

Comparative study of long short-term memory (LSTM), bidirectional LSTM, and traditional machine learning approaches for energy consumption prediction

Energy Exploration & Exploitation

1–21

© The Author(s) 2024

DOI: 10.1177/01445987241269496

journals.sagepub.com/home/eea



Hamed Alizadegan¹, Behzad Rashidi Malki²,
Arian Radmehr³, Hossein Karimi⁴
and Mohsen Asghari Ilani⁵ 

Abstract

Responsible, efficient, and environmentally conscious energy consumption practices are increasingly essential for ensuring the reliability of the modern electricity grid. This study focuses on leveraging time series analysis to improve forecasting accuracy, crucial for various application domains where real-world time series data often exhibit complex, non-linear patterns. Our approach advocates for utilizing long short-term memory (LSTM) and bidirectional long short-term memory (Bi-LSTM) models for precise time series forecasting. To ensure a fair evaluation, we compare the performance of our proposed approach with traditional neural networks, time-series forecasting methods, and conventional decline curves. Additionally, individual models based on LSTM, Bi-LSTM, and other machine learning methods are implemented for a comprehensive assessment. Experimental results consistently demonstrate that our proposed model outperforms all benchmarking methods in terms of mean absolute error (MAE) across most datasets. Addressing the imbalance between activations by consumer and prosumer groups, our predictions show superior performance compared to several traditional forecasting methods, such as the autoregressive integrated moving average (ARIMA) model and seasonal autoregressive integrated moving average (SARIMA) model. Specifically, the root

¹Department of Computer and Information Technology Engineering, Qazvin Branch, Islamic Azad University, Qazvin, Iran

²Department of Computer, Islamic Azad University Bonab, Bonab, East Azerbaijan, Iran

³Department of Computer Engineering, Islamic Azad University South Tehran Branch, Tehran, Iran

⁴Department of Electrical, Computer and IT Engineering, Qazvin Branch, Islamic Azad University, Qazvin, Iran

⁵School of Mechanical Engineering, College of Engineering, University of Tehran, Tehran, Iran

Corresponding author:

Mohsen Asghari Ilani, School of Mechanical Engineering, College of Engineering, University of Tehran, Tehran, Iran.

Email: ilani.a.m1990@gmail.com



mean square error (RMSE) of Bi-LSTM is 5.35%, 46.08%, and 50.6% lower than LSTM, ARIMA, and SARIMA, respectively, on the May test data.

Keywords

Time series forecasting, long short-term memory, bidirectional long short-term memory, deep learning, autoregressive integrated moving average, seasonal autoregressive integrated moving average, energy consumption prediction

Introduction

Forecasting energy consumption is a significant and complex undertaking within both industry and academia. Precise predictions of energy consumption offer valuable insights for efficiently allocating energy resources (Liu et al., 2018), devising energy-saving strategies (Han et al., 2008), and enhancing overall energy system performance. Additionally, accurate energy predictions assist managers in conducting market research management and facilitating economic development (Pao, 2009). From an academic perspective, the advancements in energy consumption prediction can be extended to forecasting other time series, including but not limited to traffic flow (Fu et al., 2017), weather patterns (Rasp and Lerch, 2018), temperature trends (Zhang et al., 2017), stock market behavior, and solar radiation levels.

Different methods for forecasting time series have been developed in previous research (Alizadegan et al., 2024), falling into categories like statistical approaches, computational intelligence, and a blend of the two. Among the statistical techniques, autoregressive integrated moving average (ARIMA) is widely used for modeling linear time series (Kumaresan and Ganeshkumar, 2020). However, real-world time series often demonstrate non-linear characteristics (Shaeri et al., 2022), making it essential to employ non-linear modeling techniques (Panigrahi and Behera, 2017; Shaeri et al., 2022). Computational intelligence techniques, such as feedforward neural networks (NNs) (Mengesha et al., 2022; Shaeri et al., 2022), provide a means to effectively capture and model these non-linear patterns.

While conventional computational intelligence methods like feedforward neural networks can effectively model intricate patterns among samples, they struggle to capture the long-term dependencies present in time series data. As a result, recurrent neural networks (RNNs), a specialized type of artificial neural networks, have been introduced as a viable alternative for accurate time series forecasting (Chen et al., 2018).

Even though RNNs are adept at retaining sequential information, they encounter challenges such as the vanishing gradient problem, making their training difficult (Parmezan et al., 2019). Consequently, a solution has been found in the form of long short-term memory (LSTM) networks, which serve as an extension of RNNs. LSTMs have been developed to overcome the limitations of RNNs and have proven successful in processing sequence data, including applications such as natural language processing (NLP) and speech recognition (Fischer and Krauss, 2018).

Given the significant advancements observed in the aforementioned application domains and the crucial importance of precise time series forecasting in this context (Abbasimehr and Paki, 2022), the present study conducts a comparison of various time series forecasting models for predicting energy consumption prediction. The evaluated methods encompass both statistical (persistent) approaches and those rooted in artificial intelligence. The statistical models utilized in this study

fall under the category of persistence models, including Autoregressive Moving Average (ARMA), ARIMA, and Seasonal AutoRegressive Integrated Moving Average (SARIMA). Additionally, six different types of neural network (NN) models are considered: bidirectional long short-term memory (Bi-LSTM), LSTM (Alizadegan et al., 2024), fuzzy C-mean clustering, multi-layer perceptron (MLP) and feedforward neural networks (Sharadga et al., 2020).

This study addresses a gap in existing literature on energy consumption forecasting by comparing traditional statistical models (ARIMA, SARIMA) and advanced deep learning approaches (LSTM, Bi-LSTM). Traditional models often fail to capture the non-linear, dynamic patterns of energy consumption data influenced by variables like weather and socio-economic factors. This research aims to provide empirical evidence of the relative strengths and weaknesses of these methodologies for more accurate predictions. For this purpose, we contribute to the field of energy consumption forecasting through a systematic evaluation of LSTM and Bi-LSTM deep learning models against traditional ARIMA and SARIMA methods.

This study is guided by the following research questions: How do deep learning models like LSTM and Bi-LSTM compare to traditional statistical methods such as ARIMA and SARIMA in forecasting energy consumption? What are the strengths and weaknesses of each approach in accurately predicting complex, non-linear patterns in time series data?

Additionally, we hypothesize that incorporating temporal features enhances the predictive accuracy of LSTM and Bi-LSTM models for energy consumption forecasting. We also propose that the performance of these models remains consistent across various datasets, regardless of the scale and type of energy consumption patterns they represent.

In our comparative analysis of energy consumption prediction models, the utilization of LSTM and Bi-LSTM networks consistently demonstrated superior predictive performance over traditional statistical models such as ARIMA and SARIMA. LSTM and Bi-LSTM models exhibited enhanced accuracy in capturing intricate, non-linear patterns and prolonged dependencies within the data, as evidenced by notably lower mean absolute error (MAE) and root mean square error (RMSE) metrics. Particularly noteworthy was Bi-LSTM's ability to leverage bidirectional processing, resulting in the highest accuracy among the models tested. In contrast, ARIMA and SARIMA, while effective in modeling linear trends and seasonal variations, struggled to adequately handle the non-linear dynamics inherent in energy consumption data, thereby yielding higher prediction errors compared to the deep learning approaches. This underscores the superior capability of LSTM and Bi-LSTM networks in addressing the complexities and nuances present in real-world energy consumption forecasting tasks.

While existing literature has extensively discussed individual methodologies, there is a clear need for comprehensive comparative analyses to quantitatively assess the effectiveness of these models in capturing complex temporal patterns and enhancing forecasting accuracy. This comparative study aims to provide empirical evidence and insights that can guide decision-making in energy management and inform future advancements in predictive modeling techniques. The results of our study are compared with the recent publish papers in Web of Science (WoS) in section of "Results and Discussion". In the "Methodology" section, we detail the approaches and frameworks utilized to investigate our research questions. Following this, the "Results and Discussion" section presents the empirical findings and their interpretations. Finally, the "Conclusion" section synthesizes these findings, discusses their implications, and suggests avenues for future research.

Related work

In recent years, RNNs have gained considerable traction in the realm of time series forecasting due to their inherent suitability for sequence modeling tasks. The effectiveness of RNNs, particularly

the LSTM network, has been validated in several recent forecasting studies (Abbasimehr et al., 2020; Fischer and Krauss, 2018; Gundu and Simon, 2021; Kulshrestha et al., 2020; Law et al., 2019). For instance, Gundu and Simon (Gundu and Simon, 2021) proposed an LSTM model for electricity price forecasting, leveraging particle swarm optimization (PSO). Abbasimehr et al. (Abbasimehr et al., 2020) developed an optimized stacked LSTM model for demand forecasting in a furniture company, surpassing conventional benchmark models. Similarly, Fischer and Krauss (Fischer and Krauss, 2018) investigated LSTM networks' performance in financial market forecasting, demonstrating their superiority over standard methods. Law et al. (Law et al., 2019) proposed a deep learning framework applied to tourism demand forecasting. In a related study, Kulshrestha et al. (Kulshrestha et al., 2020) presented a combined model integrating bidirectional LSTM and Bayesian optimization (BO) for tourism demand forecasting, showing improved performance compared to popular methods such as support vector regression (SVR), radial basis function neural networks (RBFNN), and autoregressive distributed lag model (ADLM) (Hosseini Rad et al., 2022; Ilani and Khoshnevisan, 2021, 2022). In another study compared two deep learning models, LSTM and Bi-directional LSTM (Bi-LSTM), for short-term univariate electric consumption forecasting across diverse datasets. Statistical evidence, including Friedman's test, indicates that Bi-LSTM significantly outperforms LSTM, demonstrating its robustness across different scales of electric power consumption (da Silva and de Moura Meneses, 2023).

Moreover, various models, including neural networks and decomposition techniques, have been developed to forecast solar irradiance. The WPD-based model achieved the lowest RMSE and MAE for Indian locations (Singla et al., 2023a). A hybrid RLMD Bi-LSTM model integrating RLMD and bidirectional LSTM significantly improved RMSE and MAE, outperforming traditional and RLMD-based models for short-term forecasts in Hisar and Jaipur (Singla et al., 2023b). A dual decomposition-based error correction model using CEEMDAN and VMD with bidirectional LSTM networks achieved lower RMSE and MAPE, validated by statistical tests, demonstrating robustness across multiple locations (Singla et al., 2022a).

Methodology

Predicting time series data that exhibits chaos, uncertainty, randomness, periodicity and nonlinearity is a significant challenge. This section presents the proposed framework for accurately predicting long-term energy consumption, with a specific focus on addressing distinct periodic patterns. The methodologies utilized are detailed below.

ARIMA and SARIMA

The ARIMA model proves valuable in time series forecasting by leveraging past values in the series. Accurate forecasting holds significance for cost saving measures, effective planning and production activities. When forecasting future values using historical data from a time series, it's known as univariate time series forecasting. On the other hand, if the series isn't utilized for prediction, it's termed as multivariate time series forecasting. ARIMA is proficient in predicting future values based on its own historical data, integrating lagged values and forecast error lags.

The ARIMA model comprises three key components: $ARIMA(p, d, q)$ which known as ARIMA notation. The order of the autoregressive (AR) term is represented by p , indicating a linear regression model that includes its own lagged values as predictors. The inclusion of differencing (d) is crucial for making predictors independent and ensuring that the series achieves stationarity. The differencing parameter is set to 0 when the series is already stationary. The order of the MA

component is denoted by q , which represents the number of lagged forecast errors (LFEs). If p represents the lag numbers of Y , where Y is utilized as the predictors, the ARIMA model can be expressed for time series prediction at time t , as illustrated in Equation (1):

$$Y_t = \text{Constant} + LY + LFE \quad (1)$$

where, LY: Lags of Y and LFE: Lagged forecast errors.

The utilized combination involves a linear combination of lags, with the primary objective being the identification of appropriate values for p , d , and q . The selection of the minimum difference d is crucial, and it should be chosen based on achieving zero autocorrelation (AC). The determination of p is associated with the order of the autoregressive (AR) component, which should equal the lags in the partial autocorrelation (PAC) surpassing the set significance threshold. PAC represents a conditional correlation.

Equation (2) demonstrates the PAC, where y represents the response variable and x_1 , x_2 , and x_3 denote predictor variables. More precisely, Eq. (2) delineates the PAC between y and x_3 , computed as the correlation between the regression residuals of y on x_1 and x_2 , and the residuals of x_3 on x_1 and x_2 .

$$PAC = \frac{\text{cov}(y, x_3|x_1, x_2)}{\sqrt{\text{var}(y|x_1, x_2)\text{var}(x_3|x_1, x_2)}} \quad (2)$$

Equation (3) denotes the h^{th} order PAC for time series data.

$$\text{PAC for the } h\text{th order} = \frac{\text{cov}(y_i, y_{i-h}|y_{i-1}, \dots, y_{i-h+1})}{\sqrt{\text{var}(y_i|y_{i-1}, \dots, y_{i-h+1}) \text{var}(y_{i-h}|y_{i-1}, \dots, y_{i-h+1})}} \quad (3)$$

The order of the MA, denoted as (q) is determined by analyzing the autocorrelation function (ACF), which illustrates the error associated with the lagged forecast. This calculation is depicted in Equation (4).

$$AC = \frac{\sum_{i=1}^{N-k} (y_i - \bar{y})(y_{i+k} - \bar{y})}{\sum_{i=1}^N (y_i - \bar{y})^2} \quad (4)$$

The mean of the time series is denoted by \bar{y} .

The lag is represented by k .

N represents the complete series value.

In cases where there is a need to account for seasonal patterns in the time series, a seasonal term is incorporated into the ARIMA model. This results in the seasonal ARIMA model, denoted as SARIMA, as its notation is provided below:

$$\text{ARIMA}(p, d, q) \times (P, D, Q)_S$$

In this context, the notation (p, d, q) refers to the non-seasonal components, while $(P, D, Q)_S$ represents the seasonal components of the model, with S indicating the period number within a season. This study utilizes SARIMA, as opposed to ARIMA, because ARIMA does not account for time series with seasonal components. SARIMA is particularly suitable for univariate data that exhibit seasonality and trend exploration (Siarni-Namini et al., 2018).

Recurrent neural network

RNNs are a specialized class of neural networks designed to recognize patterns in sequences of data, such as time series or natural language. Among the various forms of RNNs, LSTM and

Bi-LSTM stand out due to their capability to retain and utilize information over extended sequences, making them highly effective for tasks that involve sequential data prediction.

Long short-term memory

LSTMs are designed to address the vanishing gradient problem inherent in traditional RNNs by using unique memory blocks called cells, which include three primary gates: input, forget, and output gates. These gates control the flow of information through the cell.

The forget gate determines which information from the previous cell state should be discarded, and is described by the equation: (Singla et al., 2022c):

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

where σ is the sigmoid function, W_f is the weight matrix, and b_f is the bias vector.

The input gate updates the cell state with new information, using the equations:

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i)$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C)$$

where W_i and W_C are weight matrix, and b_i and b_C are bias vectors.

The cell state is then updated by combining the previous cell state and the new candidate cell state:

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t$$

Finally, the output gate determines the output of the current cell state, using:

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o)$$

$$h_t = o_t * \tanh(C_t)$$

where W_o is the weight matrix and b_o is the bias vector.

These equations show how LSTM networks process inputs by considering both the current input and the past hidden state, allowing them to capture long-term dependencies in the data.

Bidirectional LSTM

Bi-LSTM networks extend the capabilities of standard LSTMs by processing data in both forward and backward directions, which allows the model to leverage context from both the past and the future for improved prediction accuracy.

In a Bi-LSTM, the forward LSTM processes the sequence from start to end, described by: (Singla et al., 2022b)

$$\vec{f}_t = \sigma(W_{\vec{f}} \cdot [\vec{h}_{t-1}, x_t] + b_{\vec{f}})$$

$$\vec{i}_t = \sigma(W_{\vec{i}} \cdot [\vec{h}_{t-1}, x_t] + b_{\vec{i}})$$

$$\vec{\tilde{C}}_t = \tanh(W_{\vec{C}} \cdot [\vec{h}_{t-1}, x_t] + b_{\vec{C}})$$

$$\vec{C}_t = \vec{f}_t * \vec{C}_{t-1} + \vec{i}_t * \vec{\tilde{C}}_t$$

$$\vec{o}_t = \sigma(W_{\vec{o}} \cdot [\vec{h}_{t-1}, x_t] + b_o)$$

$$\vec{h}_t = \vec{o}_t * \tanh(\vec{C}_t)$$

Simultaneously, the backward LSTM processes the sequence in reverse:

$$\overleftarrow{f}_t = \sigma(W_{\overleftarrow{f}} \cdot [\overleftarrow{h}_{t+1}, x_t] + b_{\overleftarrow{f}})$$

$$\overleftarrow{i}_t = \sigma(W_{\overleftarrow{i}} \cdot [\overleftarrow{h}_{t+1}, x_t] + b_{\overleftarrow{f}})$$

$$\overleftarrow{C}_t = \tanh(W_{\overleftarrow{C}} \cdot [\overleftarrow{h}_{t+1}, x_t] + b_{\overleftarrow{C}})$$

$$\overleftarrow{C}_t = \overleftarrow{f}_t * \overleftarrow{C}_{t-1} + \overleftarrow{i}_t * \overleftarrow{C}_t$$

$$\overleftarrow{o}_t = \sigma(W_{\overleftarrow{o}} \cdot [\overleftarrow{h}_{t+1}, x_t] + b_{\overleftarrow{o}})$$

$$\overleftarrow{h}_t = \overleftarrow{o}_t * \tanh(\overleftarrow{C}_t)$$

The final output of the Bi-LSTM, y_t , is obtained by combining the forward and backward outputs using an activation function s :

$$y_t = s(\vec{h}_t, \overleftarrow{h}_t)$$

By leveraging information from both directions, Bi-LSTM models can capture a more comprehensive understanding of the sequence, leading to improved performance on tasks involving sequential data.

Metrics

RMSE is widely used for evaluating solar forecasting models because it penalizes large errors, which are highly undesirable in this context. MAE is also used but are scale-dependent and best suited for results with Gaussian distributions (View of Review of Different Error Metrics: A Case of Solar Forecasting, n.d.). MAE measures the average absolute difference between the predicted and actual prices over the entire test dataset.

Mathematically, MAE is calculated as:

$$\text{MAE} = \frac{1}{t} \sum_{t=1}^T |y_t - y'_t|$$

RMSE is another commonly used metric that penalizes larger prediction errors more heavily than smaller errors. It is calculated as the square root of the average of the squared differences between the predicted and actual prices (Wu et al., 2018):

$$\text{RMSE} = \sqrt{\frac{1}{t} \sum_{t=1}^T (y_t - y'_t)^2}$$

Where, y_t represent the true price of Bitcoin, y'_t denote the predicted price of Bitcoin, and T indicate the specific period for prediction (Figure 1).

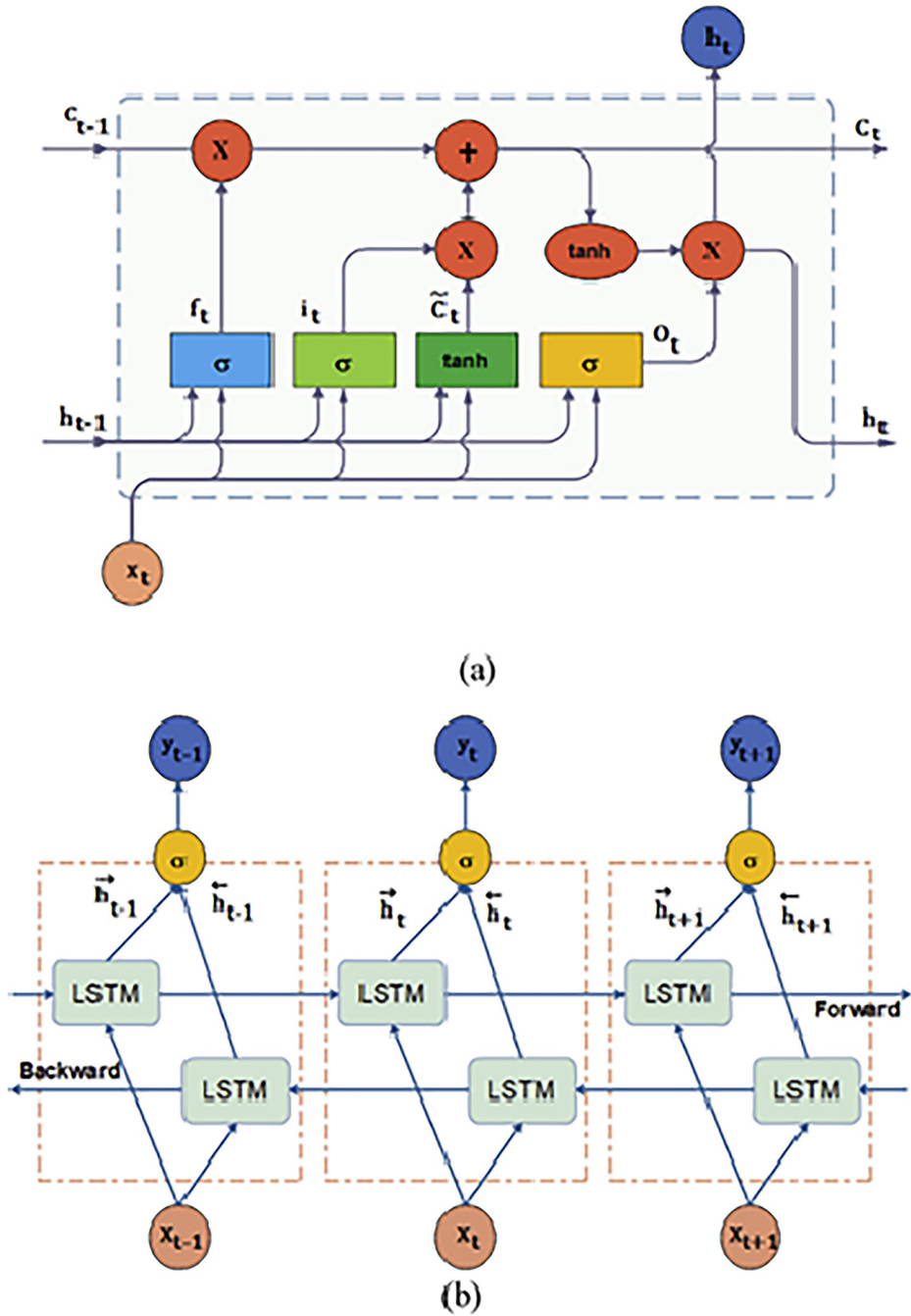


Figure 1. Architecture of (a) LSTM (b) Bi-LSTM (Sharadga et al., 2020).

Results and discussion

Data visualization

In Figure 2, the temporal dynamics of energy consumption are illustrated over the observed period. The y-axis represents energy consumption, and the x-axis represents time, offering a comprehensive depiction of the fluctuations in energy usage throughout the timeline.

The plot unveils a distinct pattern, indicating that energy consumption typically hits its lowest points from March to September. In contrast, the months spanning December to February showcase elevated energy consumption levels, reaching a peak during this timeframe. This observable seasonality implies a cyclical trend that may be influenced by diverse external factors, including weather conditions, holidays, or industrial patterns.

Insights drawn from the time series plot play a crucial role in preprocessing time series data. A comprehension of recurring patterns enables well-informed decisions regarding data transformation, feature engineering, and model selection. For instance:

- **Seasonal decomposition:** Addressing the observed seasonality can be achieved through techniques like seasonal decomposition. This process separates the time series into trend, seasonal and residual components, facilitating the isolation of patterns and enhancing the model's capability to capture underlying dynamics.
- **Feature Engineering:** Knowledge of monthly energy consumption trends allows for the creation of additional features, such as binary indicators for high or low consumption months. These engineered features contribute to the model's adaptability to specific patterns.
- **Model Selection:** The distinctive seasonality evident in the plot guides the selection of appropriate time series models. Models like SARIMA, explicitly designed to account for seasonal variations, can be considered to effectively capture the observed patterns.

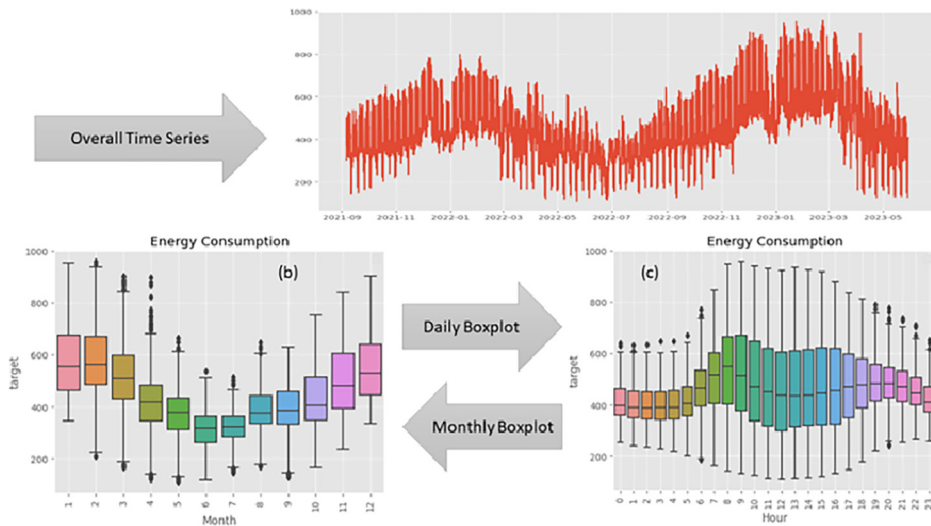


Figure 2. Time series plot of energy consumption. (a) Overall time series. (b) Daily time series boxplot. (c) Monthly time series boxplot.

The density plot discernibly exposes a conspicuous concentration of variable values within the interval [1200, 1600]. The heightened density within this range denotes a substantial aggregation of data points, suggesting a critical region within the variable’s spectrum.

This asymmetry signifies the existence of data points beyond the primary density concentration, indicating potential outliers or instances of extreme values that diverge from the central tendency (Figure 3).

In summary, the time series plot serves as a valuable tool in the preprocessing stage, enabling data scientists to make informed choices regarding feature engineering, model selection, and other relevant preprocessing steps.

Kaggle dataset

Enefit, a powerhouse in the Baltic energy sector, stands at the forefront of green energy solutions. Their expertise guides customers on a personalized and flexible journey towards sustainable practices, implementing environmentally friendly energy solutions. Despite being armed with internal predictive models and third-party forecasts to tackle energy imbalances, existing methods fall short in accurately predicting the intricate behaviors of prosumers, resulting in substantial imbalance costs.

The use of this Kaggle dataset is justified due to its quality and relevance for studying energy consumption patterns, which is critical for accurate solar irradiance forecasting. This dataset, curated from Kaggle, provides 15,312 observations featuring two key variables: date and time, and energy consumption (Enefit - Predict Energy Behavior of Prosumers | Kaggle, n.d.). This dataset offers a comprehensive perspective on Estonia’s energy dynamics, ensuring no

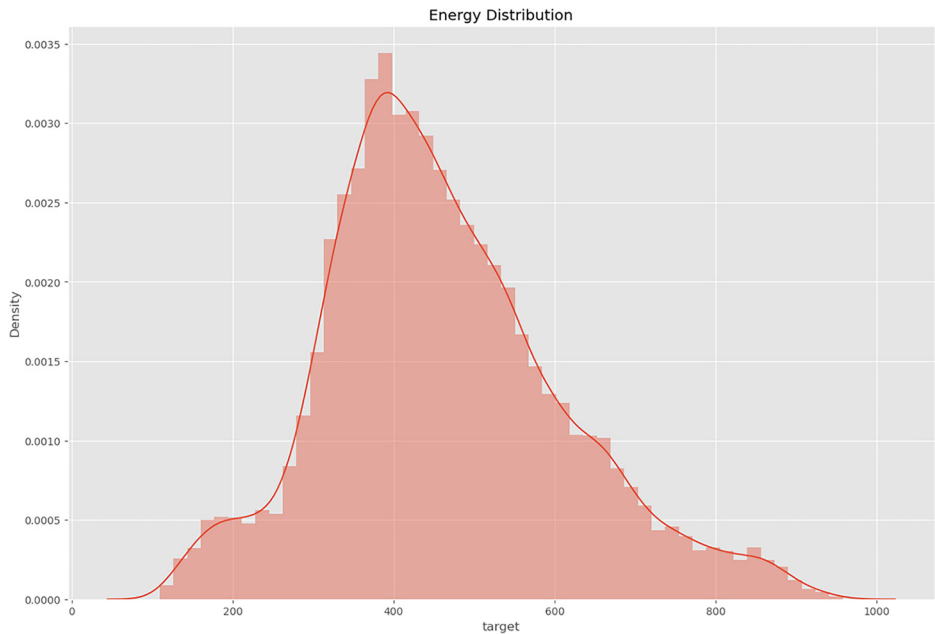


Figure 3. The density plot discernibly exposes a conspicuous concentration of variable values within the interval [1200, 1600].

duplications. The inclusion of date and time as features is crucial for capturing temporal patterns, while energy consumption serves as the target variable essential for training the forecasting models.

To ensure robust model evaluation, the dataset is split into three parts: 80% for training (12,248), 10% for validation (1532), and 10% for testing (1532). This split allows for thorough training of the model, fine-tuning through validation, and unbiased performance evaluation on the test set.

Acknowledging the constraints of existing forecasting methods, the focus shifts to machine learning (ML), specifically within the domain of time series prediction for energy scenarios. In the face of a progressively intricate energy balance landscape, this study explores the utilization of deep neural networks. Through a thorough examination of complex and imbalanced datasets, considering runtime efficiency and model precision, the research contrasts conventional ML approaches such as ARIMA and SARIMA with advanced RNNs, specifically LSTM and Bi-LSTM.

System requirements

The experiments were conducted on the Kaggle platform, an open-source notebook environment provided by OpenML. This platform offers access to both free and premium GPU (P100) and TPU resources, which are highly beneficial for academic and research purposes. The models were trained using four GPUs, each with 16 GB of RAM. The implementation was carried out in Python, utilizing TensorFlow for the backend and Keras APIs for the frontend of the system.

Prediction analysis

LSTM (long short-term memory). In our quest to decipher the intricate patterns within time series data, we harnessed the power of LSTM networks. The results, as depicted in Figure 4, showcase

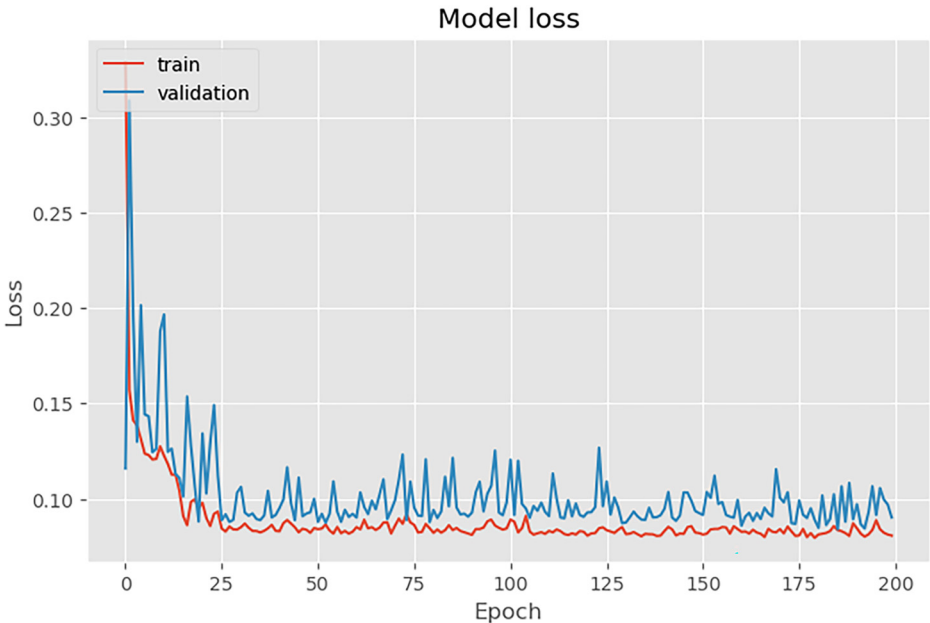


Figure 4. LSTM training and validation losses.

the convergence of training and validation losses, portraying a model finely tuned and adept at extracting intricate patterns from real-world datasets (Figure 5).

Figures 6 and 7 demonstrate the forecasting capabilities of our LSTM model. Red lines represent actual values, while blue lines depict predicted values. These figures illustrate the LSTM's effectiveness in capturing periodic sequences. The fluctuations between real and predicted values reflect the precision and accuracy inherent in our LSTM model. Our LSTM model excels in accuracy metrics when analyzing complex time series data, revealing temporal patterns with detailed insight.

The results we have obtained from our LSTM model are typically metrics used to evaluate the performance of regression models, particularly in time series forecasting tasks. Here's a brief description of each metric:

- **MAE (mean absolute error):**

- **Test Data:** This metric measures the average absolute difference between the actual and predicted values in our test dataset. In our case, the MAE for the test data is 0.097, indicating an average absolute error of approximately 0.097 units.
- **Train Data:** Similarly, for the training dataset, the MAE is 0.079. This suggests an average absolute error of around 0.079 units between the actual and predicted values during the training phase.

- **RMSE (root mean squared error):**

- **Test Data:** RMSE is a measure of the average magnitude of the errors between predicted and actual values, giving more weight to larger errors. Our RMSE for the test data is 0.1308, indicating the square root of the average squared differences between the predicted and actual values.
- **Train Data:** For the training dataset, the RMSE is 0.112, representing the square root of the average squared errors during the training phase.

These metrics provide insights into the accuracy of our LSTM model. Lower MAE and RMSE values generally indicate better performance, as they suggest smaller errors between predicted

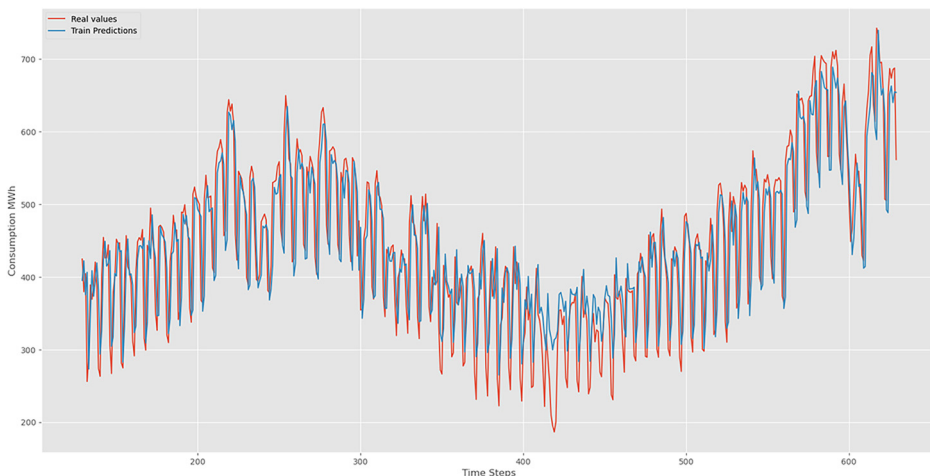


Figure 5. LSTM training prediction.

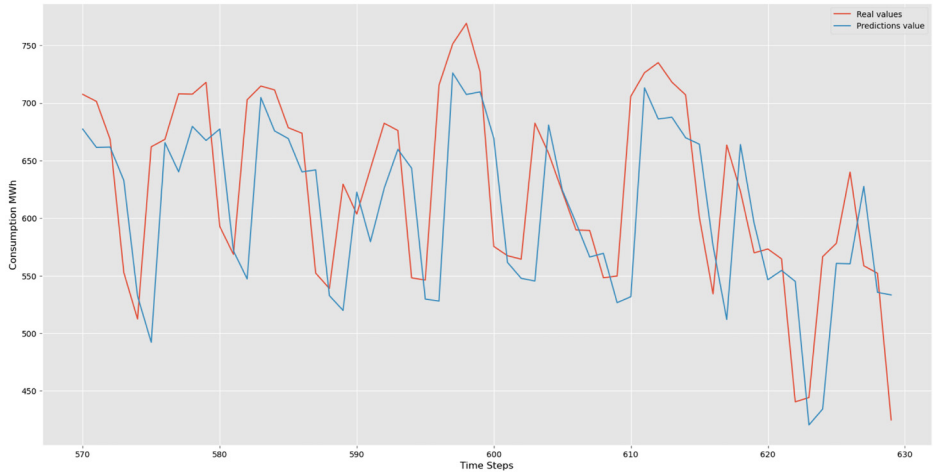


Figure 6. LSTM validation prediction.

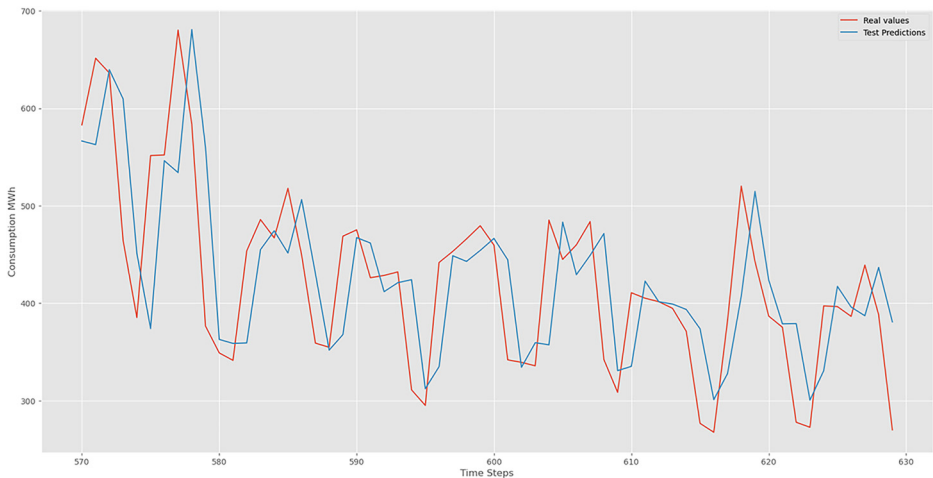


Figure 7. LSTM test prediction.

and actual values. However, the interpretation of these metrics depends on the specific characteristics and scale of our data.

Bi-LSTM (bidirectional LSTM). Training loss is compared to the validation loss. The result shows that the two values were low, and no overfitting was detected. The results are shown in Figure 8.

Examples of consumption predictions of the LSTM model were done using the training dataset (Figure 9), validation dataset (Figure 10), and test datasets (Figure 11).

The predictive capabilities of our Bi-LSTM model shine through the lens of performance metrics, revealing a nuanced understanding of its effectiveness. In the realm of time series forecasting, the Bi-LSTM model has showcased its prowess on both training and test datasets, delivering impressive results. It is provided the performance metrics as following below:

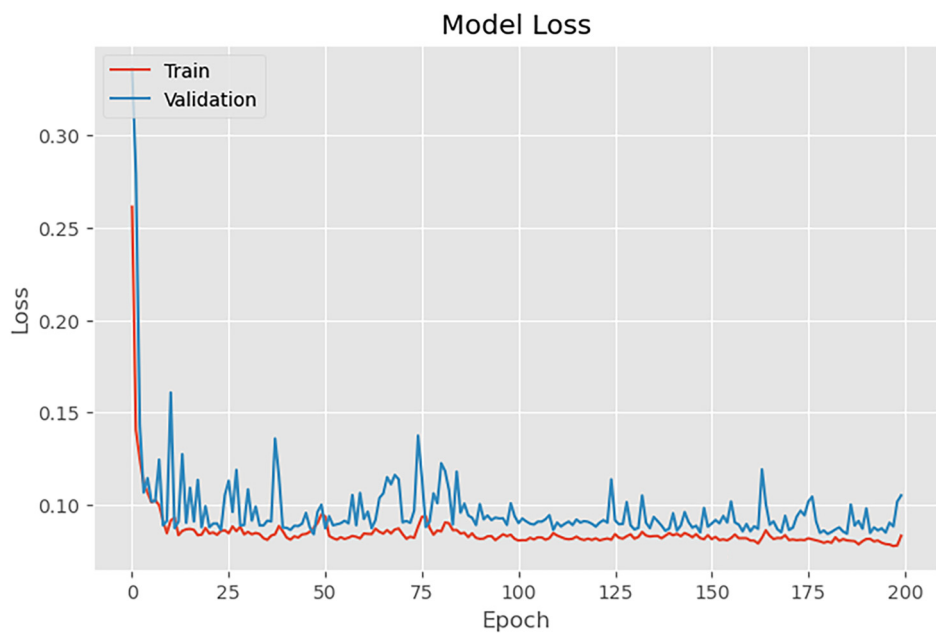


Figure 8. Bi-LSTM training and validation loss.

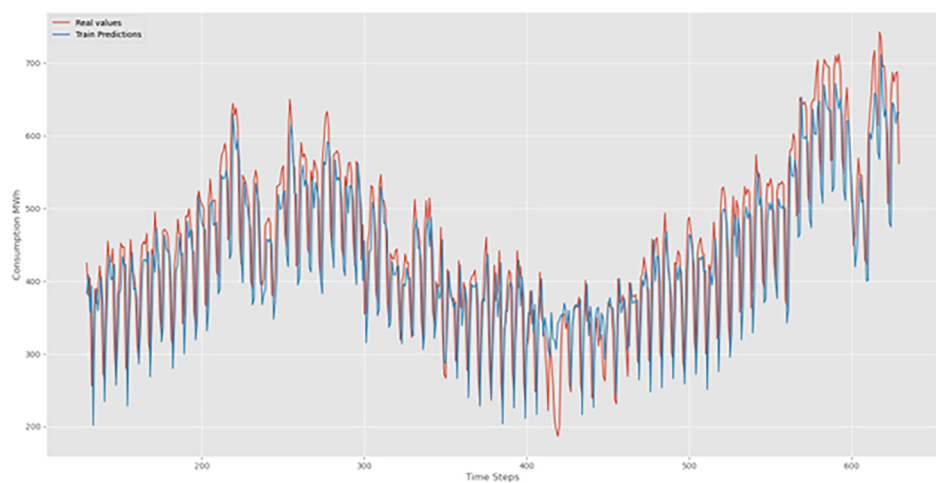


Figure 9. Bi-LSTM training prediction.

RMSE (root mean square error):

- Train Data: 0.108
- Test Data: 0.124

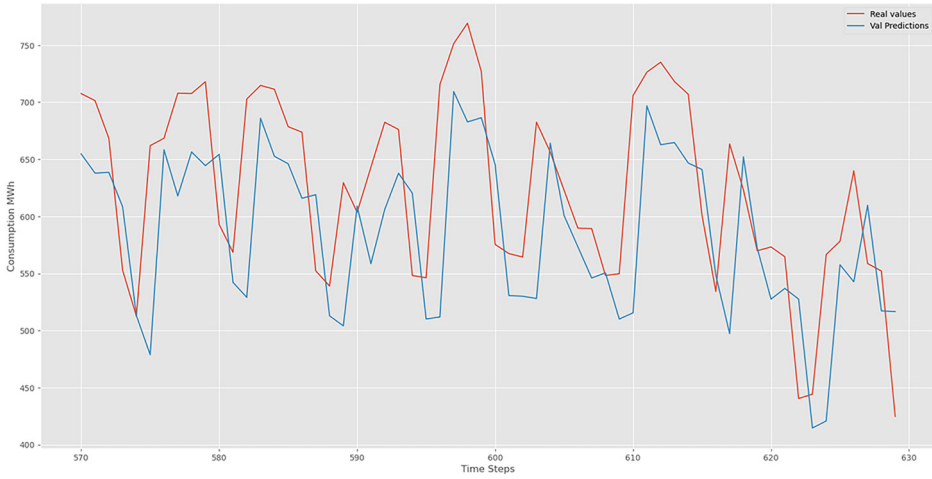


Figure 10. Bi-LSTM validation prediction.

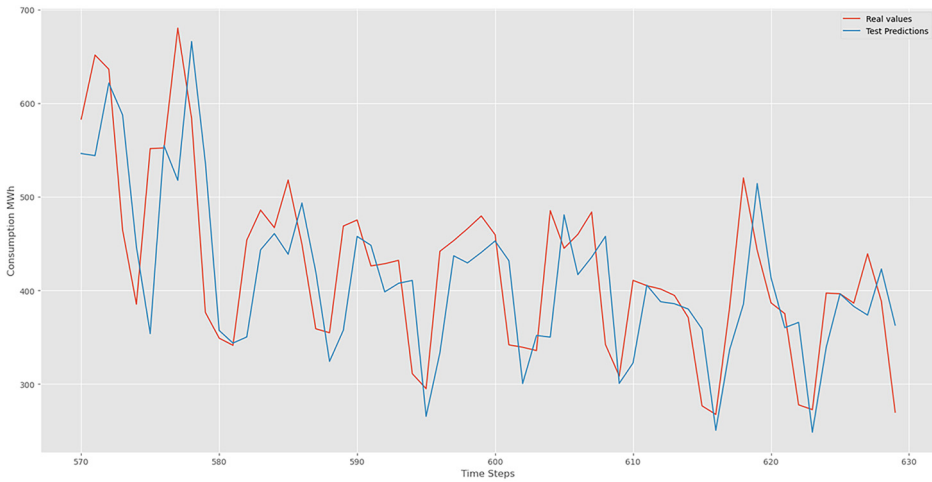


Figure 11. Bi-LSTM test prediction.

MAE (mean absolute error):

- Train Data: 0.076
- Test Data: 0.092

The evaluation metrics highlight the Bi-LSTM model's proficiency in handling the dataset's complexities. The lower RMSE and MAE values indicate its accuracy in predicting the target variable, demonstrating a robust fit to both the training and test datasets. These metrics quantify the model's performance, confirming its potential for accurate and reliable predictions in time series analysis.

SARIMA (seasonal autoregressive integrated moving average). In Figure 12, we present the captivating visual representation of predictions generated by the SARIMA model. This plot gracefully intertwines the actual observed values and the SARIMA-predicted values, painting a vivid picture of the model’s forecasting capabilities.

The red line gracefully traces the chronological path of the true values, providing a benchmark for the model’s accuracy. In harmonious contrast, the blue line intricately weaves through time, representing the SARIMA model’s predictions. The seamless alignment of these lines signifies the model’s ability to capture the inherent patterns and seasonal fluctuations present in the time series data. Noteworthy is the model’s proficiency in handling complex temporal dynamics, as evidenced by its adept prediction of future values. The periodicity and trends ingrained in the dataset are eloquently mirrored by the SARIMA model, establishing its credibility in time series forecasting. The minutely detailed interplay between actual and predicted values showcases the SARIMA model’s precision, making it a valuable tool for understanding and anticipating temporal patterns in the dataset.

The SARIMAX results (Figure 13) provide a comprehensive overview of the predictive model’s performance on the energy dataset. The model, specified as $ARIMA(2, 1, 1) \times (1, 1, 1, 12)$, showcases a detailed breakdown of coefficients, standard errors, statistical significance, and various diagnostic metrics. The SARIMAX model provides valuable insights into the temporal dynamics of the energy dataset. However, the diagnostic metrics highlight areas for improvement, such as addressing potential seasonality and autocorrelation in the residuals to enhance the model’s forecasting accuracy.

RMSE (root mean square error):

- Test Data: 0.251

MAE (mean absolute error):

- Test Data: 0.203

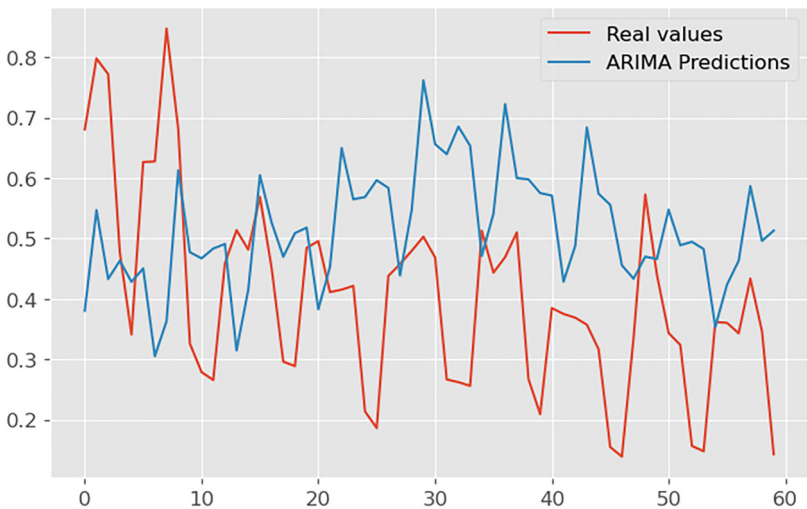


Figure 12. SARIMAX training loss.

The performance metrics of four models—LSTM, Bi-LSTM, ARIMA, and SARIMAX—are compared based on MAE and RMSE for both training and testing datasets, as presented in Table 1.

The Bi-LSTM model demonstrates slightly better performance than the LSTM model, with lower Train MAE (0.076 vs. 0.079) and Train RMSE (0.108 vs. 0.112). On the test set, Bi-LSTM also outperforms LSTM, achieving lower Test MAE (0.092 vs. 0.097) and Test RMSE (0.124 vs. 0.1308). This indicates Bi-LSTM’s enhanced ability to capture complex patterns and dependencies in the data.

Comparing LSTM and Bi-LSTM with ARIMA and SARIMAX, it is evident that LSTM and Bi-LSTM significantly outperform the traditional statistical models. For example, the Train MAE for LSTM (0.079) and Bi-LSTM (0.076) are much lower than ARIMA (0.1734) and SARIMAX (0.1823). Similarly, on the test data, both LSTM and Bi-LSTM achieve much lower Test MAE and Test RMSE compared to ARIMA and SARIMAX. Bi-LSTM has the lowest Test RMSE (0.124), followed by LSTM (0.1308), while ARIMA and SARIMAX exhibit higher errors (Test RMSE: 0.2309 and 0.2513, respectively).

Overall, Bi-LSTM emerges as the best-performing model among the four, followed closely by LSTM. Both deep learning models outperform the traditional ARIMA and SARIMAX models, which exhibit higher error rates in both training and testing phases. This analysis underscores

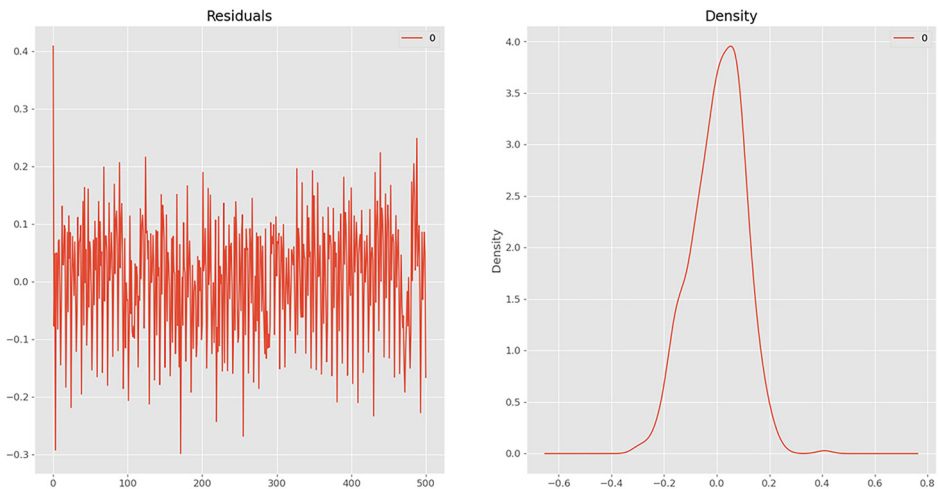


Figure 13. SARIMAX prediction: (a) residuals and (b) density.

Table 1. Comparison metrics of models performance.

Models/metrics	Train MAE	Train RMSE	Test MAE	Test RMSE
LSTM	0.079	0.112	0.097	0.1308
Bi-LSTM	0.076	0.108	0.092	0.124
ARIMA	0.1734	0.2128	0.1927	0.2309
SARIMAX	0.1823	0.2216	0.2034	0.2513

the efficacy of advanced NN architectures, particularly Bi-LSTM, in delivering more accurate energy consumption forecasts compared to traditional statistical methods. The results suggest that for applications requiring high precision in time series forecasting, leveraging deep learning approaches like LSTM and Bi-LSTM is advantageous.

On the other hand, to compare externally our work with the recent research worked on the same dataset, Chandran and Narayanan (2024) were the only researchers to have worked on this dataset, using Linear Regression, MLP, TabNet, XGBoost, LightGBM, and VotingRegressor-based ensemble models. Their work resulted in a MAE of 74.56, with the VotingRegressor ensemble being the best-performing model with an MAE of 70.16. Since this Kaggle dataset was launched only a few months ago, no other research has been published on it, and no standardized processing methods have been established. As a result, we are unable to provide a proper comparison of our results with external published papers.

Limitations

Despite the promising results, this study acknowledges several limitations that warrant future research. First, the computational complexity of LSTM and Bi-LSTM models requires significant resources and training time, and optimization techniques to reduce this overhead were not explored. Second, the models were tested on specific datasets, and their generalizability to other types of time series data remains unverified, necessitating broader dataset evaluations. Third, the study focused on univariate forecasting without incorporating additional contextual features, such as weather conditions or economic indicators, which could enhance accuracy. Lastly, deep learning models, despite their accuracy, often lack interpretability compared to traditional statistical models, highlighting the need for improved transparency and understanding in future studies.

Future study

Future research should aim to address the limitations identified in this study to improve the applicability and performance of time series forecasting models. Optimization techniques need to be explored to reduce the computational complexity and training time of LSTM and Bi-LSTM models, making them more feasible for real-time applications. Additionally, evaluating these models on a wider range of datasets will help determine their generalizability and robustness across different domains. Incorporating multivariate forecasting by including contextual features, such as weather conditions and economic indicators, could enhance predictive accuracy. Lastly, developing methods to improve the interpretability of deep learning models will be essential for making these advanced techniques more accessible and transparent to stakeholders and decision-makers.

Conclusion

Figures 6, 7, 10, and 11 shows some delay between actual and predicted values in energy consumption. This discrepancy is attributed to the dynamic nature of energy consumption patterns among consumers, which do not consistently adhere to predefined patterns due to various influencing factors. In our subsequent study, we will explore the influence of environmental parameters on energy consumption prediction. Environmental factors such as weather conditions, seasonal variations, and other external influences can significantly impact energy usage patterns.

Predicting energy consumption has become increasingly vital in our daily lives, given its significant economic implications. Various methods have been devised for energy consumption

forecasting. However, conventional techniques often fall short as they fail to capture the periodic patterns inherent in energy consumption data. This paper introduces a comprehensive approach to time series prediction with periodicity, leveraging LSTM and Bi-LSTM as RNNs, in conjunction with ARIMA and SARIMA as ML models. The comparison is applied to these models based on the datetime feature under one-step-ahead forecasting. The important findings of this study are brought below to get a look at easily:

- The proposed model presents a promising approach for forecasting the time-series energy generated by both consumers and prosumers. It offers an alternative solution for delivering reliable predictions.
- Utilizing the time variable enables precise capture of periodicity. Incorporating this variable into the LSTM model enhances accuracy in predicting energy consumption. Furthermore, the Bi-LSTM method demonstrates superior prediction performance compared to LSTM, ARIMA, and SARIMA models.
 - The RMSE of Bi-LSTM is 5.35% lower than LSTM, 46.08% lower than ARIMA and 50.6% lower than SARIMA in the forecasting of long-term time series.
 - The MAE of Bi-LSTM is 5.15% lower than LSTM, 52.08% lower than ARIMA and 54.18% lower than SARIMA in the forecasting of long-term time series.
- Optimal parameter configuration plays a pivotal role in determining the performance of the LSTM model. Careful consideration should be given to selecting the training epoch to prevent insufficient training and overfitting issues. Introducing additional hidden layers can enhance the accuracy of both Bi-LSTM and LSTM models to some degree, albeit at the expense of increased computational time. This research showcases the promising potential of the proposed approach in forecasting energy consumption. Future investigations will concentrate on developing a hybrid model that combines LSTM with other forecasting techniques for enhanced accuracy in energy consumption predictions.

Declaration of conflicting interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Mohsen Asghari Ilani  <https://orcid.org/0000-0003-0189-6956>

References

- Abbasimehr H and Paki R (2022) Improving time series forecasting using LSTM and attention models. *Journal of Ambient Intelligence and Humanized Computing* 13(1): 673–691.
- Abbasimehr H, Shabani M and Yousefi M (2020) An optimized model using LSTM network for demand forecasting. *Computers & Industrial Engineering* 143: 106435.
- Alizadegan H, Radmehr A and Ilani MA (2024) Forecasting Bitcoin prices: A comparative study of machine learning and deep learning algorithms. *Research Square*. Preprints. DOI: 10.21203/RS.3.RS-4390390/V1.
- Alizadegan H, Radmehr A, Karimi H, et al. (2024) *Solar Energy Production Forecasting: A Comparative Study of Bi-LSTM, LSTM, XGBoost Models with Activation Function Analysis*. Preprints. Epub ahead of print 15 May 2024. DOI: 10.20944/PREPRINTS202405.0994.V1.

- Chandran P and Narayanan A (2024) Ensembling is all you need? Evaluating machine learning models on predicting the energy imbalance of prosumers. In: Proceedings—11th International Conference on Signal Processing and Integrated Networks (SPIN) 2024. Institute of Electrical and Electronics Engineers Inc.: 291–296.
- Chen W, Yeo CK, Lau CT, et al. (2018) Leveraging social media news to predict stock index movement using RNN-boost. *Data & Knowledge Engineering* 118: 14–24.
- da Silva DG and de Moura Meneses AA (2023) Comparing long short-term memory (LSTM) and bidirectional LSTM deep neural networks for power consumption prediction. *Energy Reports* 10: 3315–3334.
- Enefit - Predict Energy Behavior of Prosumers | Kaggle (n.d.) Available at: <https://www.kaggle.com/competitions/predict-energy-behavior-of-prosumers/data> (accessed 27 June 2024).
- Fischer T and Krauss C (2018) Deep learning with long short-term memory networks for financial market predictions. *European Journal of Operational Research* 270(2): 654–669.
- Fu R, Zhang Z and Li L (2017) Using LSTM and GRU neural network methods for traffic flow prediction. In: Proceedings - 2016 31st Youth Academic Annual Conference of Chinese Association of Automation, YAC 2016, pp.324–328.
- Gundu V and Simon SP (2021) PSO–LSTM for short term forecast of heterogeneous time series electricity price signals. *Journal of Ambient Intelligence and Humanized Computing* 12(2): 2375–2385.
- Han B, Zhang D and Yang T (2008) Energy consumption analysis and energy management strategy for sensor node. In: Proceedings of the 2008 IEEE International Conference on Information and Automation, ICIA 2008, pp.211–214.
- Hosseini Rad R, Baniasadi S, Yousefi P, et al. (2022) Presented a framework of computational modeling to identify the patient admission scheduling problem in the healthcare system. *Journal of Healthcare Engineering* 2022: 1938719.
- Ilani MA and Khoshnevisan M (2021) Study of surfactant effects on intermolecular forces (IMF) in powder-mixed electrical discharge machining (EDM) of Ti–6Al–4V. *International Journal of Advanced Manufacturing Technology* 116(5–6): 1763–1782.
- Ilani MA and Khoshnevisan M (2022) An evaluation of the surface integrity and corrosion behavior of Ti–6Al–4V processed thermodynamically by PM-EDM criteria. *International Journal of Advanced Manufacturing Technology* 120(7–8): 5117–5129.
- Kulshrestha A, Krishnaswamy V and Sharma M (2020) Bayesian BILSTM approach for tourism demand forecasting. *Annals of Tourism Research* 83: 102925.
- Kumaresan K and Ganeshkumar P (2020) Software reliability prediction model with realistic assumption using time series (S)ARIMA model. *Journal of Ambient Intelligence and Humanized Computing* 11(11): 5561–5568.
- Law R, Li G, Fong DKC, et al. (2019) Tourism demand forecasting: A deep learning approach. *Annals of Tourism Research* 75: 410–423.
- Liu J, Chen Y, Zhan J, et al. (2018) An on-line energy management strategy based on trip condition prediction for commuter plug-in hybrid electric vehicles. *IEEE Transactions on Vehicular Technology* 67(5): 3767–3781.
- Mengesha BN, Shaeri MR and Sarabi S (2022) Artificial neural network to predict pressure drops in heat sinks. In: Proceedings of the 9th International Conference on Fluid Flow, Heat and Mass Transfer (FFHMT'22). Avestia Publishing. DOI: 10.11159/FFHMT22.202.
- Panigrahi S and Behera HS (2017) A hybrid ETS–ANN model for time series forecasting. *Engineering Applications of Artificial Intelligence* 66: 49–59.
- Pao HT (2009) Forecast of electricity consumption and economic growth in Taiwan by state space modeling. *Energy* 34(11): 1779–1791.
- Parmezan ARS, Souza VMA and Batista GEAPA (2019) Evaluation of statistical and machine learning models for time series prediction: Identifying the state-of-the-art and the best conditions for the use of each model. *Information Sciences* 484: 302–337.
- Rasp S and Lerch S (2018) Neural networks for postprocessing ensemble weather forecasts. *Monthly Weather Review* 146(11): 3885–3900.

- Shaeri MR, Randriambololona AM and Sarabi S (2022) *Prediction Accuracy of Artificial Neural Networks in Thermal Management Applications Subject to Neural Network Architectures*. Epub ahead of print 2022. DOI: 10.11159/htff22.175.
- Shaeri MR, Sarabi S, Randriambololona AM, et al. (2022) Machine learning-based optimization of air-cooled heat sinks. *Thermal Science and Engineering Progress* 34: 101398.
- Sharadga H, Hajimirza S and Balog RS (2020) Time series forecasting of solar power generation for large-scale photovoltaic plants. *Renewable Energy* 150: 797–807.
- Siami-Namini S, Tavakoli N and Siami Namin A (2018) A comparison of ARIMA and LSTM in forecasting time series. In: *Proceedings - 17th IEEE International Conference on Machine Learning and Applications, ICMLA 2018*. Institute of Electrical and Electronics Engineers Inc.: 1394–1401.
- Singla P, Duhan M and Saroha S (2022a) A dual decomposition with error correction strategy based improved hybrid deep learning model to forecast solar irradiance. *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects* 44(1): 1583–1607.
- Singla P, Duhan M and Saroha S (2022b) A hybrid solar irradiance forecasting using full wavelet packet decomposition and bi-directional long short-term memory (BiLSTM). *Arabian Journal for Science and Engineering* 47(11): 14185–14211.
- Singla P, Duhan M and Saroha S (2022c) Solar irradiation forecasting by long-short term memory using different training algorithms. In: *Renewable Energy Optimization, Planning and Control*. Singapore: Springer, 81–89.
- Singla P, Duhan M and Saroha S (2023a) A point and interval forecasting of solar irradiance using different decomposition based hybrid models. *Earth Science Informatics* 16(3): 2223–2240.
- Singla P, Duhan M and Saroha S (2023b) An integrated framework of robust local mean decomposition and bidirectional long short-term memory to forecast solar irradiance. *International Journal of Green Energy* 20(10): 1073–1085.
- View of Review of Different Error Metrics: A Case of Solar Forecasting (n.d.) Available at: <https://ajse.aiub.edu/index.php/ajse/article/view/212/116> (accessed 27 June 2024).
- Wu CH, Lu CC, Ma YF, et al. (2018) A new forecasting framework for bitcoin price with LSTM. In: *IEEE International Conference on Data Mining Workshops (ICDMW) 2018-November*. IEEE Computer Society: 168–175.
- Zhang Q, Wang H, Dong J, et al. (2017) Prediction of sea surface temperature using long short-term memory. *IEEE Geoscience and Remote Sensing Letters* 14(10): 1745–1749.