IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Date of publication 2024 09, date of current version 2024 09.
Digital Object Identifier 10.1109/ACCESS.2017.DOI

Comparative review of SARIMA and LSTM models individually and in combination, for the forecasting of electrical consumption of substations in the Chile central zone

S. CHAVEZ[1], ()
1

Corresponding author: Chavez S. Author (e-mail: sebastian.chavez@gmail.com)
https://github.com/schavezh/UNAB_pronostico_consumo_electrico/tree/main/proyecto

ABSTRACT In this study, we perform a comparative analysis of SARIMA and LSTM models, individually and in combination, to forecast the electricity consumption of substations located in the central region of Chile. Using historical consumption data from January 2018 to December 2022, we evaluate the performance of these models to predict future demand, focusing on accuracy and computational efficiency. SARIMA models are known for their ability to handle seasonal patterns in time series, while LSTMs are powerful in capturing long-term dependencies and non-linear behaviors. By implementing both models, we aim to identify which approach yields the most accurate predictions for short and long-term electricity consumption, considering various metrics such as MSE, RMSE, MAE, and R^2. Additionally, we explore a hybrid SARIMA-LSTM model, leveraging the strengths of each to improve forecast accuracy. The results show that the hybrid model outperforms individual models, particularly in capturing complex consumption patterns, offering an effective solution for energy demand forecasting in substations with high seasonality.

INDEX TERMS Electricity consumption, forecasting, LSTM, recurrent neural networks (RNN), SARIMA

I. INTRODUCTION

PREDICTING electricity consumption is important to ensure a continuous balance between production and consumption, due to the high cost of storage, which is essential to maintain the stability of the electrical system. In addition, this forecast allows for the optimization of the use of HVAC systems, which depend on electricity for their operation, contributing to the reduction of operating costs and the maximum load of the system. Finally, forecasting electricity consumption helps to implement operational strategies, benefiting both public policies and companies by managing energy demand more effectively [1].

At the national level, there is the National Energy Commission, which, as established by Decree Law No. 2,244 of 1978, must technically analyze the structure and level of prices and tariffs for energy goods and services, as well as monitor and project the current and expected operation of the energy sector, for which it must generate a forecast of demand for electric energy, which will be used to determine the short-term node price [2].

In this work, a dataset with records of electrical consumption from electrical substations in the central area of Chile will be analyzed. The objective of this is to create a machine learning model to forecast future demand for electrical consumption, using the SARIMA and LSTM models and both in combination to evaluate their performance and analyze their results.

To do this, we will perform a descriptive analysis of the information provided and develop a strategy for the implementation of some machine learning models and neural networks, evaluating their performance and seeking the best performance to identify future demand.

II. RELATED WORK

In the discussion of the reviewed papers, various deep learning and statistical techniques and models used for predicting energy consumption in different environments are identified.

One of the key approaches addressed in these works is the use of hybrid models that combine convolutional neural networks (CNN) with recurrent neural networks (RNN) and autoencoder (AE) models. This type of architecture has been used, for example, to process time series of unequal lengths and predict energy consumption in different locations. These models stand out for their ability to handle large volumes of data with irregularities in the length of the time series and non-linear consumption patterns, as shown in experiments performed with data from an energy distributor in Brazil, where their performance was compared with other statistical methods such as SARIMAX and Prophet, obtaining lower error measures in several cases [3].

Another prominent model is the Long- and Short-Term Time-series Network (LSTNet), which combines CNN and RNN to extract local dependencies and long-term patterns in multivariate time series. This approach is particularly relevant for energy data due to its ability to capture recurring daily and weekly patterns, as well as long-term trends affecting energy consumption, such as seasonal changes. Through this combination, LSTNet addresses the limitations of traditional autoregressive models (such as ARIMA), which often fail to capture complex non-linear dynamics and long-term patterns, significantly improving performance on real data [4].

Furthermore, an alternative approach, the use of SARIMA models and artificial neural networks (ANN), has been evaluated for the prediction of microgrid energy consumption in an educational setting. These models face common problems such as the lack of complete data and the need for intensive data processing to impute missing values or inconsistencies. In this case, neural networks demonstrated higher accuracy compared to SARIMA models, highlighting their flexibility to capture complex and non-linear patterns in the data [5].

In summary, deep neural network-based models, such as CNN, LSTM, and hybrid approaches, have proven to be powerful tools for energy consumption prediction compared to traditional statistical methods. However, challenges such as computational complexity, the need for hyperparameter optimization, and the correct handling of incomplete or unequal-length data remain limiting factors in their practical implementation.

III. STUDY OBJETIVE

The objective of this study is to develop an optimal model for predicting electrical energy consumption in substations in the central area of Chile, using historical data between January 2018 and December 2022. To do so, different machine learning and statistical approaches will be evaluated, such as the SARIMA model and recurrent neural networks (RNN), comparing their performance based on metrics appropriate to the data set analyzed, in order to select the most accurate and efficient model.

IV. METHODOLOGY

A. DATASET DESCRIPTION

The dataset corresponds to two files with CSV extension, which contain training and test data separately.

The training file has 244,391 observations and the test file has 61,313 observations, both with the electrical consumption of the Ajahuel, Buin, Chena, Cnavia, Elsalto, Florida and Losalme substations. In addition, each record contains the variables substation, date (date of the record) and consumption.

B. METHODOLOGY FOR DEVELOPMENT

In this work, the CRISP-DM methodology will be taken as a baseline, considering that we have a clear objective and a defined data set and we will carry out the following stages: Understanding the business, Understanding the data, Data preparation, Develop model, Evaluate models and Deployment [7].

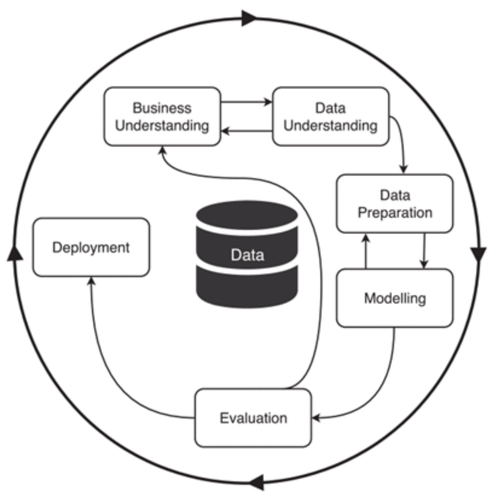

FIGURE 1. CRISP-DM data mining process model [7].

1) Business understanding

Electricity consumption forecasting is inherently a time series problem, as energy demand fluctuates over time and is influenced by various temporal patterns. The nature of electricity consumption is often characterized by recurring seasonal, daily, monthly, and even quarterly patterns, which are closely tied to factors such as temperature variations, daylight hours, and socio-economic activities. For instance, energy usage typically peaks during summer and winter months due to the increased need for heating and cooling systems, respectively. Similarly, daily consumption follows a cyclical pattern, with higher usage during working hours and lower demand at night [3]. These fluctuations make electricity consumption highly dependent on both short-term and long-term trends, creating a complex forecasting challenge [4].. Accurately capturing these patterns is crucial for energy providers to ensure that supply meets demand, optimize resource allocation, and improve grid stability [5]. Moreover, the correlation between energy consumption and external variables, such as weather conditions and holidays, introduces additional layers of complexity. Thus, effective forecasting models must not only account for these cyclical behaviors but also be robust enough to handle irregularities and sudden changes in consumption, such as those caused by extreme weather events or unexpected socio-economic shifts [6].

2) Data Preparation

Separation by substations: The dataset is divided by the different energy substations to focus the analysis and better identify patterns present in the data. With this we can see that there are different patterns for each substation, which will help the models to better identify the patterns. Also, for each substation, the missing values are interpolated, which is essential in energy prediction models, where the data is often incomplete or has "holes" due to measurement errors [3]. [5].. Resampling by hour is also justified by the need to make accurate short-term predictions (hours), as explained in the energy prediction literature [4].

Outlier detection: An outlier detection and handling process based on the calculation of the interquartile range (IQR) is included. This process is crucial, as outliers can distort the predictions of time series models, especially in scenarios such as energy consumption, where extreme variations in the data can lead to significant errors in forecasts.

RNNs, and in particular their variants such as LSTM (Long Short-Term Memory), are designed to capture temporal dependencies and sequential patterns in the data. If the data contains outliers, these extreme values can bias the activations of neurons and cause the model to focus excessively on those outlier points, distorting the representation of the underlying true patterns. This results in ineffective training and erratic predictions [3].

Another problem is that if outliers are introduced into the sequence, they can generate errors in the initial stages that propagate over time, affecting future predictions. This is especially problematic in long sequences, as errors caused by outliers can be amplified, significantly affecting the long-term performance of the model [4].

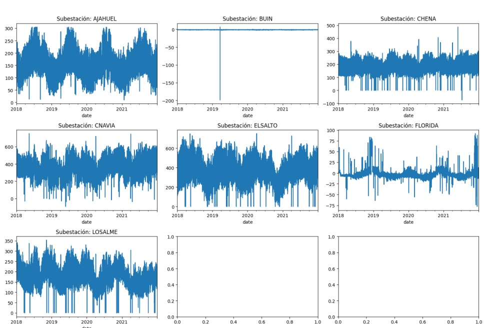

FIGURE 2. Training data on electrical consumption separated by substations

Extracting information from temporal cycles: Time series often have repetitive or seasonal patterns, such as energy consumption that varies over the year, week, or even day. To capture these temporal patterns, time-based features such as day of the year, day of the week, quarter, or time of day are extracted. These new features help the model identify patterns of periodicity that would not be apparent from the original values alone [3].

Applying sine and cosine metrics: Once these temporal features (day of the year, time of the day, etc.) are extracted, they are transformed into cyclic variables using trigonometric functions (sine and cosine). This is essential because features such as "time of the day" or "month of the year" have a cyclic nature; for example, the 24 hours of a day form a cycle where hour 0 and hour 23 are close to each other. To do this, the sine-cosine formula is applied to convert temporal features into coordinates on a circle. This creates two new features for each: one with the sine function and one with the cosine function. This allows the model to understand that time 23 and time 0 are very close together, which would not be obvious if the "time of day" value were simply used [3].

$$\text{Seno} = \sin\left(\frac{2\pi \cdot \text{hora}}{24}\right), \quad \text{Coseno} = \cos\left(\frac{2\pi \cdot \text{hora}}{24}\right)$$

FIGURE 3. Example of sine and cosine function for feature extraction for hour of day

Autocorrelation (ACF): This is a key tool to measure the correlation between observations at different lags of a time series. In the context of SARIMA models, the ACF allows us to identify how many past values have a significant relationship with the current value, which helps to determine the "q" parameter of the moving average (MA) part of the model[3]. The autocorrelation (ACF) plot shows a clear pattern of seasonality, with regular spikes indicating significant correlation between observations separated by specific time intervals. In the context of a SARIMA model, this information suggests that the data has obvious seasonal components, which is essential to properly tune the "q" parameter of the moving average (MA) part. In this case, spikes at lags 24, 48, and 72 indicate the presence of regular 24-hour cycles, which coincides with daily patterns of energy consumption. This is critical to capture the cyclical structure in the data, allowing the SARIMA model to properly account for repetitive and seasonal patterns in consumption.

Partial autocorrelation (PACF): This is used to measure the direct correlation between an observation and its lags, removing intermediate effects. In SARIMA models, this is essential to determine the parameter "p" of the autoregressive (AR) part of the model [4]. In the partial autocorrelation (PACF) plot, a sharp drop is observed in the first lags, indicating that the direct dependence between observations decreases rapidly. This is crucial to determine the parameter "p" of the autoregressive (AR) component of the SARIMA model. The first lags show significant partial correlation, suggesting that an AR model with few terms may be sufficient to capture the autoregressive relationships in the time series. In this case, it seems that a low-order autoregressive term, such as p=1 or p=2, would be suitable for the model, allowing for an efficient representation of the data.

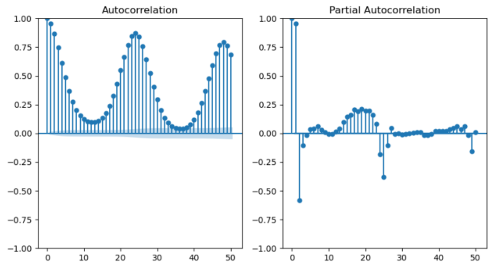

FIGURE 4. Autocorrelation and partial autocorrelation graph for the consumption data of the AJHAHUEL substation

SARIMAX Results

Dep. Variable:	consumption	No. Observations:	35064
Model:	SARIMAX(1, 0, 1)x(1, 1, 1, 48)	Log Likelihood	-127537.597
Date:	Thu, 26 Sep 2024	AIC	255085.195
Time:	12:38:11	BIC	255127.512
Sample:	01-01-2018	HQIC	255098.674
	- 12-31-2021		
Covariance Type:	opg		

	coef	std err	z	P>|z|	[0.025	0.975]
ar.L1	0.9245	0.001	644.723	0.000	0.922	0.927
ma.L1	0.1184	0.001	82.964	0.000	0.116	0.121
ar.S.L48	-0.0315	0.003	-11.985	0.000	-0.037	-0.026
ma.S.L48	-0.8454	0.002	-510.589	0.000	-0.849	-0.842
sigma2	85.1854	0.136	627.616	0.000	84.919	85.451

Ljung-Box (L1) (Q):	0.68	Jarque-Bera (JB):	14207370.45
Prob(Q):	0.41	Prob(JB):	0.00
Heteroskedasticity (H):	0.48	Skew:	-0.42
Prob(H) (two-sided):	0.00	Kurtosis:	101.68

FIGURE 5. SARIMA model results and configuration with a 48 hours window

3) Modeling

SARIMA (Seasonal AutoRegressive Integrated Moving Average) model: We use SARIMA instead of an ARIMA (AutoRegressive Integrated Moving Average) model because the data, as we have seen, show patterns that repeat at regular time intervals, such as hours of the day, days of the week, months of the year, or even year. The key difference between SARIMA and ARIMA is that SARIMA includes additional terms to model these seasonal components, which allows us to better capture cyclical fluctuations in the data. Regarding the seasonality of the data, as we could see, they show seasonality as we saw in the autocorrelation graph (ACF), since there are peaks at regular intervals (such as lags of 24, 48, 72), which indicate the presence of a daily seasonal pattern. An ARIMA model does not have the capacity to capture this seasonality, since it only takes into account linear dependencies between adjacent observations and not those that are repeated at longer intervals. SARIMA, on the other hand, includes seasonal terms that allow modeling these types of patterns [1].

In summary, the use of SARIMA is preferable to ARIMA when the data presents seasonality, as is the case with energy consumption, since it can model both short-term dependencies and seasonal cyclical patterns that are essential for obtaining accurate and robust predictions.

However, SARIMA is a model that is tuned through an iterative process, where multiple parameters are optimized, including non-seasonal and seasonal ones. As the volume of data and window length increases, the amount of calculations required to fit the model and estimate its parameters increases exponentially. This can make model training extremely slow, especially in time series with millions of data points or when analyzing long seasonal patterns (such as annual cycles) along with shorter cycles (daily or monthly). Parameter tuning can be computationally expensive in terms of time and processing, as SARIMA needs to evaluate different combinations of parameters (p, d, q) and (P, D, Q) to find the best configuration.

For this reason and based on our ACF and PACF graphs we selected the 24 hours and 48 hours windows to run our models, and since our data is seasonal we set the parameter d to 0.

LSTM (Long Short-Term Memory): For electricity consumption forecasting, they are particularly suited for time series with long-term dependencies, such as energy consumption, where consumption patterns depend not only on the last few hours, but also on behaviors that can occur over days, weeks, or even seasons. LSTMs can retain and process relevant information over long periods of time thanks to their memory architecture. This allows them to capture both short-term patterns, such as hourly fluctuations, as well as more complex long-term trends [3], something that other models, such as ARIMA or SARIMA, have difficulty handling when dependencies are non-linear or do not fit well to predetermined seasonal terms [4].

Furthermore, since the model only receives hourly electricity consumption data, the LSTM has the ability to identify implicit recurring and seasonal patterns (such as daily or weekly cycles) without the need for complex feature engineering [3].

Internally, an LSTM network processes data through memory units controlled by "gates" that regulate what information is retained or forgotten at each step. Specifically, these networks use three types of gates: input gate, forget gate, and output gate. The input gate decides what new information to add to memory, the forget gate regulates what information should be discarded from memory (e.g., less relevant consumption patterns), and the output gate determines what parts of memory are used to generate the prediction. This flexible control mechanism allows the LSTM model to better handle complex and heterogeneous temporal patterns of electrical consumption, especially when only hourly consumption data is available with no other external variables that can help model the time series [4].

For our LSTM model we have used the random search methodology from the Keras Tuner package to find the best hyperparameters. The random search selects combinations of hyperparameters at random within the minimum and maximum values specified for each hyperparameter, such as the number of LSTM layers, the number of LSTM units, the learning rate and the dropout.

Tuned hyperparameters: Number of LSTM layers: The model tunes between 1 and 3 LSTM layers, which can help find the best architecture to capture temporal dependencies. LSTM units: It varies between 32 and 128 units, tuning to find the optimal network size for predicting electricity consumption. Dropout rate: It varies between 0.1 and 0.5, testing different levels of regularization to avoid overfitting. Learning$_{r}ate : Itisoptimizedbetween1 \times 10^{-4}$ y 1×10^{-2}, which allows finding the best convergence rate.

Random Search explores only a random subset, making it more efficient when working with a large number of hyperparameters or very wide ranges. This reduces the computational cost and allows finding configurations close to the optimal ones with less training time. For the Control of the number of iterations we have specified that the search process runs for a maximum of 10 different configurations, which helps limit the execution time and resource usage, tuning each configuration twice for greater robustness.

In summary, using Random Search with Keras Tuner is an efficient and flexible methodology for tuning hyperparameters, especially in the context of models such as LSTM, where the number of hyperparameters can be high and the computational cost of testing all combinations would be very high.

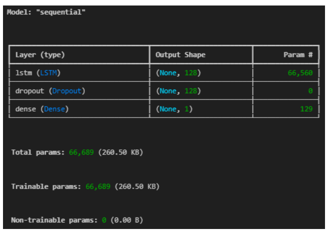

FIGURE 6. Best parameters found to LSTM model with random search of Keras Tuner

Hybrid SARIMA + LSTM model: Next we will perform a hybrid model that involves training the data with our SARIMA model with exogenous data and on the other hand calculating the differences of the SARIMA prediction and making a prediction of it with an LSTM model. The hybrid model strategy in its first phase, we will use SARIMA to model the long-term trends and seasonality of the electricity consumption data. These predictions will serve as a solid basis for the general trend. Then, in the second phase we train an LSTM model on the SARIMA residuals, which corresponds to the differences between the realvalues and the SARIMA predictions. The LSTM will then try to capture short-term patterns that SARIMA could not model well and the final predictions of the hybrid model will be the sum of the SARIMA predictions plus the predictions of the differences made with LSTM.

4) Evaluation

RMSE (Root Mean Squared Error): This metric measures the square root of the average of the squared errors between the actual and predicted values. It penalizes large errors more, so it is sensitive to large discrepancies. It is expressed in the same units as the data, so it is easy to interpret.

MAE (Mean Absolute Error): Description: It is the average of the absolute errors between the actual and predicted values, but it does not penalize large errors as much. It measures the average magnitude of the errors in the same units as the data.

R^2 (Coefficient of Determination): This measures how well the model captures the variability of the actual data. A value close to 1 indicates that the model explains the variation well, while a value close to 0 indicates the opposite. It helps to evaluate the model's ability to explain the variability of the data.

Finally, for our case we will stick with RMSE since we are looking for better accuracy over a long time horizon, and given our 1-year forecast target it is the best option, since it penalizes large errors more, which is essential in long-term predictions where small errors can accumulate.

V. EXPERIMENTS (RESULTS AND THEIR DISCUSSION)

A. SARIMA 24 HOURS AJAHUEL SUBSTATION DATA

In this first experiment, the SARIMA model is executed with a 24-hour window and with parameter q=0 as analyzed in the correlation graphs for the data from the AJHAUEL substation. Here we can identify that in the forecast, the model fails to identify the seasonal pattern of the months and year, since the window used for training was only 24 hours.

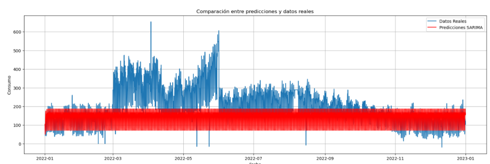

FIGURE 7. SARIMA 24-hour model forecast result graph versus real test data for AJHAUEL substation

As we can see from the results of the metrics of our first model, it confirms that it is not capturing the variability of the data well, as evidenced by the high MSE values that penalize the large differences more and the negative R2 that indicates that the model has a worse performance than a base model that simply predicted the average of the values of the series. This last value is a clear sign that the model is not capturing the variability of the data and is not suitable for this particular data set.

Métrica	Valor
Mean Squared Error (MSE)	11043.55
Root Mean Squared Error (RMSE)	105.09
Mean Absolute Error (MAE)	77.14
R-squared (R2)	-0.25

TABLE 1. Error metrics results of the SARIMA 24-hour model substation AJAHUEL

B. SARIMA 48 HOURS AJAHUEL SUBSTATION DATA

When we change the time window parameter to 48 hours, there is not much change in our forecast graph, where it is evident that the predominant series corresponds to a short time window without managing to capture the monthly or annual variability.

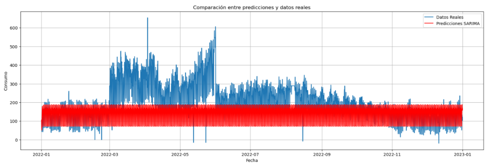

FIGURE 8. SARIMA 48-hour model forecast result graph versus real test data for AJHAUEL substation

The metrics in this case do not vary much from the previous experiment, so the same comments are maintained.

Métrica	Valor
Mean Squared Error (MSE)	10999.96
Root Mean Squared Error (RMSE)	104.88
Mean Absolute Error (MAE)	77.01
R-squared (R2)	-0.25

TABLE 2. Error metrics results of the SARIMA 48-hour model substation AJAHUEL

C. SARIMA WITH EXOGEN SUBSTATION AJAHUEL DATA

Now with our SARIMA model, including exogenous data extracted from the training dataset dates, we see that our model is able to better capture the variability and seasonality of the data over time, moving in a more similar way to the test data that is being predicted. However, due to having an apparently atypical period, our model is unable to predict accurately and deviates from its trend.

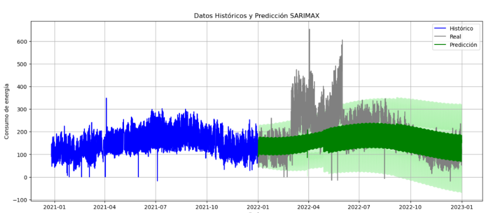

FIGURE 9. SARIMA 24-hour model + exogen data forecast result graph versus real test data for AJHAUEL substation

The above can be confirmed with our metrics, which improve compared to models that do not contain exogenous data; however, our prediction still has a considerable range of error since according to the RSME the data deviate by 91.42 points according to its scale.

Métrica	Valor
Mean Squared Error (MSE)	8358.14
Root Mean Squared Error (RMSE)	91.42
Mean Absolute Error (MAE)	62.69
R-squared (R2)	0.05

TABLE 3. Error metrics results of the SARIMA 48-hour + datos exogenos model substation AJAHUEL

D. LSTM TO SUBSTATION AJAHUEL

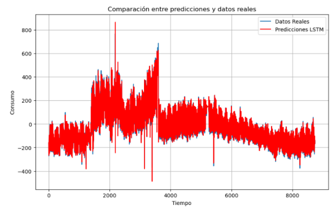

FIGURE 10. LSTM forecast result graph versus real test data for AJHAUEL substation

E. LSTM + SARIMA WITH EXOGEN SUBSTATION AJAHUEL DATA

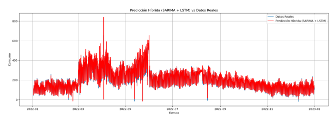

FIGURE 11. LSTM + SARIMA forecast result graph versus real test data for AJHAUEL substation

Finally, we see that our hybrid model gives us the best forecast since we use the best of the two models explored, separating the tasks of predicting short- and long-term electricity consumption data and adding them together at the end.

Métrica	SARIMA_24	SARIMA_48	SARIMA_exogen	LSTM	SARIMA_EX_LSTM
MSE	11043.55	10999.96	8358.14	935.36	405.01
RMSE	105.09	104.88	91.42	30.58	20.12
MAE	77.14	77.01	62.69	14.94	9.88
R2	-0.25	-0.25	0.05	0.96	0.95

TABLE 4. Comparación de métricas entre modelos SARIMA y LSTM

VI. CONCLUSION

Finally, we can see that there are a large number of methods and techniques to achieve a forecast in the case of electrical consumption, however, we must consider the use of computational resources and evaluate the best cost versus benefit alternative, which can give us an approximation as close to reality as possible but with coherent resources. As we saw in the experiments, data analysis and the application of statistical techniques are the basis for obtaining efficient models since just using the data we have available does not allow us to achieve good performance in the models.

REFERENCES

[1] Khan, A., Chiroma, H., Imran, M., Bangash, J. I., Asim, M., Hamza, M. F., & Aljuaid, H. (2020). Forecasting electricity consumption based on machine learning to improve performance: A case study for the organization of petroleum exporting countries (OPEC). *Computers & Electrical Engineering*, *86*, 106737.
[2] https://www.cne.cl/tarificacion/electrica/prevision-de-demanda-electrica
[3] Rick, R., & Berton, L. (2022). Energy forecasting model based on CNN-LSTM-AE for many time series with unequal lengths. Engineering Applications of Artificial Intelligence, 113, 104998.
[4] Lai, G., Chang, W.-C., Yang, Y., & Liu, H. (2018). Modeling Long- and Short-Term Temporal Patterns with Deep Neural Networks. In Proceedings of ACM Conference (SIGIR'18), 11 pages.
[5] Wasesa, M., Tiara, A. R., Afrianto, M. A., Ramadhan, F. I., Haq, I. N., & Pradipta, J. (2020, December). SARIMA and artificial neural network models for forecasting electricity consumption of a microgrid based educational building. In 2020 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM) (pp. 210-214). IEEE.
[6] Bedi, J., & Toshniwal, D. (2019). Deep learning framework to forecast electricity demand. *Applied energy*, *238*, 1312-1326.
[7] Martínez-Plumed, F., Contreras-Ochando, L., Ferri, C., Hernández-Orallo, J., Kull, M., Lachiche, N., ... & Flach, P. (2019). CRISP-DM twenty years later: From data mining processes to data science trajectories. *IEEE transactions on knowledge and data engineering*, *33*(8), 3048-3061.