Load Curve Forecasting using LSTM: a case study at Federal University of Paraiba

Gabryel M. Raposo de Alencar * **Camila Seibel Gehrke** *
Flávio da Silva V. Gomes * **Lucas Vinicius Hartmann** *
Camila Maciel *

* *Electrical Engineering Department and Renewable Energy Engineering Department, Federal University of Paraíba, PB*

E-mail: gabryel.alencar@cear.ufpb.br, camila@cear.ufpb.br, flavio@cear.ufpb.br, lucas.hartmann@cear.ufpb.br, c.andrade@cear.ufpb.br

Abstract: Energy management systems (EMS) has importance to utilities for several reasons, mainly for existing a necessity to allocate their resources in advance, requiring short, medium and long-term planning. Therefore, in this paper, a very short-term demand forecasting procedure was implemented, using a computational model based on Artificial Neural Networks (ANN) of the type Long Short-Term Memory (LSTM) to aid the Federal University of Paraíba (UFPB) analyzing peaks and off-peaks of active power throughout the past years being possible to use the forecast of One Hour Ahead as a input to a BESS for optimal battery power dispatch management, in order to reduce fines for excess contracted demand at UFPB. The study compares LSTM with a CNN model and to improve the performance of the LSTM, the network was evaluated under different aspects, as hyperparameters and considered a technique of periodicity using sines and cosines.

Keywords: Time Series, Electric Power, Artificial Neural Networks, LSTM, CNN, Energy Optimization, EMS.

1. INTRODUCTION

Throughout the world development there were plenty of energy needs, which are increasing dramatically in the past few years. Nonetheless, when contracting an energy demand from a utility, if it is exceeded, which is commonly at peak hours, a fine must be paid. However, with an ideal fixed battery storage system there is the possibility of minimizing this fine. These energy stores can be charged during off-peak hours (low electricity price periods) and discharged during peak hours (high electricity price periods). Energy management systems with their mainly functions of monitoring, optimization and use of electrical power, can reduce peaks, which helps the university to reduce energy costs and utilities to properly utilize their systems.

The optimal energy management of an electrical grid, can be assisted by a effective prediction of appropriate data, as reactive, active power and load consumption. In renewable generation, also weather variables such as wind speed, temperature, pressure are commonly used. [1]-[3]. Further, studies have proved adding the forecasting capability to an EMS enables improvement planning the power allocation [4].

Substantial body of studies in the field is concerned with using different methods of forecasting and compare the results. [5] Various forecasting methods are proposed by authors in the academic field including Convolutional Neural Networks (CNNs), fuzzy logic in [6], support vector machines (SVMs) in [7], nonlinear autoregressive neural network (NARX) in [8], and the most common LSTM Neural Network [9]. On the other hand,even many works in the literature have studied vast techniques varying the climatic variables and load demand, papers about the implementation of a forecasting module considering this seasonality as cosines and sines, related to EMS are still missing.

This article proposes a solution for this coordinated demand management being carried out using a LSTM Artificial Neural Network (ANNs) to forecast the load demand with the Very Short-Term model being implemented. The LSTM has been chosen because plenty works presented by the academic field demonstrates that it can efficaciously be used in time-series modeling thanks to its neurons configurations with gates which promotes quick learning process, and that it produces accurate results. To confirm this preposition, a comparison with CNNs has been done in this present paper.

Considering the seasonality of the time series, two new features were used as inputs in this project, the sine and cosine function, representing the variation and periodicity of the load, day after day. This new approach resulted in better results that will be discussed in the following paragraphs.

The Short-Term prediction uses 7 days to forecast the next one, whereas the very Short-Term uses 1 day to forecast the next hour. Such energy forecast is also used to economically charge and discharge battery storage to reduce electricity consumption during high price. These daily and hourly forecasts of energy demand generation help to effectively manage the real-time operation of electronic devices and the cycle of energy resources. Therefore, initially a brief analysis is made of the typical demand profile of the university, prioritizing the understanding of the growth in the use of energy over the years and subsequently implemented the application of machine learning. Next, the methodology used is described, the results obtained are analyzed and conclusions about this work are presented.

2. ARTIFICIAL INTELLIGENCE

2.1 Long Short-Term Memory Networks (LSTMs)

The LSTMs are special variations of recurrent networks, where they are able to learn long-term dependencies, remembering information for longer periods and not presenting the problem of gradients[10]. Like conventional recurrent networks, LSTMs also have their structure linked to chains of repetitions, however, they differ in complexity. The main difference is the presence of gates (input and output gate) with the ability to remove or add information to their state, in other words they are an optional way of allowing information and are composed of a sigmoid network layer. Later the configuration was improved by Gers in [11] when it was proposed a new cell gate, the forget gate. The LSTM network has several applications such as in Time Series, Natural Language Processing and Automatic Text Generation

2.2 Working of LSTM cell

The LSTM cell is shown in the following figure:

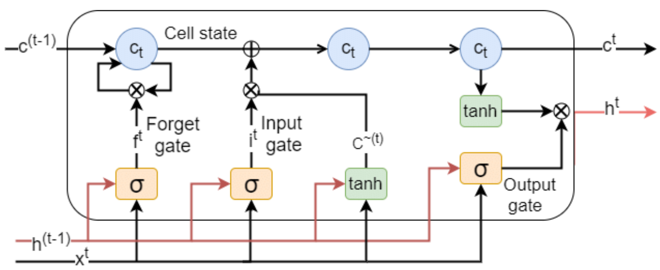

Figure 1. LSTM cell Internal Architecture

The learning process in this cell occurs in basically three steps:

[1] First Step : On the Forget Gate, happens the identification of unnecessary information, that is not useful in the state of the unit which will be discard. As it illustrates at the figure [1] the input information is a combination of the bias value bf, the input X_t at current moment t and previous output $h_{(t-1)}$ at time t − 1 (hidden state). This can be evidenced in the formula below:

$$f_t = \sigma(W_f[h_{(t-1)}, X_t] + b_f) \tag{1}$$

[2] Second Step : On the Input Gate, updating useful information to the cell state is done in two different parts. The first one regards on the sigmoid function calculating the old portion of information that will follow to the next step, and the second concerns about the tanh function. The combination of the information that passed through those phases is multiplied and added to the next cell c_t. These formulas below describes the process:

$$i_t = \sigma(W_i[h_{(t-1)}, X_t] + b_i) \tag{2}$$

$$N_t = tanh(W_n[h_{(t-1)}, X_t] + b_n) \tag{3}$$

$$C_t = C_{(t-1)} f_t + i_t N_t \tag{4}$$

[3] Third Step : On the output Gate, the task of extracting useful information from the state of the current cell (C_t) to be presented as an output value (h_t)is done. Hence, the output cell state C_t is actualized with the old information, to produce the new output for the next module. These formulas below describes the process:

$$O_t = \sigma(W_o[h_{(t-1)}, X_t] + b_o) \tag{5}$$

$$h_t = O_t tanh(C_t) \tag{6}$$

3. METHODOLOGY

In order to assess the level of load demand at the University, LSTMs, a deep learning algorithm, which is subtly used in time series forecasting, was implemented. The application of this type of technique involves carrying out a typical sequence of steps, which can be observed in the flowchart shown in Figure 2:

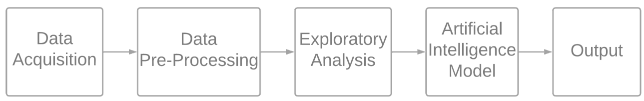

Figure 2. Diagram for application of machine learning model

3.1 Data Acquisition

The power measurement acquisition is performed by equipment from the utility at the university, which generates measurements of active and reactive power every 5 minutes.

3.2 Data Pre-Processing

At this stage, the data are processed so that their processing in Artificial Neural Networks (ANNs) can be facilitated. From this perspective, it was necessary to deal with the detection of outliers and, also, the creation of new columns that tend to be important for the interpretation of the ANN. At first, an outlier verification plotting the boxplot was necessary for understand the Data, as shown below:

This way, it was possible to infer that there were certain outliers in the active power measurement that would need

ISSN: 2525-8311

DOI: 10.20906/CBA2022/3635

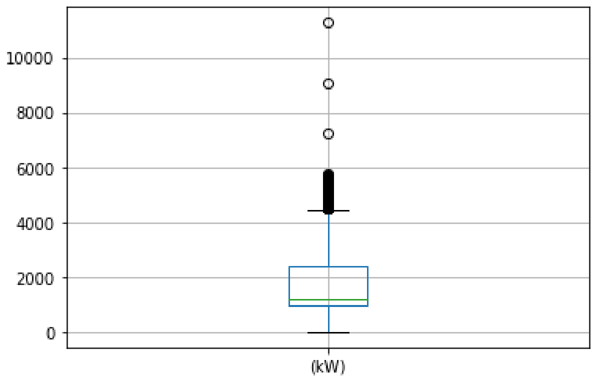

Figure 3. Boxplot Graphic showing the outliers

to be addressed, by verifying data that exceeded the upper limit of the boxplot above. In order to achieve a more concise identification of outliers, the Z-Score technique was used, which consists of measuring how much the measurement deviates from the mean in terms of standard deviation, based on the following graph:

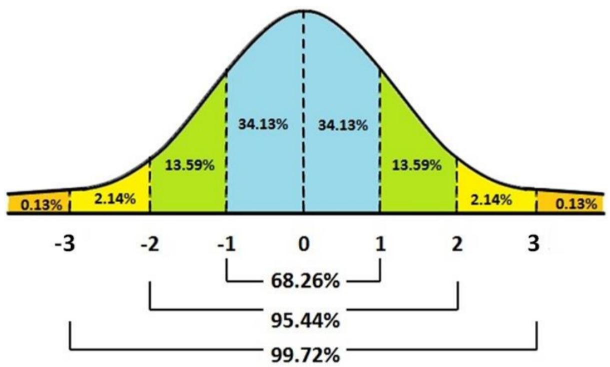

Figure 4. Table Z-score

After calculating the Z, most of the values are between -3 and 3, the classification as outliers is whoever was outside this, and, because they represent a minimal and tiny percentage of data in our DataSet (approximately 3% of the total), at first, it was chosen to just remove these values.

3.3 Exploratory Analysis

The data compile information between the years 2008 to 2019 containing information on the active and reactive power consumed by the University and were provided by the own institution.The data for the years 2020 and 2021 were not used because they were pandemic years with atypical behavior, in which the university remained closed or with little flow for most of the time. An example of their initial attributes and content are presented in the following table:

Table 1. Active and reactive power dataset

Time	Hour	(kW)	(kVAr)
2008-01-01 00:05:00	00:05	816.5	705.6
2008-01-01 00:10:00	00:10	806.4	685.4
2008-01-01 00:15:00	00:15	796.3	695.5
2008-01-01 00:20:00	00:20	786.2	685.4
2008-01-01 00:25:00	00:25	816.5	695.5

Based on the data described in the previous session, it was able to do a deeper analysis of information to characterize the problem. Therefore, using the creation of graphs, the following strategies were outlined for analysis:

Average the powers in the given time interval for all months of all years. For this, it was necessary to generate a new DataFrame in which each column represented the month and the lines represented a time in the period of 24 hours a day, thus, each value found represented the average power consumed at each time in that entire month. The graphical representation is below.

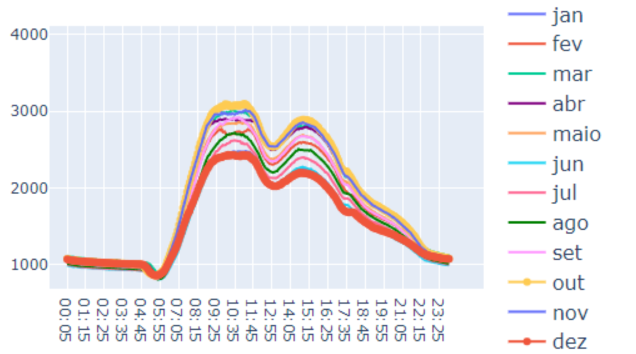

Figure 5. Average Monthly Electric Power Consumed by the University

From the Figure [5], is possible to understand the general functioning of the university in a monthly seasonality. In the stretch from 5h:20min to 6h:30min, the light poles are turned off, so there is a certain drop in power. Subsequently, there is a growth continued throughout the morning due to the arrival of students for classes, air conditioners and fans being turned on and conventional university operation, and a drop only at lunch time during the break from classes. At 2 pm, activities resume and power consumption starts to grow again. Finally, the trend at the end of the day is decrease as the flow of people on the night shift is considerably lower, with fewer evening undergraduate courses. However, in the period between 17:00 and 17:35 there is a slight increase due to the moment when the light poles are turned on again.

Another factor that can be observed when analyzing the graph was the fact that the month of October during the period of all years was the one that had the highest average power consumption and, as expected, the months with the lowest consumption were January and December due to to recess of classes at the university, with a lower flow of students. Even so, it was known that this alone was not necessary for a concise analysis of the data, then an annual perspective was started to analyze, representing the annual seasonality in another graph, using the same logic of the monthly graph. The representation is shown below:

And finally, in a third graph, the total average daily power consumed in each year:

Finally, the same daily periodic behavior seen in the first graph was confirmed, and also the situation of the growth of power expenditure at the University over the initial years due to the expansion of the Reuni program, the decrease in 2015 by the strike and, finally, a greater

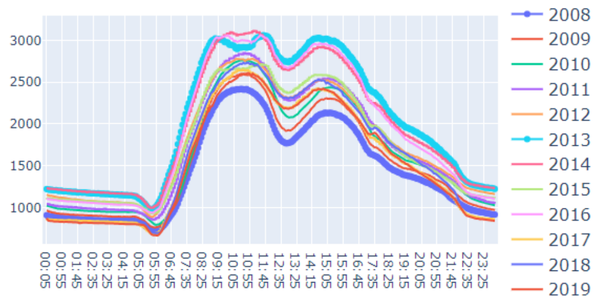

Figure 6. Average Annual Electric Power Consumed by the University

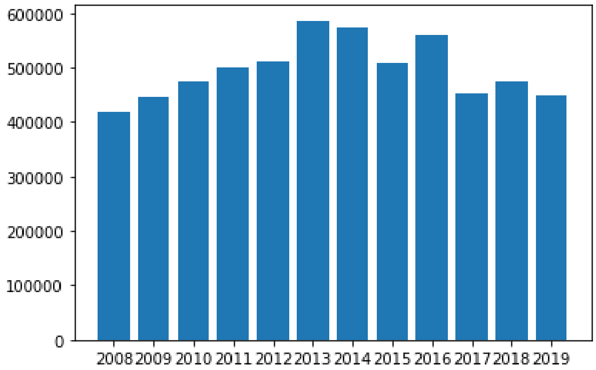

Figure 7. Total power consumed in each year considering a daily average

decrease from 2017 onwards due to the shutdown of the University Hospital (HU) in power measurements.

From this perspective, it became evident that it is a time series problem, in which a certain seasonality with a fixed trend, both daily and yearly.

3.4 Periodicity

As a time series problem in which events were repeated periodically, an annual, weekly and daily representation through sine and cosine was proposed, once these trigonometric functions are periodic functions, which are repeated over time. Further, with the addition of such columns in the DataFrame and having measurements every ten minutes, it was evident that the period of the function for representation in two day (48 hours) would be 288, as shown in the following graph:

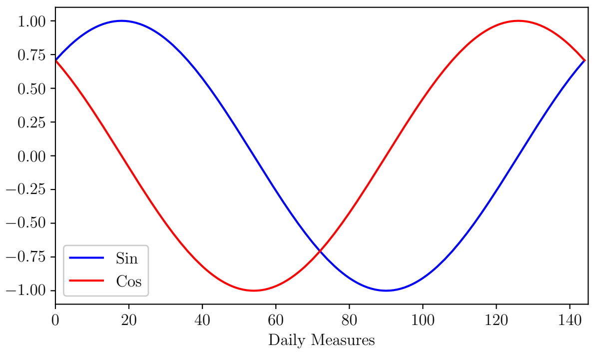

Figure 8. Sine and Cosine Periodicity.

4. SYSTEM ARCHITECTURE

In this paper, the operation and scheduling of the BESS are optimized by using the very short-term load forecast corrective actions. Moreover, by adding the real-time evaluation of the forecasted load values, it is possible to reoptimize the optimal management strategies and the operation cost would be decreased. In this section, the architecture of the model implemented is explained.

From this perspective, after understanding the initial concepts, it is possible to understand the use of resources in the construction of the ANN and the prediction of load. For similar situations in [12] the forecasting is perfomed using Convolutional Neural Networks (CNNs), in [13] with ANNs feed-forward and in [14] using LSTMs. None of them proposes the use of periodicity for forecasting. Thus, in this paper the network model used in load forecasting was the LSTM, justified by the presence of a large amount of data, since LSTM networks do not suffer from the problem of vanishing the gradient value and because they are an option with high accuracy in time series problems and compared with CNNs.

The very short-term load forecasting is performed with two days of input data in order to predict the hour-ahead. In Fig. 9, the starting procedure of the first Load Forecast and its structure have been shown.

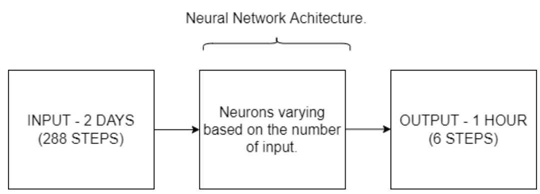

Figure 9. One Hour-Ahead Prediction - Architecture.

The project was implemented in the python programming language, and in reason of the current configuration of the computer has a less qualified GPU than the internal GPU offered by Google Colab, it was chosen to use google colab itself to create and improve the code.

Throughout the predictions, the percentage of data used in training and testing was updated and remodeled several times to identify a suitable model of ANN, until fixed it in 80 percent for training, 10 percent validation and 10 percent for testing. Thus, about 20 tests were performed using the pre-selected Data Frame, with two networks, the first one with 1 normalized input and the second with 3 normalized inputs [-1,1], opting for the Adam optimizer, varying hyperparameters such as batch size, hidden layers and neurons in the hidden layers.

5. VALIDATION

After each ANN training, a validation is performed with a metric in order to understand which network, among all the configurations, achieves the smallest error according to the metric. Among the many options to choose from, the Root Mean Squared Error (RMSE), Mean Absolute Error (MAE) and Mean Squared Error (MSE) was used. The validation consists of simulating the trained network for

ISSN: 2525-8311

DOI: 10.20906/CBA2022/3635

the data sets that were not used in the training. At each step, the value that is predicted is fed back to the input sets so that the next values to be predicted are influenced by the previous prediction, just like in the real world.

The RMSE metric is calculated by the formula:

$$RMSE = \sqrt{\frac{1}{n}\Sigma_{i=1}^{n}\Big(\frac{d_i - f_i}{\sigma_i}\Big)^2} \quad (7)$$

The MSE metric is calculated by the formula:

$$MSE = \frac{1}{n}\Sigma_{i=1}^{n}(x_i - y_i)^2 \quad (8)$$

The MAE metric is calculated by the formula:

$$MAE = \frac{1}{n}\Sigma_{i=1}^{n}|x_i - y_i| \quad (9)$$

6. RESULTS

The tests were performed using the pre-selected Data Frame, with two different machine learning network models. The first one, using Convolutional Neural Network and the second using Long Short-Term Memory Neural Networks. Firstly, merely 1 input was introduced as feature, the active power, without the periodicity configuration of the network. Confirming that the LSTM networks obtained a better result in relation to the CNNs, the last test was using the new technique in LSTMs with 3 inputs, active power, cosine and sine. The metric for the evaluation of the system was the RMSE, MAE and MSE.

Regarding on the epochs of the training models, at first it was defined as 50 epochs and tested different types of configurations with the internal layers. In view of this, it was possible to observe a certain normalization and completion of learning around the 50. Finally, the tests performed with the best results were obtained in the Adam optimizer, using tanh function.

Concerning to the number of neurons, embryonic tests were fulfilled with a high complexity and along the validation of results, it was realized reducing the numbers of neurons and hidden layers would bring a better result for the model.

In this sense, the method to define the number of neurons was to vary the first hidden layer according to the input layer, inserting multiples from 1/8x to 1x of the input layer. The results using 1 Input feature in CNNs are illustrated below:

Table 2. Results - CNN 1 INPUT

Neurons	MAE	MSE	RMSE
36	0,004921	0,000368	0,019179
41	0,006606	0.000416	0.020415
48	0,005054	0,000343	0,018526
58	0,005160	0,000364	0,019092
72	0,004793	0,000361	0,019008
96	0,004984	0,000345	0,018585
144	0.005115	0.000356	0.018886
288	0.004813	0.000337	,0.018374

In order to achieve better results, the idea of implementing dropout in the actual architecture was used. The following table demonstrates the architectures used in tests with 1 input features and 0.1 of dropout.

Table 3. Results - CNN 1 INPUT - 0.1 dropout

Neurons	Dropout	MAE	MSE	RMSE
36	0,1	0,007890	0,000441	0,021003
41	0,1	0,007140	0,000373	0,019323
48	0,1	0,006732	0,000361	0,018999
58	0,1	0,006968	0,000422	0,020539
72	0,1	0,006378	0,000431	0,020767
96	0,1	0,006260	0,000367	0,019168
144	0,1	0,006878	0,000371	0,019252
288	0,1	0,005998	0,000363	0,019045

Analyzing the two tables is evident the use of dropout worsens the results of our model. The best configuration in this situation had 48 neurons in the 1 hidden layer, without dropout, obtaining the best value of two metrics, MSE and RMSE.

Further, following the mainly models configuration for forecasting in the academic field, the training of LSTM cells were done. To compare with CNNs models, the same architecture configuration was implemented. The results are below:

Table 4. Results - LSTM 1 INPUT

Neurons	MAE	MSE	RMSE
36	0,005894	0,000337	0,018365
41	0,004834	0,000347	0,0186376
48	0,005655	0,000339	0,018425
58	0,006063	0,000343	0,018528
72	0,005751	0,000333	0,018241
96	0,006192	0,000330	0,018154
144	0,006814	0,000351	0,018746
288	0,006148	0,000340	0,018439

Confirming the performing with LSTM Networks are better than the results with CNNs, the tests with dropout were performed either:

Table 5. Results - LSTM 1 INPUT - 0.1 Dropout

Neurons	Dropout	MAE	MSE	RMSE
36	0,1	0,005045	0,000371	0,019256
41	0,1	0,004898	0,000371	0,019250
48	0,1	0,005053	0,000353	0,018778
58	0,1	0,004998	0,000362	0,019034
72	0,1	0,005487	0,000363	0,019041
96	0,1	0,004905	0,000347	0,018624
144	0,1	0,005573	0,000367	0,019175
288	0,1	0,0144254	0,002333	0,048306

The prediction graph of the best architecture metric of 1 input, LSTM Network, with 48 neurons in the 1 hidden layer, without dropout, is represented as follows:

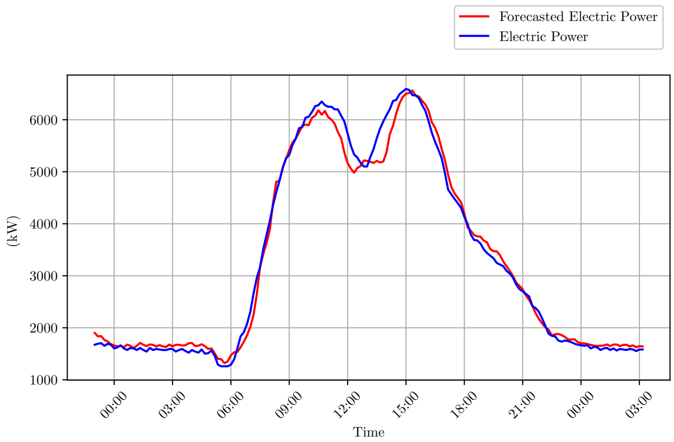

Figure 10. One day Ahead Prediction - 1 Input Feature.

Even with LSTMs the use of dropout worsens our model. Verifying the use of LSTM improve the behavior of the model, the new technique using periodicity of sines and cosines were implemented. New tests were performed using 3 inputs - active power, sine and cosine periodicity. The following table demonstrates the performance of the same architectures as used with 1 input.

Table 6. Results - LSTM 3 INPUTS

Neurons	MAE	MSE	RMSE
36	0,004642	0,000322	0,017932
41	0,004766	0,000327	0,018110
48	0,004662	0,000321	0,017912
58	0,004649	0,000309	0,017565
72	0,004446	0,000311	0,017627
96	0,004608	0,000317	0,017794
144	0,004580	0,000317	0,017813
288	0,0074892	0,000651	0,0255280

Analyzing the two tables is evident the best architecture was achieved without a dropout.The configuration was with 1 hidden layer with 58 neurons, obtaining the best score in the three metrics analyzed being 0,000309 in MSE metric, 0,017565 in RMSE and 0,004649 in MAE. By this, is evident the improvement using the new features, with the sine and cosine functions representing the seasonality and periodicity in the problem. The best result graph with 3 input features clearly demonstrates a better learning:

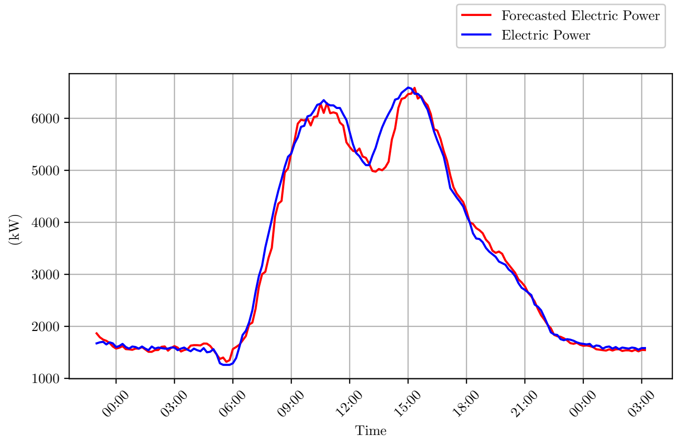

Figure 11. One day Ahead Prediction - 3 Input Features.

7. CONCLUSION

This paper proposes the use of artificial intelligence, more precisely, artificial neural networks of the LSTM type in time series to develop a very short-term forecasting tool.

Regarding the developed methodology, better results were achieved when all data processing and study techniques were used, that is, normalization, ideal definition of columns to be introduced in our neural network, ideal number of items in the test set and training. The best result of this configuration was with 3 inputs, and 3 hidden layers, with a score of 000309 in MSE metric, 0,017565 in RMSE and 0,004649 in MAE . Proving that the use of new features representing sine and cosine were effective and important in the prediction of this time series.

When comparing the use of a higher complexity, with more neurons and hidden layers, the results of the predictions were similar, however, what differentiates one from the other was mainly the simulation time and network training. Another topic is about the LSTMs architecture showing to have a higher performance in models compared to CNNs models. In addiction, an important point to report is the presence of Dropout or not, in this, it was possible to observe that as a percentage of Dropout was placed on the networks, in both cases, one input and three input features they worsened their performance. Finally, it is concluded that the new approach and algorithm is effective and can be used for time series prediction.

REFERENCES

[1] R. Palma-Behnke, C. Benavides, F. Lanas, B. Severino, L. Reyes, J. Llanos, and D. Saez, “A Microgrid Energy Management System Based on the Rolling Horizon Strategy,” IEEE Transactions on Smart Grid, vol. 4, no. 2, pp. 996–1006, jun 2013. [Online]. Available: http://ieeexplore.ieee.org/document/6417004/

[2] E. Goutard, “Expert Systems in Energy Management Systems,” Innovative Smart Grid Technologies Conference Europe (ISGT Europe), pp. 1–6, 2010.

[3]G. L. Tona, M. Luna, A. di Piazza and M. C. di Piazza, ”Development of a Forecasting Module based on Tensorflow for Use in Energy Management Systems,” IECON 2019 - 45th Annual Conference of the IEEE Industrial Electronics Society, 2019, pp. 3063-3068.

[4] D. Van Der Meer, G. R. C. Mouli, G. M. E. Mouli, L. R. Elizondo, and P. Bauer, “Energy Management System with PV Power Forecast to Optimally Charge EVs at the Workplace,” IEEE Transactions on Industrial Informatics, vol. 14, no. 1, pp. 311–320, 2018.

[5] Y. Yu, J. Cao, and J. Zhu, “An lstm short-term solar irradiance forecasting under complicated weather conditions,” IEEE Access, vol. 7, pp. 145 651–145 666, 2019.

[6] M. Rahmani and F. Faghihi, “A novel fuzzy cooperative frequency control method for islanding microgrids comprising renewable energy resources,” in 2019 5th Conference on Knowledge Based Engineering and Innovation (KBEI), Feb 2019, pp. 230–234.

[7] H. ElMoaqet, D. M. Tilbury, and S. K. Ramachandran, “Multi-step ahead predictions for critical levels in physiological time series,” IEEE Transactions on Cybernetics, vol. 46, no. 7, pp. 1704–1714, July 2016.

[8] G. L. Tona, M. Luna, A. di Piazza and M. C. di Piazza, ”Development of a Forecasting Module based on Tensorflow for Use in Energy Management Systems,” IECON 2019 - 45th Annual Conference of the IEEE Industrial Electronics Society, 2019, pp. 3063-3068.

[9] C. Liu, Z. Jin, J. Gu, and C. Qiu, “Short-term load forecasting using a long short-term memory network,” in 2017 IEEE PES Innovative Smart Grid Technologies Conference Europe (ISGT- Europe), 2017, pp. 1–6.

[10] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” Neural Comput., vol. 9, no. 8, pp. 1735–1780, Nov. 1997.

[11] F. A. Gers, J. Schmidhuber, and F. Cummins, “Learning to forget: Continual prediction with LSTM,” in Proc. 9th Int. Conf. Artif. Neural Networks: ICANN, vol. 2. Edinburgh, U.K., 1999, pp. 850–855.

[12] Aprillia, H.; Yang, H.-T.; Huang, C.-M. Short-Term Photovoltaic Power Forecasting Using a Convolutional Neural Network–Salp Swarm Algorithm. Energies 2020, 13, 1879. https://doi.org/10.3390/en13081879

[13] S. M. Elgarhy, M. M. Othman, A. Taha and H. M. Hasanien, ”Short term load forecasting using ANN technique,” 2017 Nineteenth International Middle East Power Systems Conference (MEPCON), 2017, pp. 1385-1394, doi: 10.1109/MEPCON.2017.8301364.

[14] S. Kumar, L. Hussain, S. Banarjee and M. Reza, ”Energy Load Forecasting using Deep Learning Approach-LSTM and GRU in Spark Cluster,” 2018 Fifth International Conference on Emerging Applications of Information Technology (EAIT), 2018, pp. 1-4, doi: 10.1109/EAIT.2018.8470406. *Nature*, 135:7–9, 1956.