Predicting Energy Consumption Through Deep Learning-Powered Time Series Analysis

Zineb Touati-Hamad[1] , Mohamed Ridda Laouar[2]

Laboratory of Mathematics, Informatics and Systems
Echahid Cheikh Larbi Tebessi University
Tebessa, 12002, Algeria

[1]zineb.touatihamad@univ-tebessa.dz
[2]ridda.laouar@univ-tebessa.dz

Abstract. This paper focuses on precise energy consumption prediction through AI-driven models using time series data. Addressing data quality and model selection challenges, we conduct thorough data analysis, rectifying missing values, and normalizing inputs. Leveraging LSTM due to its efficacy in capturing temporal dependencies, we surpass existing limitations. Experimental outcomes validate our approach, accentuating the significance of data analysis, highlighting LSTM's relevance for precise consumption forecasting. These findings inform effective energy management strategies across domains.

Keywords: Energy Consumption, Prediction, Time Series, Data Analysis, LSTM

1 Introduction

Energy consumption is a pressing concern with far-reaching implications for the economy, environment, and society at large. Meeting the escalating energy demand necessitates smarter, more sustainable management techniques. Artificial Intelligence (AI) and Machine Learning (ML) models offer promising avenues for tackling this challenge, ushering in innovative solutions for predicting and analyzing energy usage [1].

Recent times have witnessed a surge in enthusiasm among researchers and practitioners for integrating AI and ML models into the realm of energy consumption. These models find application across diverse energy-centric domains, spanning energy demand projection, consumption optimization, and efficiency enhancement. Their deployment has markedly enhanced the accuracy of energy consumption forecasts, empowering energy enterprises to optimize resource utilization and curtail wastage.

The potential of ML models in foreseeing energy consumption patterns is notably exciting [2]. They hold the potential to reshape energy administration fundamentally, proffering precise, streamlined resolutions. As technology advances, these models have matured, capably managing intricate energy usage trends and foreseeing future patterns. This positions them as invaluable assets for prudent energy management, judicious decision-making, waste mitigation, and bolstering energy efficiency.

This paper delves into the challenges tied to data quality and model selection within energy consumption prognosis, employing an AI model tailored for time series data. The exploration centers on how this model can enrich energy management strategies and contribute to a sustainable energy future. The objective is to furnish a comprehensive comprehension of Machine Learning model deployment, pinpointing prospects and obstacles for effective integration.

2 State of the Art

Several studies have delved into forecasting electricity consumption using various techniques and approaches. Hadjout et al [3] introduced an ensemble-based model leveraging deep learning techniques to predict monthly electricity consumption for Algeria's economic sector. Their model integrated LSTM, GRU, and CNN, and was assessed using a chronological dataset from 2000 clients spanning 14 years of monthly electricity consumption in Bejaia, Algeria. The findings revealed the proposed model achieved a MAPE of 3.04% and an RMSE of 60.66.

Shakouri and Sahed [4] proposed an electricity consumption forecasting model based on artificial neural networks (ANN) for Algeria's annual consumption prediction. Their model employed a feedforward neural network architecture with backpropagation algorithm. Evaluating the model on a multivariate dataset from Algeria's electrical system, the results indicated a MAPE of 0.033.

Bezzar et al [5] introduced a data analysis-driven time series forecasting approach for managing domestic electricity consumption. They presented an XGBoost model addressing challenges like inappropriate model selection and unanalyzed time series datasets. Experimental results on the Individual Household Electricity Power Consumption (IHEPC) dataset showcased the superiority of their proposed model over various ST and/or AI-based models in the literature, achieving RMSE and MAPE of 0.229 and 0.026, respectively.

Kim and Cho [6] proposed a CNN-LSTM neural network model to predict residential energy consumption. Trained and tested on the publicly available "individual household electricity power consumption" (IHEPC) dataset, the model achieved a MAPE of 32.83, MSE of 0.3549, RMSE of 0.5957, and MAE of 0.3317.

Han et al [7] introduced an effective deep learning framework for intelligent energy management in IoT networks. The proposed framework combined LSTM and CNN models for energy consumption forecasting, evaluated using the PJM dataset.

Results demonstrated the framework achieved MSE and RMSE of 0.15 and 3.77, respectively.

Khan et al [8] proposed an efficient short-term electrical load forecasting model for effective energy management. Their hybrid deep learning approach was evaluated on the publicly available PJM dataset, achieving a good performance with an RMSE of 3.4.

Mujeeb and Javaid [9] introduced two novel approaches, ESAENARX and DE-RELM, for predictive analysis of large volumes of load and electricity price data. These approaches employed data analysis techniques like regression, clustering, and classification to forecast energy consumption and prices. Evaluated on the publicly available PJM dataset, these approaches outperformed reference models in terms of MAPE and RMSE, with values ranging from 1.08 to 5.24.

Gao et al [10] proposed a multi-block-based prediction engine for price and load forecasting. This engine used a hybrid approach combining different neural network combinations, tested under similar prediction conditions to showcase their capabilities. Evaluated on the publicly available PJM dataset, the engine demonstrated high performance with an RMSE of 1.14 and a MAPE of 0.49.

3 Contribution

The goal of our work is to design an intelligent system capable of predicting hourly electrical energy consumption based on the hourly energy consumption database from the PJM website, which is known in related research works. The structure of the proposed model is depicted in Figure 1.

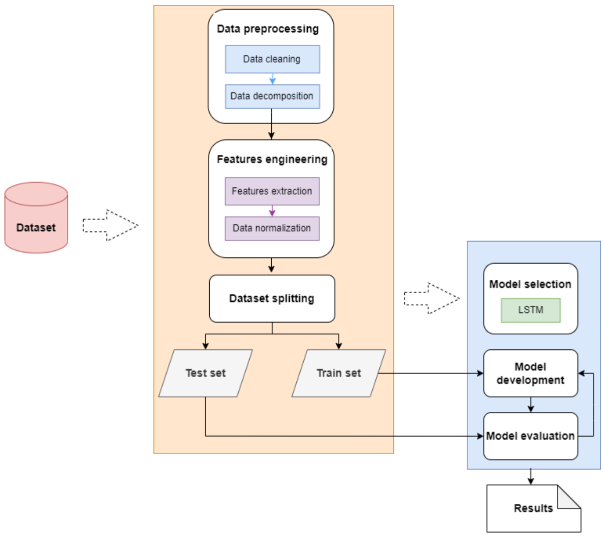

Figure 1. Proposed methodology

3.1 Dataset

The dataset utilized in this study consists of hourly energy consumption data spanning over 15 years, sourced from "PJM Interconnection LLC (PJM)," a US-based organization managing residential electrical transmission. The data is publicly accessible on the PJM website under the "CC0: Public Domain" license. It represents a significant resource of hourly energy consumption in megawatts (MW), available on Kaggle [11] [12] . This dataset is particularly valuable for our research due to its relevance, substantial size, and widespread availability. It has been extensively used in academia, making it a reliable and benchmark-worthy source for data preprocessing, analysis, and forecasting techniques.

3.2 Data preprocessing

1) Data Cleaning: As the initial phase of our work, we scrutinize the dataset to identify any occurrences of missing values, zeros, duplicates, or outliers. Our dataset does not encompass any nulls or duplicates requiring treatment. By pinpointing and addressing any potential issues, we can enhance the accuracy and reliability of our analysis.

2) Data Decomposition : Time series decomposition is a method used to break down a time series into its components, such as trend, seasonality, and residuals, while also assessing the stationarity of the time series. Analyzing the time series data is crucial to estimate and distinguish all present components, emphasizing their impact on the overall behavior of the time series.

The trend graph of our dataset is presented in Figure 2 (a). The trend component represents the long-term movement of the data. The graph provides insights into energy consumption, showing an increase during hot and cold periods and a slight decrease during other periods. Also in Figure 2 (a), the consumption pattern differs based on the considered duration. Over a longer period (such as several years), the consumption trend isn't a straight line but a curve, indicating a nonlinear relationship between time and consumption. This means that the rate of consumption change varies over time and isn't constant. However, over a shorter period, the consumption trend appears as a straight line, indicating a linear relationship between time and consumption. This suggests a constant and predictable rate of consumption change during this shorter period.

Figure 2 (b) illustrates the seasonality component. The seasonality component represents the regular pattern of data that repeats over a fixed period, such as daily, weekly, monthly, or yearly. Seasonality can be positive or negative, indicating whether data tends to be higher or lower during specific periods. This information can be useful for understanding and predicting future electricity consumption values.

Figure 2 (c) displays the residuals. Residuals represent the unexplained variation in the data that cannot be attributed to the trend or seasonality components. They are the random and unpredictable fluctuations in the data.

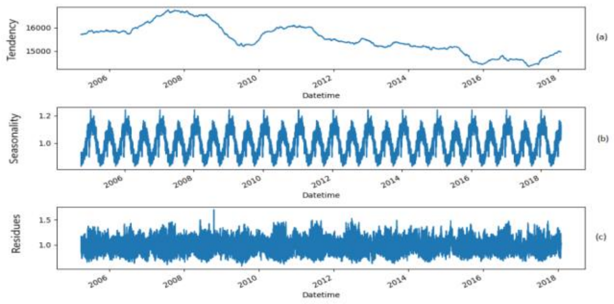

Figure 2. Decomposition of the time series (Trendency, seasonality and residues)

3.3 Feature Engineering

1) Feature Extraction : Feature extraction from time series involves extracting relevant information or features from time series data. The goal is to identify and capture patterns, trends, or other characteristics within the time series that can be used in prediction tasks. Effective feature extraction techniques can enhance the accuracy and efficiency of these models.

This process holds significant importance in any data science project involving machine learning. The chosen features serve as input to the learning model utilized.

We extracted the following features from our data:

• Hour: values of the day's hour (00:00=0, 23:00=24) for each data point.

The time of day can be useful for identifying daily trends in the data, such as activity spikes at certain times of the day.

• Dayofweek: week: values of the day of the week (Monday=0, Sunday=6) for each data point. The day of the week can help capture weekly trends in the data, such as consumption changes during weekdays versus weekends.

• Quarter: values of the year's quarter (1-4) for each data point. The quarter of the year can be useful for identifying seasonal trends in the data, such as increased consumption at different times of the year.

• Month: values of the year's month (1-12) for each data point. The month of the year can also be useful for identifying seasonal trends in the data.

• Year: values of the year (2004-2018) for each data point. The year can be useful for identifying longer-term trends in the data, such as gradual changes in consumer behavior over time.

• Dayofyear: values of the day of the year (1-365 or 1-366) for each data point. The day of the year can be useful for identifying cyclic trends in the data, such as energy consumption changes related to shifting seasons.

	EnergyConsumption_MW	hour	dayofweek	quarter	month	year	dayofyear
Datetime							
2004-10-01 01:00:00	12379.0	1	4	4	10	2004	275
2004-10-01 02:00:00	11935.0	2	4	4	10	2004	275
2004-10-01 03:00:00	11692.0	3	4	4	10	2004	275
2004-10-01 04:00:00	11597.0	4	4	4	10	2004	275
2004-10-01 05:00:00	11681.0	5	4	4	10	2004	275
...	...	...	...	...	...	...	...
2018-08-02 20:00:00	17673.0	20	3	3	8	2018	214
2018-08-02 21:00:00	17303.0	21	3	3	8	2018	214
2018-08-02 22:00:00	17001.0	22	3	3	8	2018	214
2018-08-02 23:00:00	15964.0	23	3	3	8	2018	214
2018-08-03 00:00:00	14809.0	0	4	3	8	2018	215

Figure 3. Features extraction

2) Data Normalization : data normalization is crucial to ensure a consistent scale across different variables within a time series and mitigate the impact of outliers. The Min-Max scaling is a common method, mapping features to a fixed range [0, 1]. Min-Max normalization is performed using the following formula:

$$\hat{x} = (x - xmin) / (xmax - xmin)$$

Here, x represents the value of the feature we are normalizing, while $xmin$ and $xmax$ respectively denote the smallest and largest observed values for that feature [5].

3.4 Dataset Splitting

Data was split into training and testing sets during the study. Figure 4 depicts the division of data into training and testing sets, based on the time series of 01-01-2015. As a result, the training and testing set split ratio was approximately 70-30%.

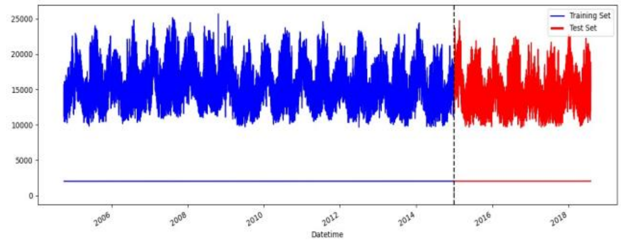

Figure 4. Dataset splitting

3.5 Model Selection

Our LSTM model consists of a total of four layers: The first layer is an LSTM layer with 40 memory cells, serving as the initial input layer, expecting sequential data with a sequence length and seven features. This layer processes the input sequence, captures temporal dependencies, and generates output sequences. Following the first LSTM layer, a dropout layer is added to mitigate overfitting, randomly setting a fraction (10%) of input units to zero during training. This regularization technique enhances the model's ability to generalize to unseen data.

The second layer is another LSTM layer with 40 memory cells, mirroring the architecture of the first LSTM layer. Its purpose is to further process sequential information and generate output sequences. It is succeeded by another dropout layer to alleviate overfitting.

The third layer is the final LSTM layer of the model, also comprising 40 memory cells. However, unlike the previous two LSTM layers, this layer does not return sequences. Instead, it summarizes the information learned from the sequence and produces a single output. Once again, a dropout layer is included after this layer for regularization.

The last layer is a dense layer with a single unit, performing a linear transformation on the input and producing a single output value. This layer acts as the model's output layer and is responsible for generating predictions based on the patterns learned from sequential data.

3.6 Model Evaluation

In the context of energy consumption forecasting, having an accurate predictive model is crucial. Therefore, assessing model performance using measures that effectively capture forecast accuracy and precision is essential. Through a combination of RMSE, MAE, MAPE, and R^2, we can comprehensively evaluate model performance and pinpoint areas where it might underperform. RMSE is valuable when larger errors impact overall performance more than smaller errors. MAE is useful when all errors, whether big or small, equally influence overall performance. MAPE is beneficial when evaluating model performance in terms of percentage errors. R^2 indicates how well the LSTM model fits the prediction problem. Employing multiple evaluation metrics provides a holistic view of the model's strengths and limitations, allowing us to leverage this information for performance enhancement.

These outcomes highlight the robustness and reliability of our LSTM model in electricity consumption prediction. The low RMSE value signifies predictions closely align with actual values, while high MAPE and MAE indicate precision with minimal errors. Moreover, the elevated coefficient of determination (R^2) at 0.96 indicates our model's capacity to explain 96% of the observed variance in the data. These encouraging observations instill strong confidence in employing our

LSTM model for accurate electricity consumption forecasts, holding significant potential for energy resource planning and optimization.

3.7 Results & Discussion

Upon analyzing Figures 5, it is evident that our LSTM model exhibits remarkable capability in making accurate predictions, closely tracking real data points.

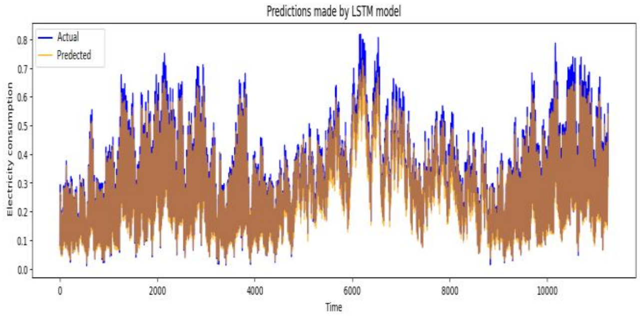

Figure 5. Prediction results with LSTM on the test set

Model performance on the testing set can be evaluated using various metrics: an RMSE of 0.03, MAPE of 0.08, MAE of 0.02, and a coefficient of determination (R^2) of 0.96, as shown in Figure 6.

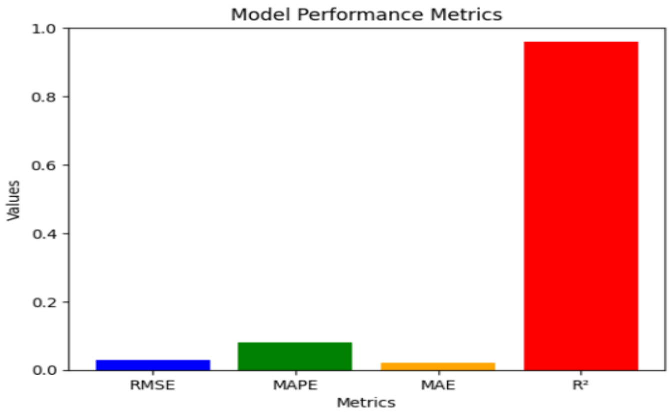

Figure 6. Performance of our LSTM model

To ensure a focused and concise comparison, we will exclusively evaluate the performance of our LSTM model in relation to studies that have used the same dataset as ours.

The table below illustrates the performance comparison of our model with previous works, using performance metrics such as RMSE, MAPE, MAE, and R^2.

Analyzing the results, it's evident that the work [3] demonstrates a high RMSE of 60.66, indicating considerable prediction error compared to other studies. Conversely, our model achieves an impressively low RMSE of 0.03, suggesting high accuracy in energy consumption prediction.

In terms of MAPE, our model achieves a result of 0.08, significantly lower than most previous works. The work [9] presents the highest MAPE of 1.09, indicating greater deviation from other models. Concerning MAE, our model achieves a very low result of 0.02, showcasing high precision in energy consumption predictions. In contrast, work [6] has the highest MAE of 0.3317, signifying a larger absolute error compared to other studies.

Lastly, the coefficient of determination (R^2) assesses the proportion of variance in the data explained by the prediction model. Our model demonstrates an exceptional R^2 coefficient of 0.96, implying excellent capability in explaining energy consumption variation compared to other models.

This comparative analysis compellingly highlights the superior performance of our model over prior works. The attained results indicate high accuracy, low error, and a strong ability to explain energy consumption variation. These performances enhance the credibility and effectiveness of our energy consumption prediction model within the context of our contribution.

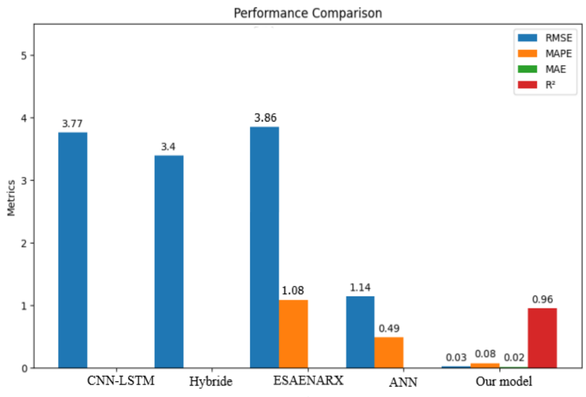

Figure 7. Comparison with related works

4 Cocnlusion & Perspectives

This study delves into the utilization of an AI model for predicting energy consumption, specifically employing time series data and the Long Short-Term Memory (LSTM) model. Our aim was to address existing limitations in the literature related to data quality and model choice. We conducted a thorough data analysis to ensure input data reliability, opting for the LSTM model due to its prowess in handling sequential problems.

Through our research, we have made notable strides in accurately forecasting energy consumption. Performance evaluation metrics, including Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), Mean Absolute Error (MAE), and Coefficient of Determination (R^2), indicate strong model performance and dependability. Notably, we achieved exceptional outcomes with RMSE = 0.03, MAPE = 0.08, MAE = 0.02, and $R^2 = 0.96$, effectively showcasing the efficacy of our approach.

By surmounting challenges related to data quality and model selection, we have contributed to the field of energy consumption prediction. Our research underscores the significance of meticulous data analysis and highlights the pertinence and efficacy of LSTM models in achieving precise energy consumption forecasts. The findings of our study hold noteworthy implications for energy management, resource allocation, and informed decision-making across diverse domains.

Subsequent work will be geared towards refining model accuracy, exploring their application in developing nations, and examining the influence of data quality on model performance. We are confident that these advancements will foster an improved grasp of and application of machine learning models in the realm of energy demand forecasting.

Acknowledgement

We would like to express our gratitude for the support extended by the Algerian General Direction of Scientific Research and Technological Development (DGRSDT) as well as the invaluable assistance provided by the LAMIS Laboratory. Their contributions have been instrumental in the realization of this work.

References

[1] S. Bourhnane, M. R. Abid, R. Lghoul, K. Zine-Dine, N. Elkamoun, and D. Benhaddou, “Machine learning for energy consumption prediction and scheduling in smart buildings,” SN Appl. Sci., vol. 2, no. 2, p. 297, Jan. 2020, doi: 10.1007/s42452-020-2024-9.

[2] O. Shaikh et al., “EnergyVis: Interactively Tracking and Exploring Energy Consumption for ML Models.” arXiv, Mar. 30, 2021. doi: 10.48550/arXiv.2103.16435.

[3] D. Hadjout, J. F. Torres, A. Troncoso, A. Sebaa, and F. Martínez-Álvarez, “Electricity consumption forecasting based on ensemble deep learning with application to the Algerian market,” Energy, vol. 243, p. 123060, Mar. 2022, doi: 10.1016/j.energy.2021.123060.

[4] S. M. Chekouri and A. Sahed, “Forecasting Electricity Consumption in Algeria Using Artificial Neural Networks,” مجلة رؤى اقتصادية, vol. 12, no. 1, pp. 247–261, Jun. 2022.

[5] N. E.-H. Bezzar, L. Laimeche, A. Meraoumia, and L. Houam, "Data analysis-based time series forecast for managing household electricity consumption," Demonstr. Math., vol. 55, no. 1, pp. 900–921, Jan. 2022, doi: 10.1515/dema-2022-0176.

[6] T.-Y. Kim and S.-B. Cho, "Predicting residential energy consumption using CNN-LSTM neural networks," Energy, vol. 182, pp. 72–81, Sep. 2019, doi: 10.1016/j.energy.2019.05.230.

[7] T. Han, K. Muhammad, T. Hussain, J. Lloret, and S. W. Baik, "An Efficient Deep Learning Framework for Intelligent Energy Management in IoT Networks," IEEE Internet Things J., vol. 8, no. 5, pp. 3170–3179, Mar. 2021, doi: 10.1109/JIOT.2020.3013306.

[8] Z. A. Khan et al., "Efficient Short-Term Electricity Load Forecasting for Effective Energy Management," Sustain. Energy Technol. Assess., vol. 53, p. 102337, Oct. 2022, doi: 10.1016/j.seta.2022.102337.

[9] S. Mujeeb and N. Javaid, "ESAENARX and DE-RELM: Novel schemes for big data predictive analytics of electricity load and price," Sustain. Cities Soc., vol. 51, p. 101642, Nov. 2019, doi: 10.1016/j.scs.2019.101642.

[10] W. Gao, A. Darvishan, M. Toghani, M. Mohammadi, O. Abedinia, and N. Ghadimi, "Different states of multi-block based forecast engine for price and load prediction," Int. J. Electr. Power Energy Syst., vol. 104, pp. 423–435, Jan. 2019, doi: 10.1016/j.ijepes.2018.07.014.

[11] "Hourly Energy Consumption | Kaggle." https://www.kaggle.com/datasets/robikscube/hourly-energy-consumption (accessed Sep. 10, 2023).

[12] "PJM - Home." https://www.pjm.com/ (accessed Sep. 10, 2023).