BIOGAS ELECTRICITY PRODUCTION FORECASTING IN LIVESTOCK FARMS USING MACHINE LEARNING TECHNIQUES: A CASE STUDY IN VIETNAM

DỰ BÁO SẢN LƯỢNG ĐIỆN KHÍ SINH HỌC Ở CÁC TRANG TRẠI CHĂN NUÔI SỬ DỤNG CÁC THUẬT TOÁN HỌC MÁY: MỘT NGHIÊN CỨU TẠI VIỆT NAM

Nguyen Duy Hieu[1], Nguyen Vinh Anh[1], Hoang Anh[2], Hoang Duc Chinh[1,*]

DOI: https://doi.org/10.57001/huih5804.2023.060

ABSTRACT

Biogas energy is considered a renewable energy source. The efficient usage of biogas resources can help reduce greenhouse gas emission, especially methane, generate electricity to power farms' loads, and decrease load demand on grids. We first present the data acquisition scheme of self-developed biogas generation systems, complete with a description of the farm architecture and load estimation. Then, with the necessary data collected, five machine learning techniques are then explored and adopted to process the data and forecast energy production at several livestock farms in practice. Comparisons are made among these techniques, which includes RNN, MLP, polynomial regression, decision trees and random forest regression, to evaluate the accuracy of the predictions. It was concluded from the comparisons that Polynomial Regression performed the best in predicting the energy production at the hog farm, while random-tree-based methods performed the worst.

Keywords: *Biogas energy, machine learning, energy forecasting.*

TÓM TẮT

Khí sinh học biogas có thể được coi là một nguồn năng lượng tái tạo. Việc sử dụng các nguồn khí sinh học một cách hiệu quả có thể giúp giảm lượng khí thải nhà kính, đặc biệt là methane, phát điện để đáp ứng một phần nhu cầu năng lượng ở các trang trại, và giảm chi phí sử dụng điện lên lưới điện. Bài báo này trình bày một hệ thống thu thập dữ liệu của hệ thống phát điện khí sinh học đã được xây dựng, bao gồm kiến trúc của hệ thống và ước lượng tải của trang trại. Chúng tôi cũng tiến hành thử nghiệm năm thuật toán học máy khác nhau là RNN, MLP, hồi quy đa thức, và hai thuật toán dẫn xuất của cây quyết định để xử lí thông tin thu thập được và dự báo sản lượng điện ở các trang trại trong thực tế. Kết quả áp dụng các thuật toán này được so sánh với nhau để đánh giá tính chính xác của dự báo. Qua kết quả thu được, có thể thấy rằng phương pháp hồi quy đa thức có độ chính xác cao nhất, và các mô hình dẫn xuất của cây quyết định có độ chính xác kém nhất.

Từ khóa: *Năng lượng khí sinh học, học máy, dự báo năng lượng.*

[1]School of Electronics and Electrical Engineering, Hanoi University of Science and Technology

[*]Email: chinh.hoangduc@hust.edu.vn

Received: 22/10/2022

Revised: 05/02/2023

Accepted: 15/3/2023

1. INTRODUCTION

Energy is the fuel of civilization. It is part of the fundamental high-resolution foundation that upholds the lower-resolution, more abstract functioning of our society, and it was estimated that the total electricity consumption of the world was around 25 TWh in 2019 [1]. The demand for energy is ever-growing, with primary energy having experienced an estimated 31-exajoule increase in 2021 [2]. Although most of the energy demand was met with fossil fuel, which accounted for 59% of 2021's energy generated, renewable energy had nevertheless assimilated a considerable 13% share of global power generation, which, remarkably, was higher than that of nuclear energy, which was 9.8% [1]. The sources of renewable energy that constitutes this growth included solar, wind energy, biofuels, etc.

Biogas, which is a form of biofuel, is a gaseous fuel obtained from the anaerobic digestion of organic material. The composition of a biogas mix typically includes methane, carbon dioxide, hydrogen sulfide, ammonia, and hydrogen [3]. Energy is obtained from the combustion of methane in the biogas mix. Biogas is a renewable, environmentally friendly source of energy that has an advantage over other sources of renewable energy in terms of ease of control. Since biogas can be obtained from the anaerobic digestion of organic waste, which is often in steady supply, it can be considered renewable. Biogas energy generation can be labeled as carbon-neutral because the carbon dioxide that the combustion of biogas produces has been fixed from the atmosphere by the plants from which the organic waste originates. Biogas energy is a dispatchable source of energy, which means that electricity generation from biogas can be activated or deactivated on command [4]. This adjustability of operation of biogas energy presents an opportunity for control and optimization that is less directly viable in non-dispatchable sources of energy like solar or wind. Because these methods of energy generation are more reliant on external non-operational factors like the weather, scheduled operation is achieved through the use of an energy imbalance market, or energy storing systems. This makes them less flexible and less efficient than a dispatchable source of energy, which bypasses such necessities, and whose operation hours can be directly adjusted according to energy demand. Moreover, with recent advances in technology, the production of biogas can be predicted to an extent with the use of machine learning [5], making biogas systems relatively more stable compared to less predictable sources of energy.

With the rise in renewable energy comes necessity for the adoption of smaller, more local energy frameworks for more efficient distribution, storage and consumption. One of such energy frameworks is the microgrid. A microgrid structure works to provide users in a small geographic area with electricity generated from renewables or pulled from the utility grid if necessary. However, electricity distribution in a microgrid system needs to be intelligently controlled for its potential to be fully realized. As one strong tool that facilitates the efficient operation microgrid is energy forecasting, this study will explore the predictive power and accuracy of different machine learning algorithms in forecasting electricity production and the benefits that the use of such algorithms in a microgrid structure, a farm, may bring.

Some similar works have been done in the past as energy forecasting with the use of machine learning has been studied for many years. For example, the use of advanced neural network models was examined in 1996 in [6]. Most of the volume of research focusing on energy forecasting was done on the topics of load forecasting, price forecasting, and wind and solar energy forecasting. Studies on machine learning models for solar energy forecasting extends to recent time in [7], where the accuracy of four different machine learning models, which are linear regression (LR), random forest (RF), Support Vector Regression (SVR), and (Artificial Neural Network) ANN was tested with real data and evaluated on six metrics. Data of wind energy was also used to test the models, with [8] testing xGBoost, SVR, and RF.

The energy forecast in a building microgrid structure is also thoroughly studied. A bibliometric analysis of building energy prediction using artificial neural network was conducted in [9], and it was found that towards the recent years, both the publication and citation counts of building energy prediction has been experiencing strong increases, with there being over 100 publications in this topic in 2020. It is difficult to evaluate the benefit that the use of energy forecasting may bring to an energy system. In [10], Zhou et al. investigated the use of a theoretical game model to describe energy management, tested three different short-term wind energy forecasting algorithms, and simulated the effects the algorithms may have on a generic energy framework that includes a microgrid. The results included a verification of the game-theoretical model and that the proposed algorithm used, genetic SAE, outperforms two other algorithms that were tested. It can be observed that although research works in the topic of energy prediction have been done before, most of them differs from this one either in terms of type of forecast target, type of microgrid structure, or the type of data utilized (simulation data or real data).

For our contribution, in this work, we first develop a data acquisition scheme for biogas energy generation systems and accumulate their data into datasets over time. Subsequently, because in practice, the operation of these systems are much different from one another due to the size of the farms, the biogas production capacity of the systems, the types of electric loads, the habit of operation, etc., machine learning techniques such as multiple linear regression (MLR), polynomial regression, decision tree (DT), random forest regression and recurrent neural network (RNN) are explored and applied to understand the energy production of those generators. The structure of our paper is as follows: The background of the adopted machine learning techniques is presented briefly in Section 2. Section 3 introduces the system description of the biogas-based generators in livestock farms and the data acquisition scheme. Section 4 discusses the metrics of performance evaluation and the results of the studies. We conclude our works in Section 5.

2. MACHINE LEARNING TECHNIQUES FOR ENERGY FORECASTING

As mentioned in Section 1, machine learning has been widely applied in energy forecasting problems. Listed below are a few algorithms well-known in various prediction problems and adopted in our works.

Multiple linear regression. MLP is a technique that attempts to model a response variable (dependent variable) based on two or more explanatory variables (independent variables) [11]. Assuming that this relationship can be represented by a linear model the observed data is fit into a linear equation to construct the model. This model can then be used to predict the response values from some additional data collected. The general model of MLP with given n observations have the form of

$$y_i = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_p x_{ip} + \varepsilon_i \tag{1}$$

Where:

y_i is a value of the dependent variable y, and $x_{i1}, x_{i2}, \ldots, x_{ip}$ are the values of the p independent variables $x_1, x_2, \ldots, x_p$ in the data set.

$\beta_0, \beta_1, \ldots, \beta_p$ are the regression coefficients obtained once the model has been developed.

ε_i is the error term or the disturbance term, which represents the difference between the estimated value achieved by the model and the actual one due to factors other than the independent variable x and should be selected with an appropriate estimation method.

Polynomial regression. Different from linear regression, polynomial regression models the dependent variable as a polynomial function of the independent variable, so it can be considered a non-linear modeling approach [12]. The relationship between the dependent and independent variables are shown below:

$$y = b_0 + b_1x + b_2x^2 + b_3x^3 + \ldots + b_nx^n \quad (2)$$

The high order terms of the independent variables are introduced with the expectation that the accuracy of the model can be improved.

Decision trees. Decision trees (DTs) are non-parametric supervised learning methods that are commonly used for classification and regression problems. This group of learning methods aims to predict the value of a target variable by interpreting data features to get simple decision rules [13]. The DT has some advantages such as simplicity, little data preparation, relatively fast execution.

Random Forest Regression. Random forests (RF) regression is a generally superior form of decision trees regression that has a lower probability of overfitting than normal decision trees regression. RF regression achieves many different purposes by generating multiple decision trees during the training phase. If the goal is to classify, the mode of all the decision trees' final selections will be the output of the forest. Besides, if the purpose is regression, the mean of all the decision trees' output will be the output of the forest [14].

Recurrent neural network (RNN). A recurrent network is a neural network capable of working with input temporal or sequence data, so it is suitable for handling tasks such as voice recognition, language processing, etc. The difference between an RNN and a feedforward neural network is that the middle layer of a recurrent network feeds information not only forward to the output layer but also back to itself in the next time step in the sequence and thus enables the processing of information in the time domain [15].

3. SYSTEM DESCRIPTION

3.1. Biogas based generation system in livestock farms

The biogas-based generation systems in this study are self-developed and deployed in a serval livestock farms in Vietnam. The whole farm grid is illustrated in Fig. 1. The main components of such a system are biogas tanks, which collect the waste, a filter system, which removes unwanted gases, a fuel tank, a mixing tank, a biogas electrical generator, and a control and supervisory system. According to local regulation, the generation system must be connected to the farm distribution grid in the island mode (off-grid) and serve as an alternative source beside the main grid. The farm owners intend to maximize biogas consumption to either avoid releasing unburned biogas into the atmosphere or generate electricity to power the farm loads. Typical loads in the livestock farms include the pump, the cooling fans of each barn, the lighting system, and other miscellaneous loads. Amongst these, the pump, the manure, the biogas dehydrators, and the cooling fans are heavy loads that can consume up to a few kW.

In reality, the average power consumption in a hog farm is around 4.5 ÷ 6.2kW per 1000 pigs. This power consumption depends on the electrical equipment or appliances used in the farm. It is also affected by the way the distribution system is installed in the farm, i.e. whether the system has its own transformer substation or it needs long transmission lines, which may entail significant voltage drop or power loss along the line. Furthermore, farm operators may not pay attention to the maintenance the generation system, thus the energy efficiency tends to decrease over the time.

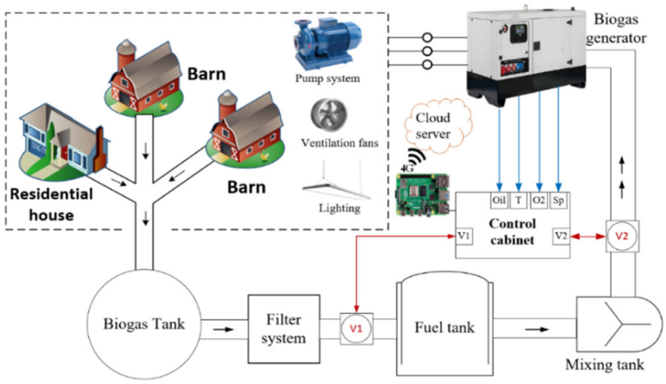

Figure 1. The livestock farm equipped with biogas-based generation systems and typical electrical loads.

3.2. Data acquisition scheme

Various parameters of the generation system are being measured by the corresponding sensors connected to the control and supervisory system. They are the cooling water temperature of the engine, the oil pressure, the speed, the oxygen concentration in the exhaust fumes, and the electrical parameters like three phases voltage, current, active power, active energy, power factor, etc, of which all the electrical parameters are used for prediction. The control and supervisory system are equipped with an embedded computer as shown in Fig. 1 which enables it to acquire sensing data and store it locally. The data is also transferred to a cloud server over the Internet for further processing. As all of the developed biogas-based energy generation systems in this research have been deployed in rural areas, it is essential to support collecting data remotely to facilitate the developer team in studying the systems' operation more efficiently. The sensing data is post-processed at the server to filter out outliers caused by sensor noise or failures of the system before being used for prediction. The outliers are nonetheless still useful for the analysis of the condition of the generation system, but this is not covered in the scope of this work.

4. RESULTS AND DISCUSSION

We have deployed more than ten biogas-based generation systems over the last one year. Amongst those, the collected data of four systems which are being operated more frequently are selected to show in this research. These systems are installed in different hog farms in the northern part of Vietnam, where the winter has more influence on their operation. Information of the farms and the respective generation system is shown in Table 1. The scale, the capacity, the generator ratings, and the electrical loads are properties unique for each farm.

Table 1. Information of the livestock farms and their biogas generators

Farm ID	Size (m^2)	Number of pigs	Rated power of generator (kW)	Power consumption of heavy loads (kW)
06	5000	5000	90	Water treatment system (20kW)
09	2000	2000	80	Cooling fans, office building
11	7000	4000	120	Water treatment system equipped with high power motor of 450kW
14	63000	20000	120	Cooling fans

4.1. Energy forecasting

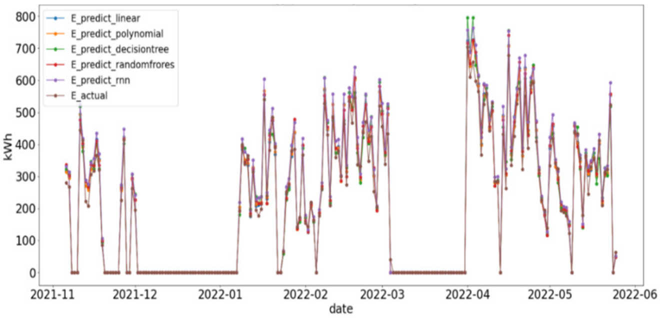

Figure 2. Energy forecasting of the biogas generation system in farm ID 06 over 8 months

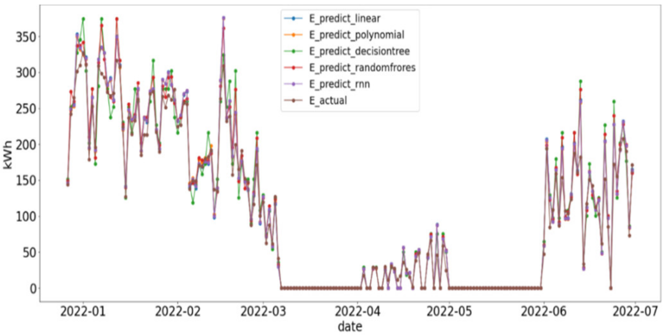

Figure 3. Energy forecasting of the biogas generation system in farm ID 09 over 6 months

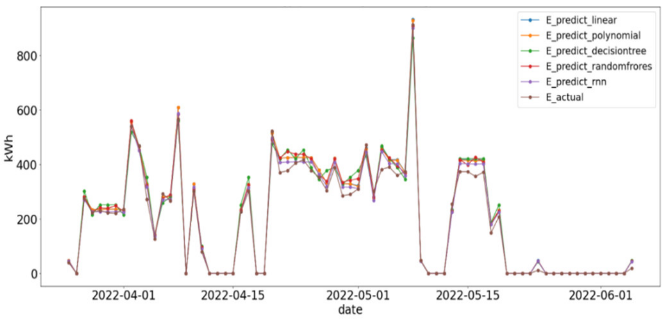

Figure 4. Energy forecasting of the biogas generation system in farm ID 11 over 3 months

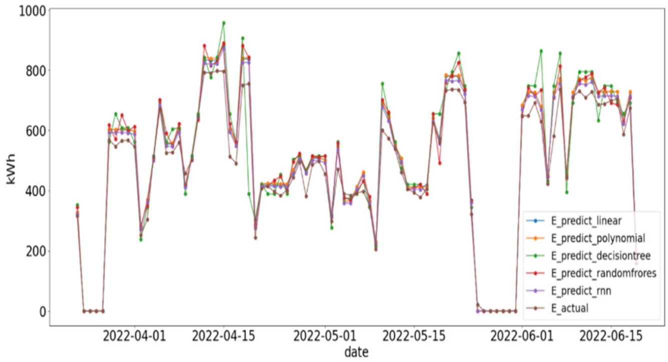

Figure 5. Energy forecasting of the biogas generation system in farm ID 14 over 3 months

The five techniques presented in Section 2 are applied to forecast biogas energy production in different farms using the past data. The data set is divided into the train set and the test set which are 80% and 20% of the original set.The Scikit-learn library is used to train the data with different algorithms as mentioned in session 2. Fig. 2, 3, 4, 5 show the energy forecasting in the four selected farms compared with the actual consumption. Fig. 2 and 3 show the energy usagein the two individual farms 06 and 09 over a period of eight and six months respectively leading up to the study. It can be observed that the biogas energy consumption here was intermittent due to the fact that most of the farm's pigs may have been sold during certain periods, and thus no biogas was produced, and energy demand decreased dramatically. Generators in farm ID 11 and 14 are newly installed, so there is less data collected, and the energy forecasting have been performed only for the three months in summertime.

4.2. Performance evaluation:

Evaluation metrics. In general, energy consumption trends predicted with the machine learning algorithms are more or less similar to the actual one as presented in the figures in Section 4.1. However, the accuracy varies in the cases of different farms. Three different metrics are used to evaluate the precision of the forecasting results shown above quantitatively.

Mean absolute error (MAE). This is a simple metric that reflects the difference between the predicted values $\hat{y}_i$ and the actual values y_i. When using this metric, all the data points are considered the same without any exceptions, and thus the influence of outliers is not included.

$$\mathrm{MAE} = \frac{\sum_{i=1}^{n} \left| y - \hat{y} \right|}{n} \quad (3)$$

Mean square error (MSE). The MSE is most widely used for regression models. The MSE is computed below

$$\mathrm{MSE} = \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{n} \quad (4)$$

It is expected that good models have the smallest values of MAE and MSE possible.

R^2 score. R^2 score is an evaluation metric that describes to what extent is the variation in the dependent variable can be attributed to the independent variables. An R^2 score of a model reflects how closely it can estimate the data trend and thus how well it can be used to make predictions. The R^2 score value can be either positive and less than 1.0, which is the highest score possible, negative, which signifies bad modelling, or 0.0. If a model has an R^2 score of 0.0, it invariably predicts the expected value of the output regardless of the input. The formula for the R^2 score is given below:

$$R^2_{(y,\hat{y})} = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2} \quad (5)$$

Where $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$ and $\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \sum_{i=1}^{n}\varepsilon_i^2$

The evaluation. Evaluation of the models obtained with different algorithms for different generation system is shown in Table 2, 3 and 4.

Table 2 presents the precision of the models measured by the R^2 score. It can be observed from the table that the polynomial model made predictions with the highest precision for all 4 generators. The Decision Tree and Random Forest models were less accurate with R^2 scores of 0.6 - 0.8 with generator 06 and 09 and they are even negative in the case of generator 11 and 14.

In table 3 and 4 are evaluation results obtained using the MAE and the MSE performance metrics. Both metrics also have the polynomial model as the most precise model out of the 5 models considered, producing the lowest error as well as having the highest overall accuracy rate. As a result, it can be a good candidate to perform energy generation forecasting using the polynomial model.

Nevertheless, precision evaluation results of the RNN and the MLP were only marginally inferior to that of the polynomial model, so they can be viable alternatives. The Decision Tree and Random Forest models' precision is low, and their error rates are high, so they are less suitable for use in this system.

Table 2. R^2 scores of models obtained with different algorithms

Algorithms	Generator 06	Generator 09	Generator 11	Generator 14
MLP	0.841	0.894	0.287	0.322
Polynomial	0.843	0.895	0.289	0.326
Decision Tree	0.635	0.786	-0.444	-0.349
Random Forest	0.784	0.866	0.119	0.1723
RNN	0.827	0.893	0.291	0.310

Table 3. MAE of models obtained with different algorithms

Algorithms	Generator 06	Generator 09	Generator 11	Generator 14
MLP	0.00593	0.0281	0.0396	0.0371
Polynomial	0.00592	0.0278	0.0392	0.0365
Decision Tree	0.00593	0.0281	0.0396	0.0371
Random Forest	0.00593	0.0281	0.0396	0.0371
RNN	0.00625	0.0279	0.0389	0.0364

Table 4. MSE of models obtained with different algorithms

Algorithms	Generator 06	Generator 09	Generator 11	Generator 14
MLP	0.00833	0.0369	0.0505	0.0477
Polynomial	0.00827	0.0367	0.0504	0.0476
Decision Tree	0.01262	0.0509	0.0719	0.0672
Random Forest	0.00970	0.0402	0.0561	0.0527
RNN	0.00868	0.0360	0.0504	0.0481

5. CONCLUSION

In this paper, we have presented the data collection scheme of biogas generation system in livestock farms. The data set is helpful for the community to understand the biogas energy production and usage in the rural areas in Vietnam. Machine learning techniques have also been explored to forecast and help understand the energy demand of the livestock farm. It is also suggested which techniques should be good options to apply in the case of biogas energy production in livestock farms. These initial analyses enable the farm owners to adjust the generator operation plan so that the usage of biogas produced for generating electricity is optimized. Subsequently, it would result in the reduction of electrical bills and maximizing the profit of the business. There are still challenges and uncertainties affecting the prediction, such as weather conditions, livestock diseases, the change in livestock market demand, etc. In future works, we would like to include more input information to improve the prediction models and provide recommendation services to the users of the biogas generation system for better operation.

REFERENCES

[1]. International Energy Agency.: *Electricity - Fuels & Technologies*, https://www.iea.org/fuels-and-technologies/electricity, last accessed 2022/07/09.

[2]. BP plc.: *Statistical Review of World Energy 2022* 71st edition, https://www.bp.com/content/dam/bp/business-sites/en/global/corporate/pdfs/energy-economics/statistical-review/bp-stats-review-2022-full-report.pdf, last accessed 2022/07/09.

[3]. Wellinger A., Murphy J., Baxter D., 2013. *The Biogas Handbook: Science, Production and Applications.* Elsevier.

[4]. Hung D. Q., Mithulananthan N., Lee K. Y., 2014. *Optimal placement of dispatchable and nondispatchable renewable DG units in distribution networks for minimizing energy loss.* International Journal of Electrical Power & Energy Systems, 55, pp. 179–186.

P-ISSN 1859-3585 | E-ISSN 2615-9619

[5]. Djavan De Clercq, Devansh Jalota, Ruoxi Shang, Kunyi Ni, Zhuxin Zhang, Areeb Khan, Zongguo Wen, Luis Caicedo, Kai Yuan, 2019. *Machine learning powered software for accurate prediction of biogas production: A case study on industrial-scale Chinese production data.* Journal of Cleaner Production vol 218, 390-399.

[6]. Kariniotakis G. N., Stavrakakis G. S., Nogaret E. F., 1996. *Wind power forecasting using advanced neural networks models.* IEEE Trans. Energy Convers., vol. 11, no. 4, pp. 762–767.

[7]. Jebli I., Belouadha F. Z., Kabbaj M. I., Tilioua A., 2021. *Prediction of solar energy guided by pearson correlation using machine learning.* Energy, vol. 224, 120109.

[8]. Demolli H., Dokuz A. S., Ecemis A., Gokcek M., 2019. *Wind power forecasting based on daily wind speed data using machine learning algorithms*. Energy Conversion and Management, vol. 198, 111823.

[9]. Hong T., Pinson P., Wang Y., Weron R., Yang D., Zareipour H., 2020. *Energy Forecasting: A Review and Outlook.* IEEE Open Access Journal of Power and Energy, vol. 7, pp. 376–388.

[10]. Zhou Z., Xiong F., Huang B., Xu C., Jiao R., Liao B., Yin Z., Li J., 2017. *Game-theoretical energy management for energy internet with big data-based renewable power forecasting.* IEEE Access, vol. 5, pp. 5731–5746.

[11]. Freedman D. A., 2009. *Statistical models: theory and practic*e. 2nd edn. Cambridge university press.

[12]. Ostertagová E., 2012. *Modelling using Polynomial Regression.* Procedia Engineering, vol. 48, pp. 500–506.

[13]. Lozano-Medina J.I., Hervert-Escobar L., Hernandez-Gress N., 2020. *Risk profiles of financial service portfolio for women segment using machine learning algorithms*. In Computational Science - ICCS 2020. Lecture Notes in Computer Science, vol. 12143. Springer, Cham.

[14]. Tin Kam Ho, 1995. *Random decision forests*. In Proceedings of 3rd International Conference on Document Analysis and Recognition, vol.1, pp. 278-282.

[15]. Medsker L. R., Jain L. C., 1999. *Recurrent neural networks: Design and Applications* 1st edn. CRC Press.

THÔNG TIN TÁC GIẢ

Nguyễn Duy Hiếu, Nguyễn Vinh Anh, Hoàng Anh, Hoàng Đức Chính

Trường Điện - Điện tử, Đại học Bách khoa Hà Nội