CHEMICAL ENGINEERING TRANSACTIONS

VOL. 103, 2023

A publication of AIDIC
The Italian Association of Chemical Engineering
Online at www.cetjournal.it

Guest Editors: Petar S. Varbanov, Panos Seferlis, Yee Van Fan, Athanasios Papadopoulos

ISBN 979-12-81206-02-1; **ISSN** 2283-9216

DOI: 10.3303/CET23103116

Smart Energy Strategy - A Comparative Study of Energy Consumption Forecasting Machine Learning Models

Anastasia Zafeiriou[a,*], Georgios Chantzis[a], Tomas Jonkaitis[b], Paris Fokaides[c], Agis Papadopoulos[a]

[a]Aristotle University of Thessaloniki, Process Equipment Design Laboratory, School of Mechanical Engineering, 54124 Thessaloniki, Greece
[b]UAB Caverion Lietuva, Skersinės g. 9, LT-08449 Vilnius, Lithuania
[c]Frederick University, School of Engineering, 7, Frederickou Str., 1036, Nicosia, Cyprus
akzafeir@meng.auth.gr

The energy sector is going through a very turbulent period since 2021. The energy crisis, which started before the war in Ukraine but was magnified by it, led to highly volatile energy prices that coincided with the post-pandemic economic recovery and the expansion of economic growth, resulting in a rather unstable situation. Still, European countries reacted with a surge in awareness of energy efficiency and in adopting energy-saving measures whilst trying to accelerate the propagation of renewable energy sources. The latter can only be successful if it is based on a cohesive energy strategy. An appropriate energy strategy has to be based on managing the demand, which in turn leads to the necessity to predict consumption values. Nowadays, this can be done with various models, which, however, must be trained. One of the tools for successfully creating energy consumption forecasts is supervised machine learning (ML). In this research, 7 different ML models were examined and trained, with the help of Python and with an input of 5 y of data or, alternatively, with real hourly energy consumption data of a tertiary sector's building. This analysis aids the selection of the appropriate model considering internal and external factors, with a special focus on cyclicality, seasonality, and trend in both hourly and daily predictions. The results indicate that a deep learning architecture based on an artificial recurrent neural network (RNN) is chosen to better deal with hourly predictions of energy consumption. The effectiveness of the models applied to the time series data decreased in the evaluation of the daily data. However, Gradient Boosting frameworks based on decision trees show that valuable predictions are feasible. Lastly, it was proven that the ambient temperature improves the effectiveness of forecasts both in hourly and daily models.

1. Introduction

Pandemics, conflicts, a shortage of resources, and inefficiencies in the electrical system are all key events that highlight the importance of an effective energy management plan. The pillars of the Sustainable Energy Action Plan are energy conservation and energy efficiency, renewable energy development, and education and awareness. An efficient strategy for energy conservation and improving the energy efficiency of electrical energy consumption is its accurate forecast (Shklyarskiy and Batueva, 2019). When considering the problem from the viewpoint of a building, energy management systems must include a reliable method for estimating the energy consumption of the building (Karijadi and Chou, 2022).
Predictive analysis, which uses computer tools capable of seeing patterns in the examined data based on the same rules that may be used to elaborate forecasts, is one of the issues that organizations confront today (Candanedo et al., 2018). Accurate predictions go a long way in understanding the causes of high-power consumption and making early decisions. Machine learning (ML) methods are known to be the best approach to achieving desired results in predictive tasks in this domain (El Maghraouix et al., 2022). As a result, ML has been used in several studies related to energy prediction in conventional and especially residential buildings.
Many have examined the reliability of simple models in time series data, like Hewamalage et al. (2022), who proved that when patterns become complex, nonlinear global models such as RNNs, LGBMs are superior to

Paper Received: 12 April 2023; Revised: 27 June 2023; Accepted: 28 July 2023
Please cite this article as: Zafeiriou A., Chantzis G., Jonkaitis T., Fokaides P., Papadopoulos A., 2023, Smart Energy Strategy - A Comparative Study of Energy Consumption Forecasting Machine Learning Models, Chemical Engineering Transactions, 103, 691-696 DOI:10.3303/CET23103116

simple linear global models. Shapi et al. (2021) studied six months of energy demand data from two commercial buildings by applying 3 ML methods. The SVM method, in this case, shows the most promising result. Khan et al. (2020) proposed an ML-based hybrid approach, combining multilayer perceptron, support vector regression, and CatBoost for power forecasting. On the other hand, Chou et al. (2018) proved that the single machine learning models are the best choice for users who would like to promptly estimate the energy consumption of buildings. Recent studies have examined the various types of machine learning techniques and their connection to energy consumption data, but few have compared the effectiveness of the models, both on hourly and daily levels, within a specific case study with real (more than a year) hourly data. There is a research gap in the application of external parameters to optimize the performance of models.
In this paper, the goal is to create prediction models for the energy consumption of one building using single ML models considering both internal and external factors. Historical data covers 5 y (00:00 January 5, 2017, to 09:00 November 15, 2022) kWh of consumption of a business building. The external factors refer to the air temperature of the environment, relative humidity, and wind speed. The data were evaluated on an hourly and daily basis, and the models were applied to both time series. The following models were developed: Autoregressive integrated moving average (Arima), Light Gradient-Boosting Machine (LGBM), Long Short-Term Memory networks (LSTM), Prophet, Random Forest (RF), Ridge and Extreme Gradient Boosting (XGBOOST). The effectiveness of the models was compared and evaluated by applying the index Root Mean Square Error (RMSE). The prediction accuracy depends on the structure and patterns of investigated values. The best-suited prediction model could be different if, for example, heating or water consumption were investigated for the same building.

2. Method and theory

2.1 Data preparation and pre-processing

When preprocessing the data, the issues faced must deal with datetime indices, missing data, and duplicates. To apply ML models to forecasting problems, the time series must be transformed into a matrix in which each value is related to the time window (lags) that precedes it (Rodrigo and Ortiz, 2021). Lastly, the data were separated into 4 y of training and 1 y of testing. A time series graph, both historically and at random time intervals, visually highlights the behavior and patterns of the data and can lay the foundation for building a reliable model. In this step, the seasonal decompose results object was used and produced the results depicted in Figure 1, which shows the a) daily, b) weekly, and c) monthly seasonality of the data. There is a daily seasonality, with demand decreasing between 08:00 and 18:00 h. Weekly seasonality shows lower demand values during the weekend. It is observed that there is an annual seasonality, with higher median demand values in October, January, and July.

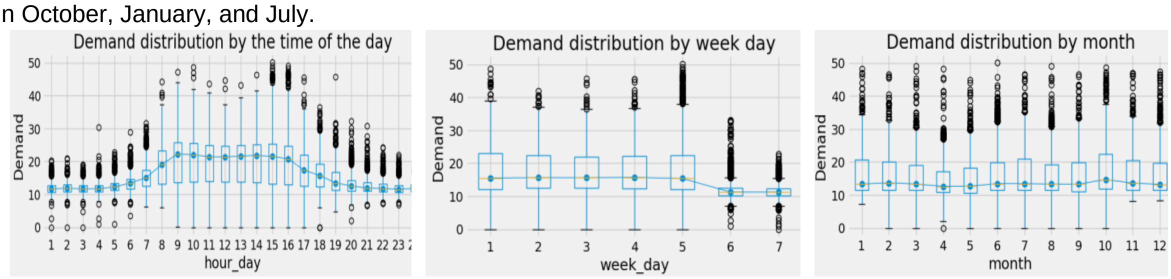

Figure 1: Boxplots for a) daily, b) weekly, c) annual seasonality

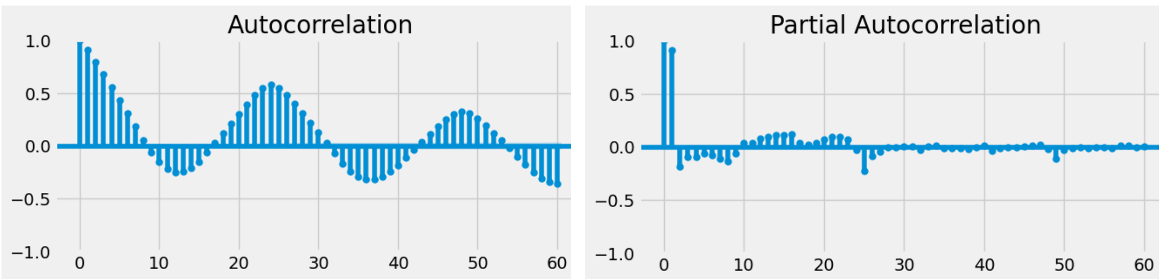

Figure 2: Autocorrelation and Partial Autocorrelation graphs

The autocorrelation and partial autocorrelation graphs (Figure 2) clearly demonstrate a relationship between the demand for 1 h and previous hours and the demand for that same hour on the previous day. Evidence that autoregressive models can function well is this kind of correlation.

2.2 Model creation and training

Eight models were created and trained with different regressors and lags in both hourly and daily data. Skforecast library was a significant tool for the creation of most of the models. In the case of the LSTM neural network and XGBOOST, the tensorflow library also played an important role. Hyperparameter tuning was used to further improve the performance of RF and Ridge machine learning models. In this step, the relationship between energy consumption and ambient temperature, humidity, and windspeed was examined. Of the external factors, it was shown that only the ambient temperature can be used to optimize the predictions. Therefore, its correlation with the time series data is chosen to be presented below (Figure 3). In this diagram, consumption values are in kWh, and temperature values are in °C.

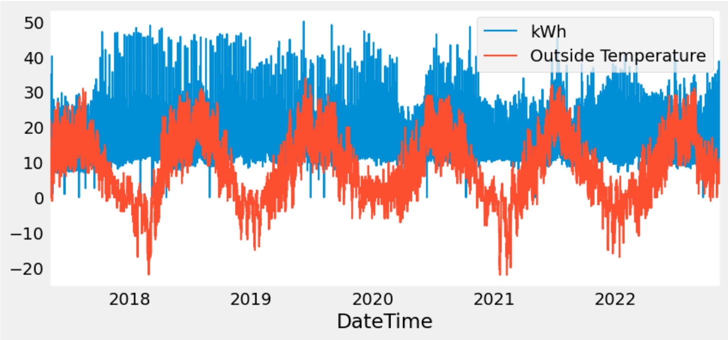

Figure 3: Electricity Hourly Demand and ambient temperature

There are many error metrics that can be used to compare the accuracy of the forecasting models, such as Average Forecasting Error (AFE), Mean Absolute Error (MAE), Mean Absolute Deviation (MAD), Performance Parameter (PP), and Root Mean Square Error (RMSE). In this paper, the accuracy of the forecasting results was compared using the RMSE. The choice of this particular error metric is because RMSE is considered the most widely used error measure in regression situations, and it's recommended for energy demand data (Lee et al., 2022).

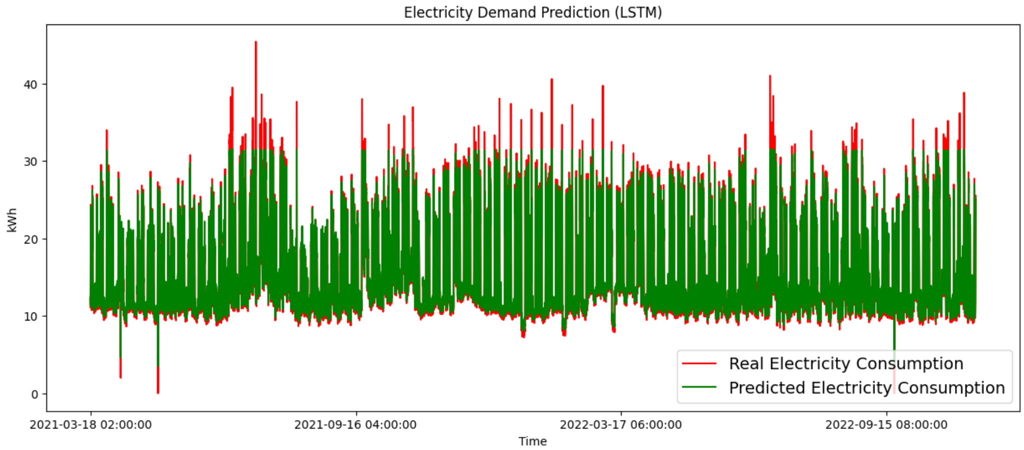

Figure 4: Electricity Hourly Demand Prediction - LSTM Model

2.3 Model testing and final model

The main task of this part was the evaluation of the models by the root mean squared error (RMSE) index using the sklearn.metrics class. The hourly performance of the models was clearly more efficient than that of the daily models. Figure 4 shows the noteworthy application of the LSTM model to the hourly data (RMSE: 0.524). LSTM is one of the most common RNN architectures, which produces results for sequential data tasks because of its ability to capture long-time dependencies (Crisóstomo et al., 2020). RNNs are neural networks for predicting outcomes from sequence data. Input, hidden, and output layers make up traditional neural networks, which use a feedforward process to function. The hidden state, which holds data from all previous steps, is a looping mechanism introduced by RNNs that allows past information to be passed forward. In contrast to conventional architectures, this enables information to remain in RNNs. RNNs have a feedback mechanism that allows them to train their memory for time-varying patterns either from the hidden to the input layer or from the output to the input layer (Ewuzie et al., 2022).

According to the daily forecasts, as it is shown in Figure 5, the LGBM model fits the time series data better (RMSE: 47.73). The LGBM model uses a histogram-based approach to reduce the effects of high dimensional data, speed up computation, and guard against overfitting the forecasting system (Massaoudi et al., 2021). The LGBM models also contain relatively higher complexity since, according to our setup, one model is built per each step in the horizon (Hewamalage et al., 2022).

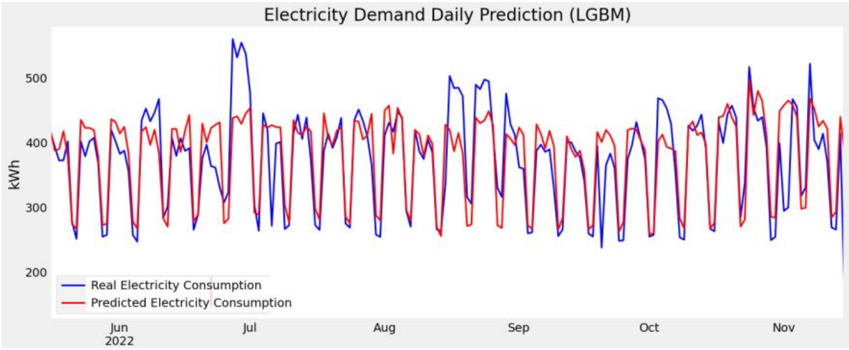

Figure 5: Electricity Daily Demand Prediction - LGBM Model

There is a bigger margin for improvement in the daily forecasts and the RF model appears to respond considerably better to the daily frequency compared to its hourly state. The external components were also included to enhance the models' fit with the series data. The exterior temperature was the only variable that had an impact on the model's performance (enhancing the RMSE). For comparison, in Figure 6, graphs representing models a) before (RMSE: 49.9) and b) after the introduction of the current outdoor temperature are displayed (RMSE: 45.3).

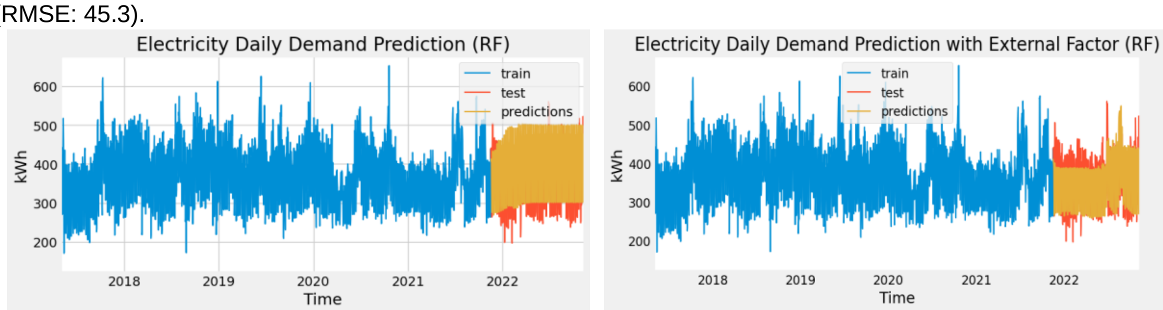

Figure 6: Electricity Daily Demand Prediction a) Simple RF model, b) Rf model with external factor

A bagging framework and an independent Decision Tree (DT) are the two components of the supervised method known as Random Forest. The essential component of RF is the integration of DTs, which generates several DTs by randomly generating the dataset's column variables and row values before averaging the DTs' outputs. A single DT is challenging to predict precisely, but when all DTs are combined to form a forest, the aggregated findings can consider the results of all DTs, leading to a more precise overall prediction. Finally, a backtesting with a refit strategy is followed. The model was trained each time before making predictions. It is a variation of the standard cross-validation. Instead of making a random distribution of the observations, the training set increases sequentially, maintaining the temporal order of the data (El Maghraoui et al., 2022).

3. Results

Analysis of the time series data identified a trend and a seasonality in the data. Seasonality appeared with a daily, weekly, and annual pattern. Their extensive evaluation showed the critical points. According to the data analysis, the UCM, Arima, LGBM, LSTM, Prophet, RF, Ridge, and XGBOOST models were examined. In Figure 7, it is easy to benchmark each model according to its performance in the RMSE index, as shown both on the left figure with the hourly forecasts and on the right with the daily predictions. Their application showed that significant predictions could be made for the energy consumption of the business building. Considering the data in an hourly format, the LSTM model seems more reliable, showing the smallest and particularly satisfactory RMSE index (0.53).

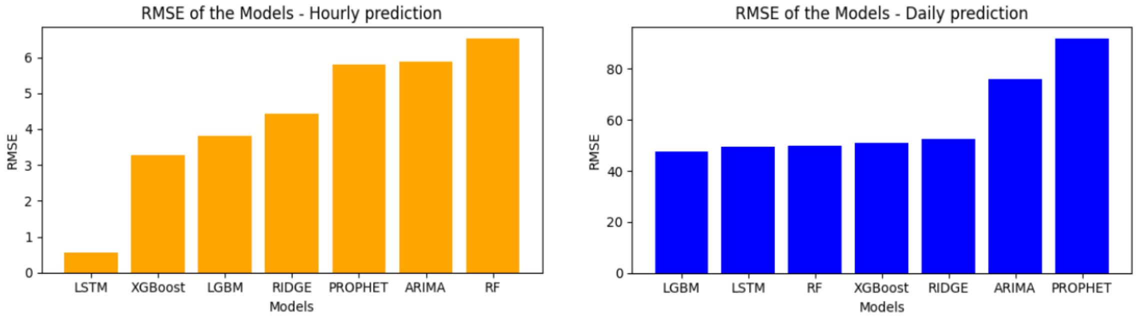

Figure 7: Comparison of the models' effectiveness - RMSE index in a) hourly and b) daily predictions

The effectiveness of the models to be applied to the time series data decreased in the evaluation based on the daily data. However, again according to the RMSE index, the fit of the model to the data is significant and show that valuable predictions can be made. In the daily form, the LGBM, LSTM and RF model showed the best performance. Arima and prophet models seem to show the least satisfactory performance in both cases. To improve mostly the performance of the daily predictions, the addition of external factors was implemented, trained and tested in the RF model. Humidity and wind speed had no influence on energy consumption so including them in predictions simply increases the computational complexity without improving the performance of the prediction model. Nevertheless, it was shown that the ambient temperature improves the effectiveness of forecasts in the hourly model by 7 %, while in the daily model, the improvement reaches 10 %.
The main problem with predictions based on external factors is, for example, in this case, the need for a reliable source for correct future temperature predictions. There are studies based on real-time data and existing buildings that prove that the electricity consumption in high-performance buildings is independent of ambient air temperature (Rios et al., 2017). On the other hand, research in a larger field analysis of the correlation of factors shows that seasonality, ambient air temperature, and daylight hours directly affect the change in electricity volumes (Shklyarskiy and Batueva, 2019). The dependence of consumption on the ambient temperature is due to a variety of factors, such as the type and age of the buildings, the climate of the area, and the cooling and heating systems.

4. Conclusions

The changes in electricity demand profile are directly affecting the efficiency and, in some cases, the stability of the systems. At the same time, the increased penetration of Renewable Energy Sources in Europe's electrical systems makes the latter face new challenges. It is necessary to precisely forecast the consumption patterns to develop suitable energy management systems. The current ML study was developed in order to train the corresponding prediction models. With five years of historical data on energy demand of a business building, various model applications were examined. On hourly and daily form, with and without external factors, the algorithms were tested. Significant findings highlighting the role of ML in energy demand forecasting, and comparisons were made to choose the best model based on how well it performed in each scenario.
The findings suggest that to better handle hourly predictions in energy usage, LSTM model, a deep learning architecture based on an artificial recurrent neural network (RNN), is selected. In the analysis of the daily data, the performance of the models to be applied to the time series data dropped. Gradient Boosting frameworks, which are based on decision trees, demonstrate that it is possible to make worthwhile predictions. Finally, it was shown while checking the performance of the RF model that the ambient temperature increases forecast accuracy in the hourly model by 7 % and in the daily model by 10 %. The importance of outdoor temperature in the performance of the model is affected by many factors, such as the climate of the area, the type of building, and its heating/ cooling systems. In the next steps of the research, it will be crucial to develop a protocol for the best ML model based on the forecast horizon, the quantity of historical data, and the type of building. Future work will eventually aim to generate predictions with real-time data from a smart meter platform, incorporating additional external parameters.

Acknowledgments

This research has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie Actions (MSCA) Research and Innovation Staff Exchange (RISE) grant agreement No.101007641, project eUMaP - Development of Utilities Management Platform for the case of Quarantine and Lockdown.

References

Candanedo I.S., Nieves E.H., González S.R., Martín M.T.S., Briones A.G., 2018, Machine Learning Predictive Model for Industry 4.0. In: Uden, L., Hadzima, B., Ting, IH. (Eds) Knowledge Management in Organizations, KMO 2018, Communications in Computer and Information Science, 877, Springer, Cham, Switzerland.

Chou J.S., Tran D.S., 2018, Forecasting energy consumption time series using machine learning techniques based on usage patterns of residential householders, Energy, 165, 709-726.

Crisóstomo H., Abílio O, Ferreira O.L., Pozzobon P., Rebeca M., Olino A., Rosa S.C., Guimarães Ferreira P.H., Fontes Guimarães R., Trancoso Gomes R.A., 2020, Rice Crop Detection Using LSTM, Bi-LSTM, and Machine Learning Models from Sentinel-1 Time Series, Remote Sensing, 12, 2655.

El Maghraoui A., Ledmaoui Y., Laayati O., El Hadraoui H., Chebak A., 2022, Smart Energy Management: A Comparative Study of Energy Consumption Forecasting Algorithms for an Experimental Open-Pit Mine, Energies, 15, 4569.

Ewuzie U., Bolade O.P., Egbedina A.O., 2022, Chapter 9 - Application of deep learning and machine learning methods in water quality modeling and prediction: a review, In Intelligent Data-Centric Systems, Current Trends and Advances in Computer-Aided Intelligent Environmental Data Engineering, Academic Press,185-218, Cambridge, Massachusetts, United States of America.

Hewamalage H., Bergmeir C., Bandara K., 2022, Global models for time series forecasting: A Simulation study, Pattern Recognition, 124, 108441.

Karijadi I., Chou S.Y., 2022, A hybrid RF-LSTM based on CEEMDAN for improving the accuracy of building energy consumption prediction, Energy and Buildings, 259, 111908.

Khan P.W., Byun Y.C., Lee S.J., Kang D.H., Kang J.Y., Park H.S., 2020, Machine Learning-Based Approach to Predict Energy Consumption of Renewable and Nonrenewable Power Sources. Energies, 13, 4870.

Lee M.H.L., Ser Y.C., Selvachandran G., Thong P.H., Cuong L., Son L.H., Tuan N.T., Gerogiannis V.C., 2022, A Comparative Study of Forecasting Electricity Consumption Using Machine Learning Models, Mathematics, 10, 1329.

Massaoudi M., Refaat S.S., Chihi I., Trabelsi M., Oueslati F.S., Abu-Rub H., 2021, A novel stacked generalization ensemble-based hybrid LGBM-XGB-MLP model for Short-Term Load Forecasting, Energy, 214, 118874.

Rios C.F., Naganathan H., Chong W., Lee S., Alves A., 2017, Analyzing the Impact of Outside Temperature on Energy Consumption and Production Patterns in High-Performance Research Buildings in Arizona, Journal of Architectural Engineering, 23, C4017002.

Rodrigo J.A., Ortiz J.E., 2021, Skforecast: time series forecasting with Python and Scikit-learn, available under an Attribution 4.0 International license (CC BY 4.0), <cienciadedatos.net/documentos/py27-time-series-forecasting-python-scikitlearn.html>, accessed 23.02.2023.

Shapi M.K.M., Ramli N.A., Awalin L.J., 2021, Energy consumption prediction by using machine learning for smart building: Case study in Malaysia, Developments in the Built Environment, 5, 100037.

Shklyarskiy J.E. Batueva D.E., 2019, The influence of external climatic factors on the accuracy of the forecast of energy consumption, EECE, E3S Web of Conferences, 140, 04014.