**Hungarian Electricity Demand Forecasting using LSTM and ARIMA Methods**

Shahab Sabzi

Budapest University of Technology and Economics

**Abstract**:

Forecasting electricity demand has an important role in scheduling future programs in the network and regulating the capacity, dealing with demand-side management, and future market regulation. Therefore, In this study, two forecasting models are implemented on the Hungarian electricity demand dataset. First, a deep learning LSTM model is created to forecast the next sample, and then the obtained results are compared with an ARIMA model. In the end, the simulation results verify the superiority of LSTM over the ARIMA model, with better precision.

**Introduction (and literature review)**

[this section is not complete now and will be completed later with a full literature review]

Data collected over a period of time are called time series. Trend (long term pattern), cyclical (repeated ups and downs), seasonal (regular fluctuations occurring within the same month or quarter) and irregular components (unexplained random fluctuations) are all components of time series. [1].

Using historical and current data, forecasting makes predictions about future values. Planning and controlling future power systems require a forecast of future load demand. Load forecasts can be classified into three categories: short-term load forecasts (STLF), medium-term load forecasts (MTLF), and long-term load forecasts (LTLF). LTF forecasts more than a year in advance, while STLF forecasts for a day to a week at most [2]. In STLF, an electric utility schedules generation and transmission of electricity, while in MTLF, fuel is purchased, and LTLS is responsible for the development of the power distribution and delivery systems (generation, transmission, and distribution) [3].

Electricity demand forecasting is not a new subject and has been carried out many times in the literature. There are various forecasting methods available in the literature such as multiple linear regression, time series model, Autoregressive integrated moving average (ARIMA) process, artificial neural network, Fuzzy time series, Fuzzy neural network, etc. Article [4] proposes an Adaptive Neuro-Fuzzy Inference System (ANFIS) to predict the electricity demand. While the used dataset is the electricity demand of Ghana from 2003 to 2018, different scenarios are implemented for validation. In [5], authors used time series models, including simple moving average (SMA), weighted moving average (WMA), simple exponential smoothing (SES), Holt linear trend (HL), Holt-Winters (HW) and centred moving average (CMA) have been used for the electricity forecasting action.

An empirical mode decomposition and state-space model are combined in [6] in order to improve the robustness and accuracy of electricity demand forecasting, for which In order to forecast all sub-series (feature extraction), empirical mode decomposition is applied to the entire time series first (noise filtering). Then state-space model parameters are optimized using maximum likelihood via a Kalman filter. A paper [7] models electricity demand in European countries using ARIMA, periodic AR, a method based on the principal component analysis (PCA) of demand profiles, and an extension for double seasonality of Holt-Winters exponential smoothing. Authors in [8] use the SARIMA model for weekly peak demand forecasting in Australia.

Several modelling techniques are compared in [9], including techniques that can capture the specific dynamics of demand time series. In this regard, there are two major components of electricity demand time series: deterministic and stochastic. Authors estimate the two components using regression and time series techniques with both parametric and nonparametric approaches. The study also provided several linear regression methods (such as tri-cubic, Gaussian, and Epanechnikov), smoothing splines, regression splines, as well as traditional time series models (autoregressive moving averages and nonparametric autoregressions). Furthermore, they apply several spline function-based models (such as smoothing splines, regression splines, and vector autoregressions).

**Data collection and preprocessing**

The data for this study is obtained from the [ENTSO-E Transparency](https://transparency.entsoe.eu/load-domain/r2/totalLoadR2/show?name=&areaType=CTY&biddingZone.values=CTY|10Y1001A1001A83F!CTY|10Y1001A1001A83F) of Europe [9]. The data is made of 15-minute intervals from the beginning of 2020 until Dec. 31 2020, made of 35136 data points in total, which is then converted to hourly data by neglecting data points between hours, leaving us with 8784 data points altogether. Figure 1 shows some useful information about the dataset.

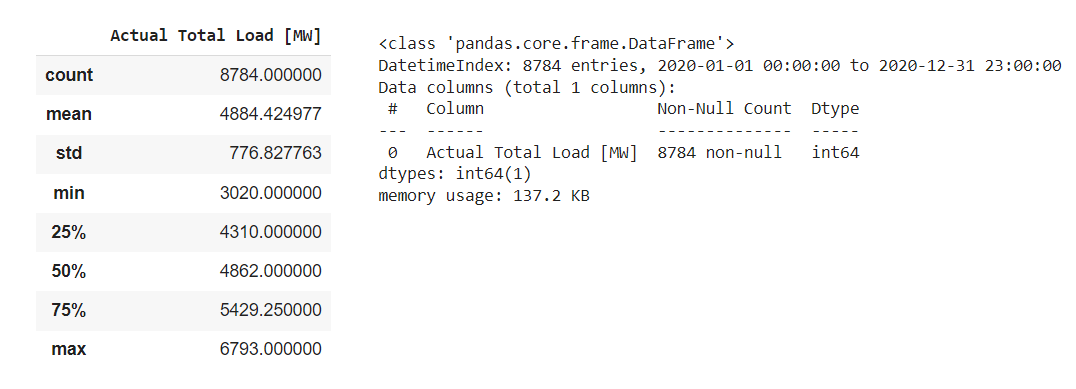


Figure 1. Some information about the dataset

Figure 2 shows the demand curve for the year 2020 based on hourly data and its moving average with averaging of 24\*7 data points, which is weekly.

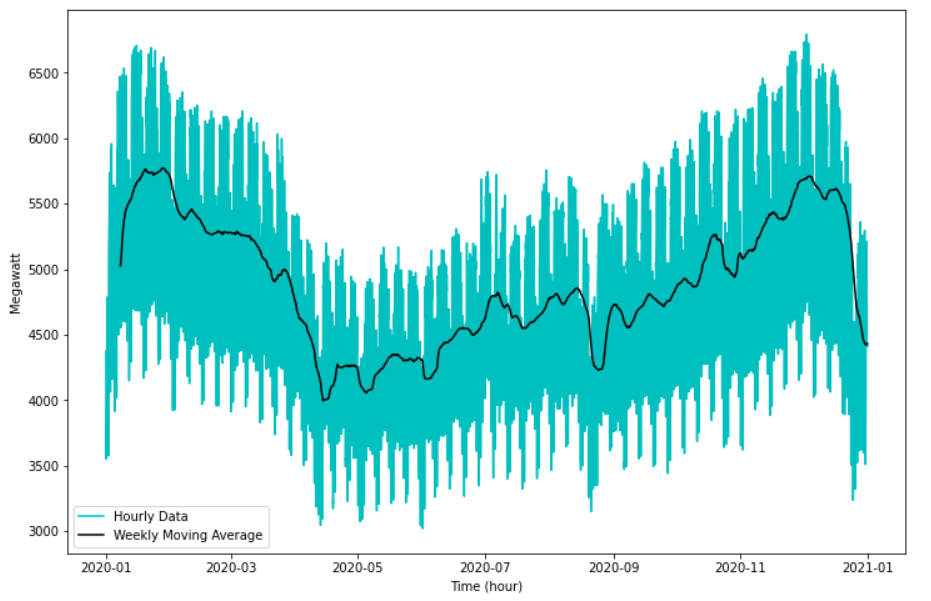


Figure 2. demand curve for the year 2020 based on hourly data and its moving average

**LSTM model**

[Formulation and mathematical modelling will be added later in the full article]

For the LSTM, windows with a size of 60 sample have been used, meaning that the next sample is forecasted based on the past 60 samples. Model is sequential with 16 layers in the input and 1 layer in the output. The used activation function is linear and for the compiler, loss model is Mean Squared Error, the optimizer is Adam, and metric for evaluation is MAE. Early stopping is applied to the model to stop the algorithm whenever there is no improvement in the validation loss after 5 epochs. Figure 3 shows the Loss curve for test data and train data. As seen, both curves converge to around zero after about 5 to 10 epochs, and after 23 epochs, early stopping is activated and stops the algorithm from running further.

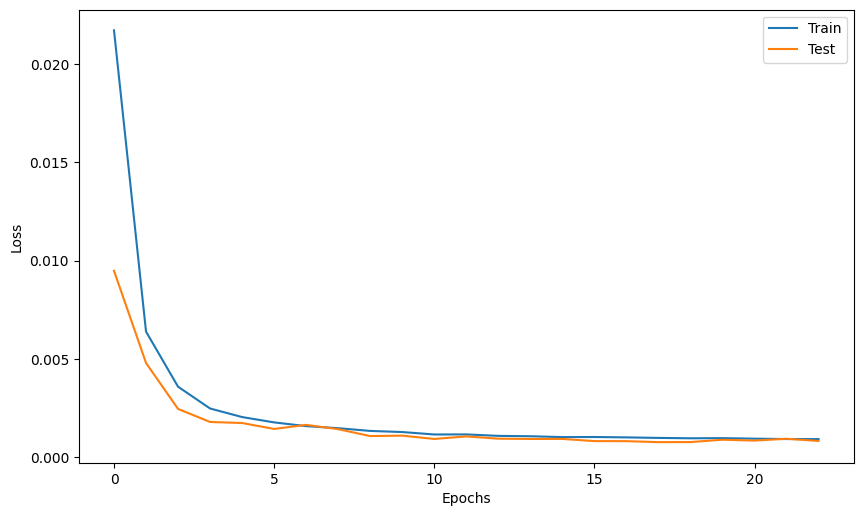


Figure 3. Loss curve for test data and train data

Figure 4 show the real curve and predicted curve for the whole year period.

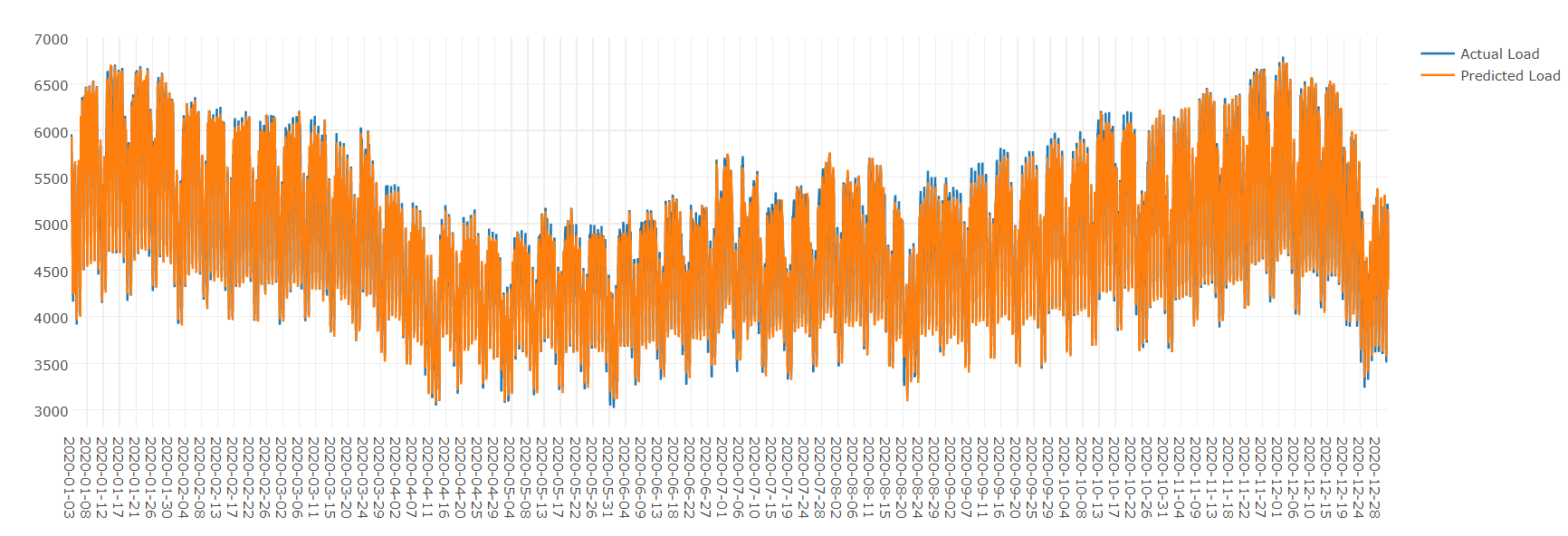


Figure 4. Real demand and predicted demand

Figure 5 shows the predicted curve and real data for the last three days of the year.

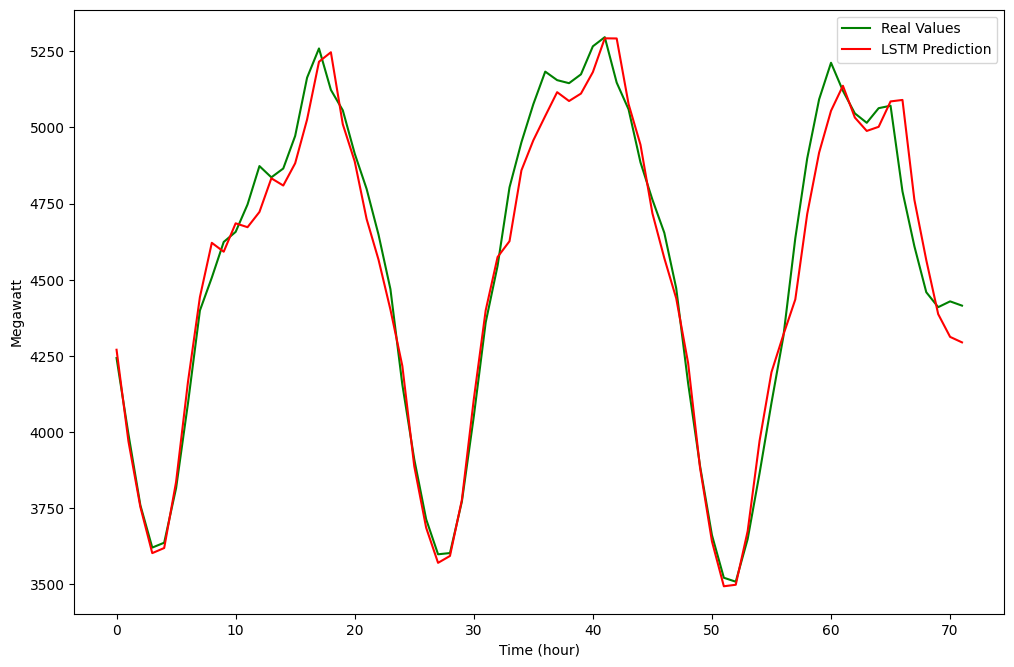


Figure 5. predicted curve and real data for the last three days of the year

**ARIMA model**

[Formulation and mathematical modelling will be added later in the full article]

For the ARIMA model, only last month's data (December 2020) is used. 28 days are used to train the model and the last three days (made of 72 samples) are forecasted. Figure 6 demonstrates the train data and test data for the ARIMA model. Also, figures 7-9 show some attributes of the data, including seasonality, trend and residuals.

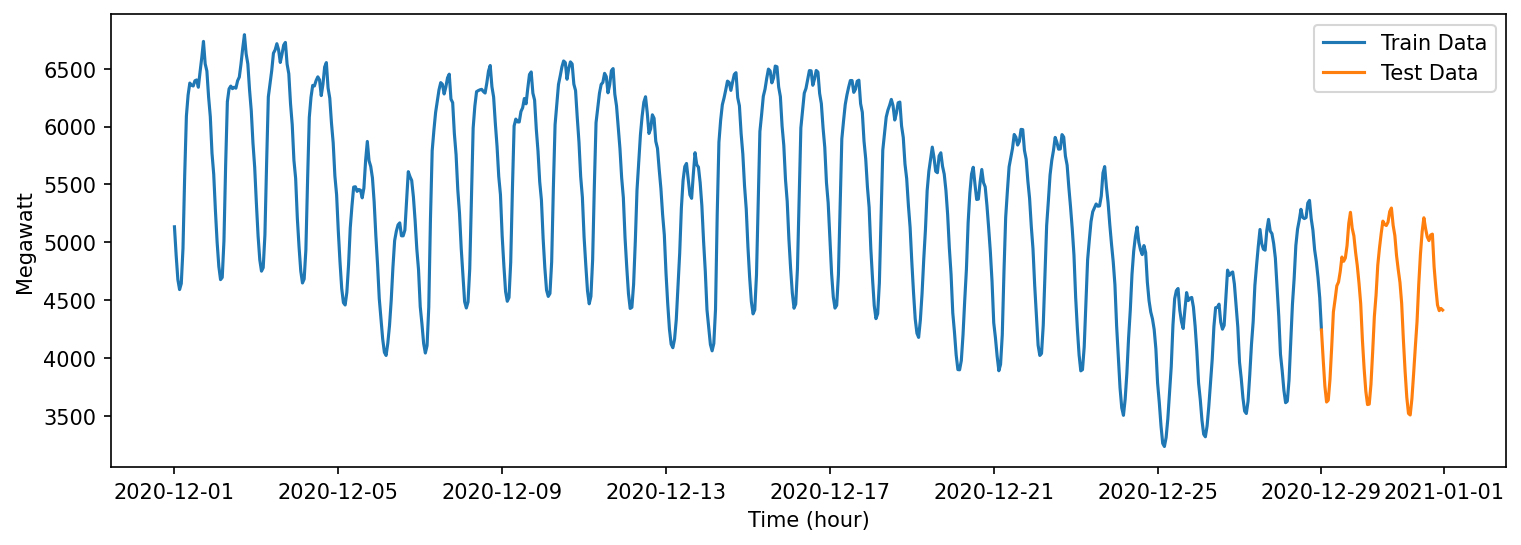


Figure 6. train data and test data for the ARIMA model

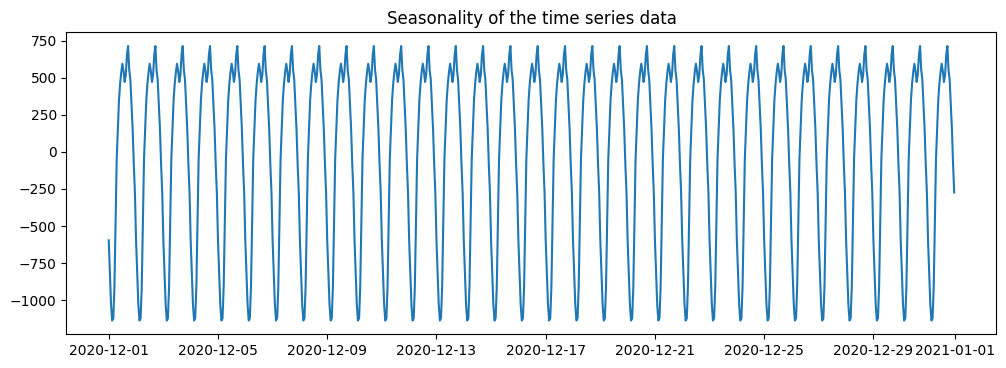


Figure 7. Seasonality of the input dataset

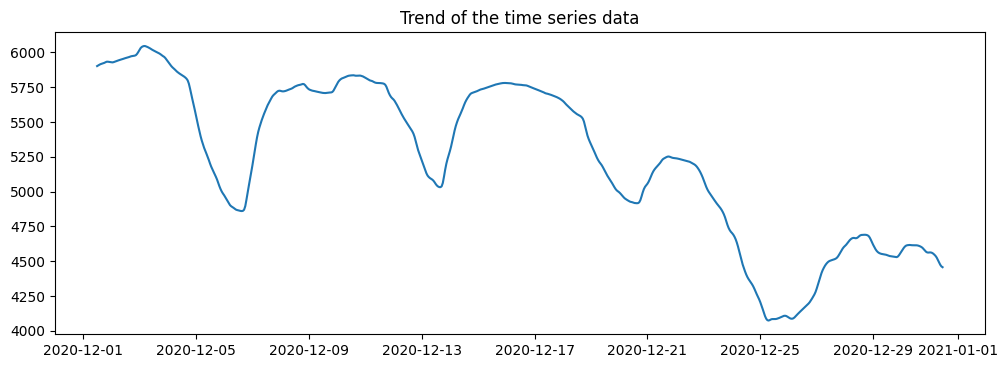


Figure 8. Trend of the input dataset

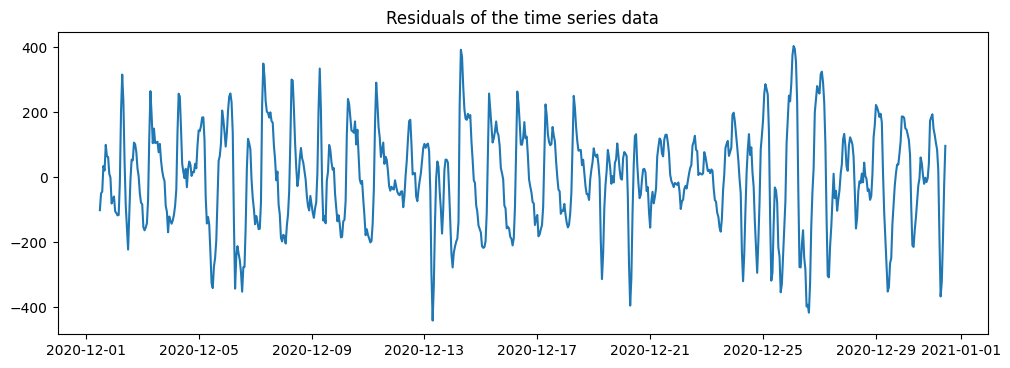


Figure 9. Residuals of the input dataset

As seen, the model is seasonal, and method Seasonal ARIMA (SARIMA) must be used to forecast the future.

In order to find the coefficients of the SARIMA model, the auto\_arima function in pmdarima package is used, and parameters are obtained as ARIMA(3,1,1)(1,1,2)[12]. Then the trained model is used to predict the next 72 samples in the future. Figure …. Shows the real values and predicted values using the ARIMA model.

Figure 10 show the real values and predicted values of the trained ARIMA model.

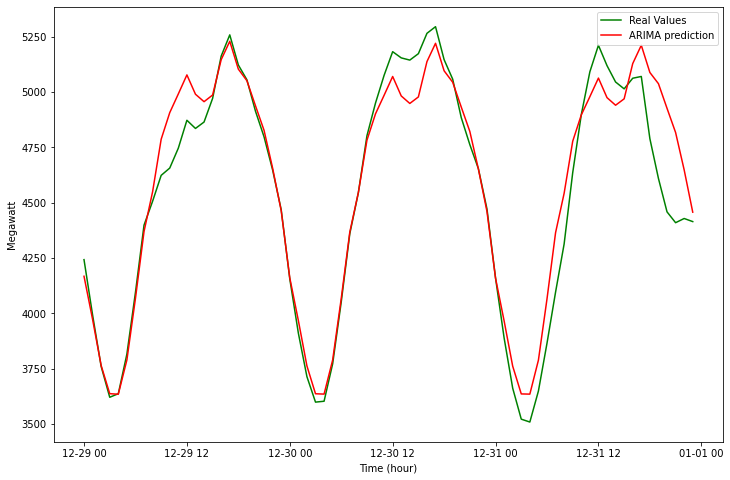


Figure 10. real values and predicted values with ARIMA model

**Comparison of LSTM and ARIMA models**

Figure 11 shows the real values and predicted values by LSTM and ARIMA models.

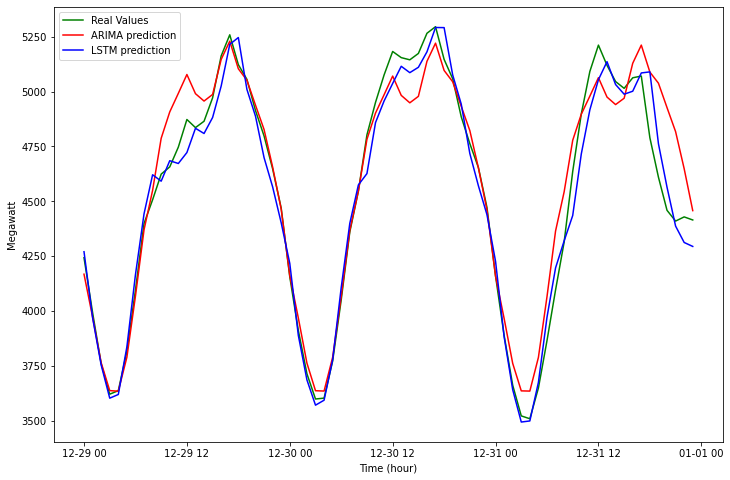


Figure 11. real values and predicted values by LSTM and ARIMA models

The figure shows that both models are doing relatively well in predicting the future; however, LSTM is relatively better. Also, in terms of Root Mean Squared Error (RMSE), LSTM has an RMSE of 89.405, while this value for ARIMA is 142.267, indicating that LSTM is doing better.

**Conclusion**

In this study, a time series forecasting of Hungarian electricity demand was carried out using the LSTM method, and then the results were then compared with the seasonal ARIMA method. The dataset included one year of hourly data for the LSTM and one month for the ARIMA model. The LSTM model forecasted the next sample about the past 60 samples, while in the ARIMA model, 28 days of data were used for the training, and the last three days were used for the validation. While both methods provided acceptable results, LSTM presented better accuracy by a lower root mean squared error.

**References**

[1] G. Tunnicliffe Wilson, “Time Series Analysis: Forecasting and Control,5th Edition, by George E. P. Box, Gwilym M. Jenkins, Gregory C. Reinsel and Greta M. Ljung, 2015. Published by John Wiley and Sons Inc., Hoboken, New Jersey, pp. 712. ISBN: 978-1-118-67502-1,” *J. Time Ser. Anal.*, vol. 37, p. n/a-n/a, Mar. 2016, doi: 10.1111/jtsa.12194.

[2] D. Pedregal and J. Trapero, “Mid-term hourly electricity forecasting based on multi-rate approach,” *Energy Convers. Manag. - ENERG CONV MANAGE*, vol. 51, pp. 105–111, Jan. 2010, doi: 10.1016/j.enconman.2009.08.028.

[3] E. Almeshaiei and H. Soltan, “A methodology for Electric Power Load Forecasting,” *Alex. Eng. J.*, vol. 50, no. 2, pp. 137–144, Jun. 2011, doi: 10.1016/j.aej.2011.01.015.

[4] A. Acakpovi, A. T. Ternor, N. Y. Asabere, P. Adjei, and A.-S. Iddrisu, “Time Series Prediction of Electricity Demand Using Adaptive Neuro-Fuzzy Inference Systems,” *Math. Probl. Eng.*, vol. 2020, p. e4181045, Aug. 2020, doi: 10.1155/2020/4181045.

[5] Y. W. Lee, K. G. Tay, and Y. Y. Choy, “Forecasting Electricity Consumption Using Time Series Model,” *Int. J. Eng. Technol.*, vol. 7, no. 4.30, Art. no. 4.30, Nov. 2018, doi: 10.14419/ijet.v7i4.30.22124.

[6] Z. Hu, J. Ma, L. Yang, L. Yao, and M. Pang, “Monthly electricity demand forecasting using empirical mode decomposition-based state space model,” *Energy Environ.*, vol. 30, no. 7, pp. 1236–1254, Nov. 2019, doi: 10.1177/0958305X19842061.

[7] J. W. Taylor and P. E. McSharry, “Short-Term Load Forecasting Methods: An Evaluation Based on European Data,” *IEEE Trans. Power Syst.*, vol. 22, no. 4, pp. 2213–2219, Nov. 2007, doi: 10.1109/TPWRS.2007.907583.

[8] A. Eshragh, B. Ganim, T. Perkins, and K. Bandara, “The Importance of Environmental Factors in Forecasting Australian Power Demand,” *ArXiv191100817 Stat*, Oct. 2021, Accessed: Dec. 11, 2021. [Online]. Available: http://arxiv.org/abs/1911.00817

[9] “ENTSO-E Transparency Platform.” https://transparency.entsoe.eu/dashboard/show (accessed Dec. 11, 2021).