continuous classes. 1996.

[13] Lantz, B., *Machine learning with R*. 2013, Birmingham-UK: Packt Publishing Ltd.

[14] Hornik, K., et al., *Rweka: R/weka interface*. 2016.

[15] Chai, T. and R.R. Draxler, Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE in the literature. Geoscientific model development, 2014. 7(3): p. 1247-1250.

[16] Kankal, M., et al., Modeling and forecasting of Turkey’s energy consumption using socio-economic and demographic variables. Applied Energy, 2011. 88(5): p. 1927-1939.

[17] Kialashaki, A. and J.R. Reisel, Development and validation of artificial neural network models of the energy demand in the industrial sector of the United States. Energy, 2014. 76: p. 749-760.

[18] Hwang, J.-H. and S.-H. Yoo, Energy consumption, CO 2 emissions, and economic growth: evidence from Indonesia. Quality & Quantity, 2014. 48(1): p. 63-73.

[19] Kurniawan, R., Y. Sugiantan, and S. Managi, *Cleaner energy conversion and household emission decomposition analysis in Indonesia*. Journal of Cleaner Production, 2018. 201: p. 334-342.

[20] Kurniawan, R. and S. Managi, Coal consumption, urbanization, and trade openness linkage in Indonesia. Energy Policy, 2018. 121: p. 576-583.

**Acknowledgements**

This paper was supported by Grant-in-Aid for Specially Promoted Research (26000001) by the Japan Society for the Promotion of Science. Any opinions, findings, conclusions, and recommendations expressed in this study are those of the authors and do not necessarily reflect the views of the institutions and funding agencies. We acknowledge comments from the reviewer and the editor of International Conference on Future Environment and Energy 2019.