V International Scientific and Technical Conference Actual Issues of Power Supply Systems

Application of artificial intelligence models for detecting and predicting anomalies based on electricity consumption data

AIPCP25-CF-ICAIPSS2025-00299 | Article

PDF auto-generated using **ReView**
from
RIVER VALLEY
TECHNOLOGIES

Application of artificial intelligence models for detecting and predicting anomalies based on electricity consumption data

Yusuf Avazov[1,a], Umidjon Ruziev[1], Ibragim Karabaev[2], Kamola Abdullaeva[1], Golibjon Rashidov[1], Toyir Mamatqulov[2]

[1] *Tashkent state technical university named after Islam Karimov, Tashkent, Uzbekistan*
[2] *Termez State Engineering and Agrotechnologies University*

[a)] *Corresponding author: yusufbek_avazov@mail.ru*

Abstract. The article analyzes problems of detecting anomalies and predicting time series based on electricity consumption data from smart meters. The study uses the Isolation Forest, XGBoost, and LSTM algorithms, and their effectiveness in anomaly detection, energy consumption forecasting, and modeling dynamic relationships between complex variables is evaluated based on comparative analysis. The obtained results show that the proposed approaches allow for the timely detection of abnormal behavior in energy consumption, optimization of the power grid load, and ensuring the efficient use of resources. As a result, the foundation will be laid for the formation of integrated energy management systems, the stability of the power supply infrastructure will be strengthened, and energy efficiency will be significantly increased.

INTRODUCTION

In recent years, the rapid development of digital technologies has created the basis for fundamental changes in the management of energy networks, real-time monitoring of energy consumption, and increasing the efficiency of resource use. In particular, smart meters are of particular importance as a modern system capable of collecting high-precision and frequently updated data on electricity consumption. The sharp increase in the volume of data obtained from these devices increases the need for automated methods aimed at in-depth analysis of energy consumption patterns, identification of unusual behavior, and forecasting future trends [1-3].

Time series data from smart meters have a complex structure, incorporating seasonality, trend components, as well as the influence of external factors such as temperature, humidity, and wind speed. Proper modeling of these factors allows for a deeper understanding of the dynamic nature of energy consumption, reliable forecasting of future consumption, as well as predicting possible loads in the electrical grid and periods when demand may increase. At the same time, identifying anomalies associated with unusual consumption - excessively high or low consumption, technical failures, incorrect calculations, or the likelihood of fraud - plays an important role in ensuring the reliability and security of the system [4-17].

The relevance of this research lies in the analysis of data obtained from smart meters using modern artificial intelligence technologies, the identification of seasonality, relationships, and structural features in them, as well as the creation of effective models capable of finding unusual situations in energy consumption and reliably predicting future consumption. In this process, the study of comprehensive data (EDA analysis), the analysis of correlation relationships between variables, the formation of lag properties over time (feature engineering), and the comparison of the results of various models form the methodological basis of the research [18-26].

This study was conducted on data consisting of 5000 records based on the "Smart Meter Electricity Consumption Dataset" [27]. This collection includes electricity consumption indicators recorded at 30-minute time intervals, weather parameters (temperature, humidity, wind speed), as well as average values calculated based on historical consumption and anomaly labels. Such a broad and multifaceted data structure allows for comprehensive analysis by performing extended EDA analysis, visual analysis of anomalies, feature engineering, and comparing the effectiveness of various models (Isolation Forest, XGBoost, and LSTM).

ADVANCED EDA ANALYSIS

At the initial stage of the study, we conducted an extended exploratory analysis (EDA) on the Smart Meter Electricity Consumption Dataset. This dataset consists of 5,000 observations and 7 main columns, allowing us to understand the dynamics of consumption over time. Table 1. The table view of the electricity consumption data set of the Smart meter is presented.

Table 1. Table view of Smart Meter Electricity Consumption Dataset

#	Timestamp	Electricity_Consumed	Temperature	Humidity	Wind_Speed	Avg_Past_Consumption	Anomaly_Label
0	2024-01-01 00:00:00	0.457786	0.469524	0.396368	0.445441	0.692057	Normal
1	2024-01-01 00:30:00	0.351956	0.465545	0.451184	0.458729	0.539874	Normal
2	2024-01-01 01:00:00	0.482948	0.285415	0.408289	0.470360	0.614724	Normal
3	2024-01-01 01:30:00	0.628838	0.482095	0.512308	0.576241	0.757044	Normal
4	2024-01-01 02:00:00	0.335974	0.624741	0.672021	0.373004	0.673981	Normal
…	…	…	…	…	…	…	…
4995	2024-04-14 01:30:00	0.366839	0.701004	0.362397	0.509174	0.490516	Normal
4996	2024-04-14 02:00:00	0.493568	0.258212	0.677895	0.627889	0.535212	Normal
4997	2024-04-14 02:30:00	0.893818	0.431739	0.688926	0.508038	0.681099	Abnormal
4998	2024-04-14 03:00:00	0.509673	0.592927	0.366151	0.668218	0.710599	Normal
4999	2024-04-14 03:30:00	0.233656	0.612872	0.473481	0.268306	0.555286	Normal

In the time series graph (Electricity_Consumed - Timestamp), although electricity consumption in the period from January to April 2024 remained generally stable, significant fluctuations in values showed high dispersion. The daily or weekly seasonality curve is not observed on the graph, but high random fluctuations indicate that consumption may be significantly dependent on external factors (temperature, humidity, and wind speed) (Figure 1).

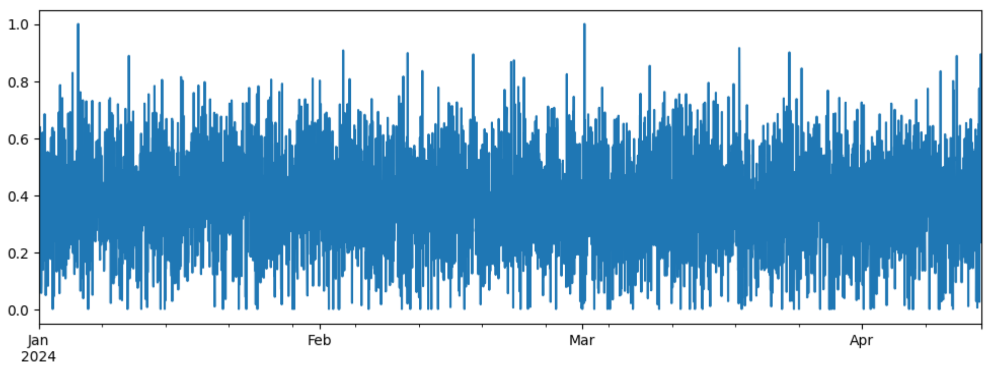


FIGURE 1. Graph of the time series of electricity consumption.

Therefore, a special approach is necessary when choosing a time-stable forecasting model, and the creation of additional characteristics (lag and seasonal characteristics) is important for increasing the stability of the models.

ABNORMALITY ANALYSIS

According to the results of this analysis, the number of normal cases was 4750 and anomalies - 250. Although anomalies constitute 5% of the total data, such a share corresponds to the natural frequency of unusual consumption cases in real energy systems and indicates the presence of the problem of class imbalance in model training. This discrepancy necessitates the use of special strategies such as oversampling, threshold adjustment, or weighted loss for models aimed at detecting anomalies.

The linear graph of time consumption showed that consumption indicators frequently change from the lowest values to sharply high values, and the noise percentage of the signal is high. This means that the seasonal component for this season is weak, and consumption changes mainly under the influence of external factors. These features indicate that simple regression models may have a high risk of overfitting, and therefore increase the likelihood of the dominance of in-depth learning models, such as LSTM.

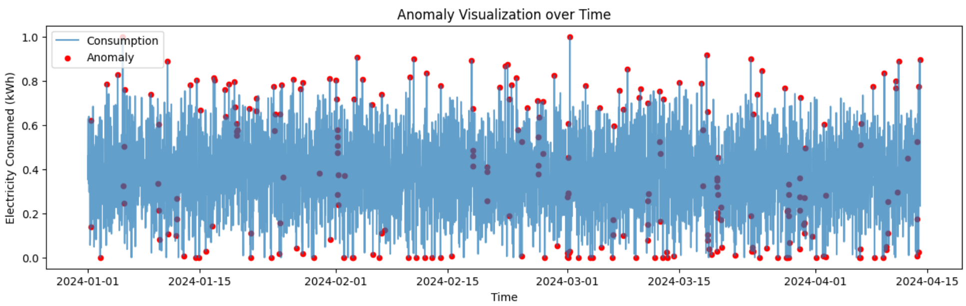


FIGURE 2. Time anomaly trend.

Below, using a graph of dispersion by temperature and consumption characteristics, we distinguish anomalous states (Fig. 3).

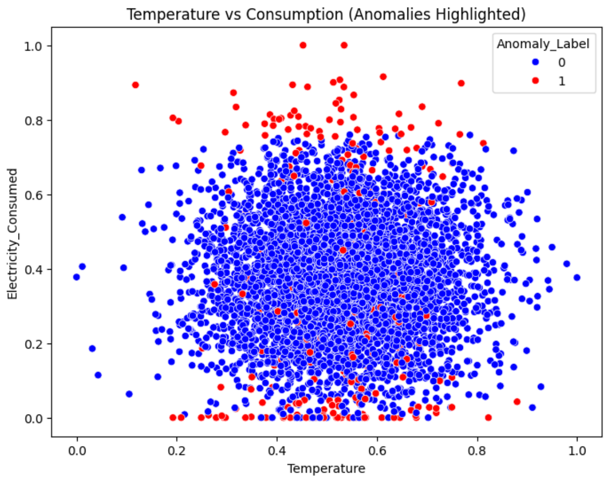


FIGURE 3. Visual classification of anomalies

Using pair plot analysis, we visually represent multidimensional relationships between energy consumption, weather factors, and historical average consumption indicators, as well as the spatial location of anomalies (Fig. 4).

As can be seen from the graphs, anomalies in most cases are located in the same distribution area as normal points and do not form a clearly linear or distinct cluster. This indicates the complexity of detecting anomalies using simple threshold or linear separation methods. Also, although a relatively more pronounced relationship is observed between Avg_Past_Consumption and Electricity_Consumed, there is no clear structural separation between weather factors and consumption. These results confirm the expediency of multidimensional and nonlinear approaches to detecting anomalies, in particular, such algorithms as Isolation Forest.

The results of the graphical analysis allow us to draw the following conclusions. Anomalies are mainly observed at very low or very high values of electricity consumption, indicating that abnormal consumption patterns typically manifest in extreme consumption ranges. At the same time, a direct and strong influence of the temperature factor on the occurrence of the anomaly was not revealed, since the anomalous points are scattered across different temperature ranges.

This means that anomalies are mainly caused not by one factor, but by a combination of several variables and complex multifactorial conditions. The results of this multidimensional analysis serve as a methodological basis for a deeper study of correlational relationships, assessment of seasonality, and formation of lag characteristics in time at subsequent stages.

CORRELATION ANALYSIS

We form the results of this analysis based on the correlation matrix (thermal map) (Fig. 5). This matrix shows that the relationship between Electricity_Consumed and Avg_Past_Consumption is 0.32. This indicator means that historical consumption is an important factor in the formation of future consumption, and there is an average significant direct correlation. This result once again confirms the importance of lag properties in time series modeling.

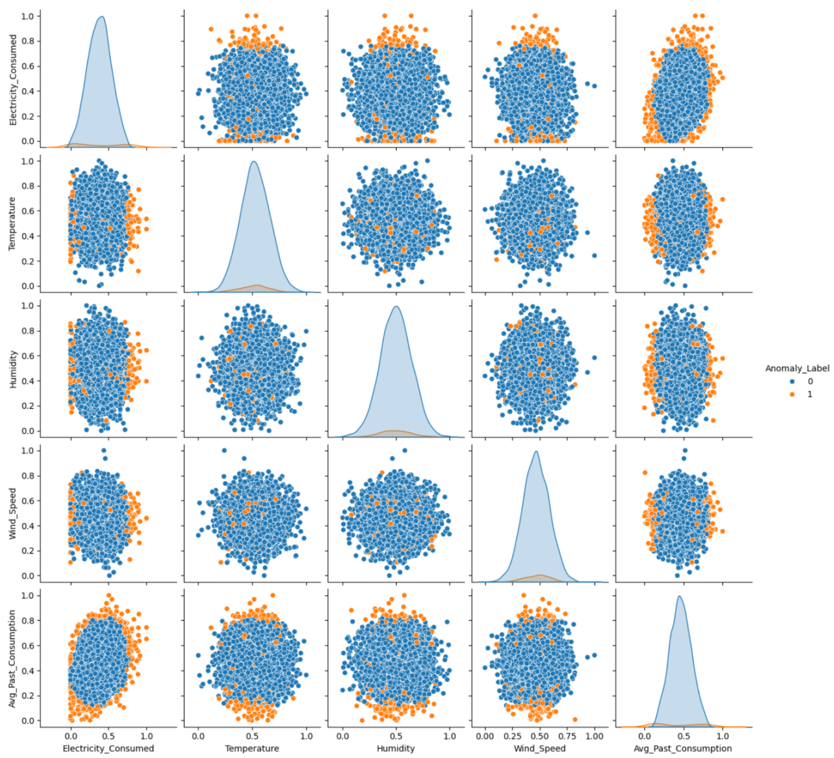


FIGURE 4. Multidimensional relationships between average consumption indicators and spatial distribution of anomalies.

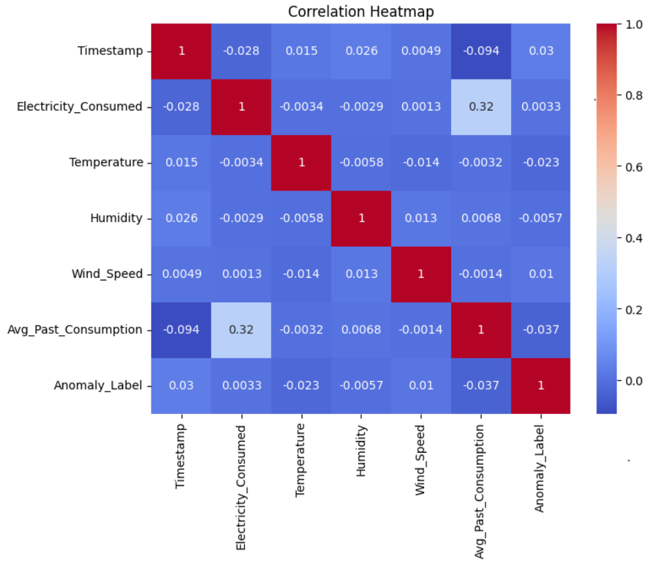


FIGURE 5. The relationship between properties is a correlation matrix.

The correlation of weather-related variables (Temperature, Humidity, Wind_Speed) with consumption is practically zero (≈0.00-0.03). This confirms that the main factor forming energy consumption in this dataset is not the external environment, but internal consumption indicators and time components.

The Anomaly_Label column also showed a very low correlation with all properties. This means that it is impossible to identify anomalies using simple linear relationships, and more nonlinear approaches are needed for their identification, in particular, Isolation Forest, deep learning models, or complex multidimensional analysis methods.

In general, correlation analysis shows that simple statistical methods are insufficient in the process of detecting anomalies, it is advisable to use advanced ML/DL models for the effective study of hidden structures in energy consumption graphs.

ANALYSIS OF SEASONALITY

Daily and weekly seasonal graphs revealed even more deeply the temporal structure of consumption.

Daily seasonality (Hourly Avg Consumption): Electricity consumption fluctuates several times throughout the day. In some hours (4-6 in the morning, in the evening) consumption is relatively high, which corresponds to the usual energy loads in real life (Fig. 6).

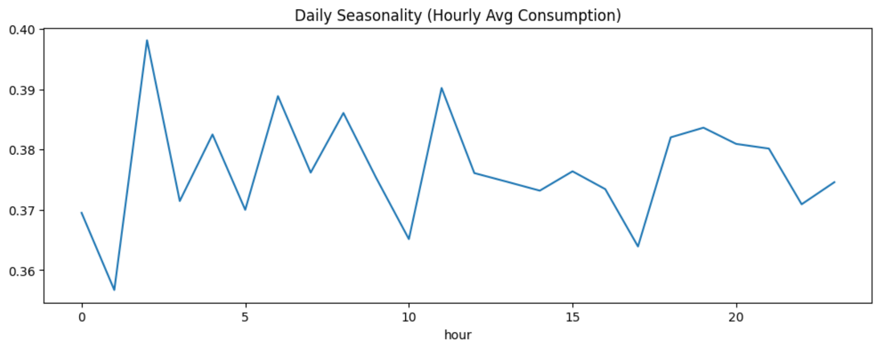


FIGURE 6. Daily seasonal electricity consumption.

Weekly seasonality (Day of Week Avg): On the 4th day of the week (Thursday), consumption had the highest average value. On the 2nd day (Tuesday), the lowest indicator was recorded. This shows that energy consumption has differences between working days and weekends (Fig. 7).

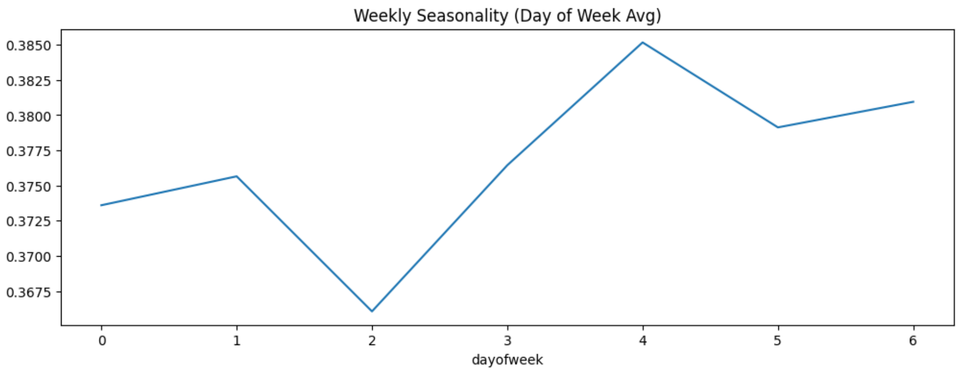
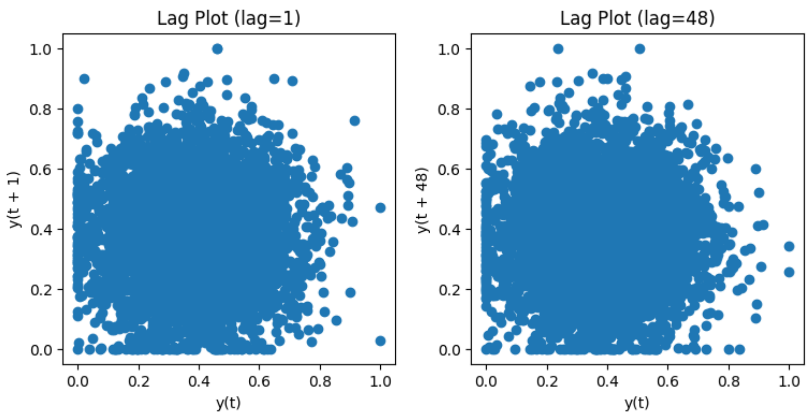


FIGURE 7. Weekly seasonal electricity consumption.

AUTOCORRELATION AND LAG-PLOT ANALYSIS

The lag plots compiled for lag-1 and lag-48 showed a weak internal dependence of the data on time. In both graphs, the points are concentrated around the center and do not form a definite diagonal or a specific figure (Fig. 8).



a) b)

FIGURE 8. Lag-plots compiled according to Lag-1 and Lag-48. a) The points are denser on the Lag 1 diagram, b) Points are not tightly packed on the Lag 48 diagram.

This means that there is no stable trend or regular cyclical structure in the time series, and consumption values are mainly characterized by random fluctuations. The absence of strong repetition in the Lag-48 period (24 hours) also confirms the insufficiency of accurate daily seasonality in the data. Among these common features, the Avg_Past_Consumption column stands out as the most important variable forming the time dependence.

Results of the autocorrelation function (Autocorrelation Function - ACF). In the ACF graph, no significant autocorrelation was observed, except for the 0-lag, and in long lags, the correlation approaches almost zero (Fig. 9).

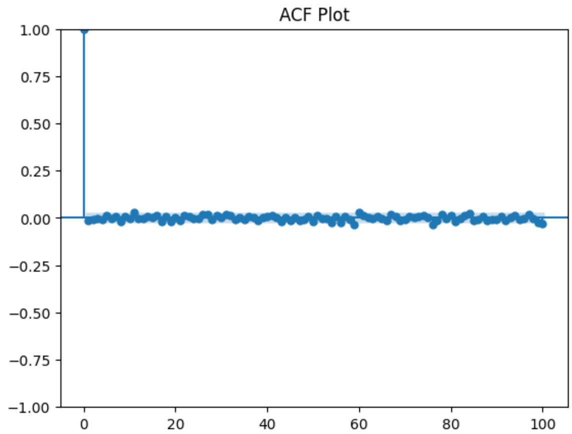


FIGURE 9. Graph of the autocorrelation function.

This means that the process of energy consumption has a strong stochastic nature, and there are no strictly repeating or periodic patterns. The obtained results show that models based on the study of long-term memory of time series (for example, LSTM) show that the cases that can be dominant in this dataset are limited, and the internal autocorrelation of the data is very low. However, models enriched with lag properties (such as XGBoost) can work more effectively with signals of stochastic nature.

FEATURE ENGINEERING

In order to increase the accuracy of the model results, a number of additional features related to the time series (feature engineering) were created:

Lag-1, Lag-48, Lag-96 - consumption values 30 minutes, 24 hours, and 48 hours ago, respectively;

Day and week characteristics (hour, dayofweek) - reflect the daily and weekly rhythms of consumption;

Moving average (Avg_Past_Consumption) - an indicator that mitigates short-term fluctuations and highlights the overall trend.

The combination of these features gave priority to the use of the following models:

- significantly increases the prediction accuracy for the XGBoost model, the model effectively absorbs lag signals;
- for the LSTM model, the time dependence is strengthened, and the ability to study sequential patterns is improved;
- in the anomaly detection process, additional features help stabilize the signal, making it easier to identify unusual points.

As a result, the feature engineering process creates a reliable foundation for subsequent modeling stages, increasing the efficiency of various models.

BUILDING MODELS OF AI INTELLIGENCE

Isolation Forest model. This is one of the fastest and most effective ensemble models specializing in detecting anomalies within multidimensional data. In this study, the model is trained with the following hyperparameters:

- *n_estimators=200* - more trees increased the accuracy due to the complexity of the data;
- *contamination=0.03* - close to the proportion of anomalies in the dataset (5%), but selected slightly lower for model stability;
- the default value of max_samples helped reduce overfitting in large datasets.

The model was trained based on consumption, temperature, humidity, wind speed, and historical consumption characteristics. When comparing the model results with actual values, the following indicators were obtained.

Table 2. Results of comparing the values obtained by the model with the actual values

Indicator	Normal	Anomaly
Precision	0.96	0.53
Recall	0.99	0.32
F1-score	0.97	0.40

The model detected normal states very well (Recall=0.99) and did not fully cover anomalies, showing the presence of difficulties in detecting anomalies (Recall=0.32). We can assume that this is due to the small number of anomalies in the dataset (5%). Also, due to the dispersion of anomalies, the model became complicated.

In the graph below, we highlight the points classified by the model as "normal" (0) and "anomalous" (1) in the space of consumer values (Fig. 10).

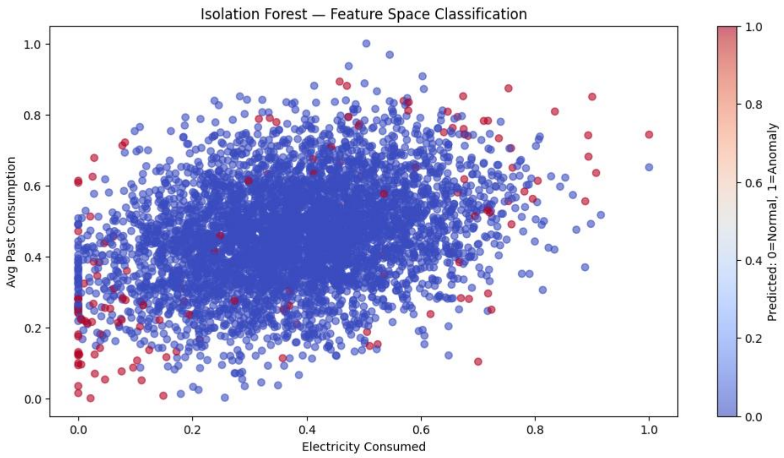


FIGURE 10. Feature Space Classification

The graph clearly shows what decisions the Isolation Forest model makes in the property space. Here, the blue dots represent normal consumption states, and the red dots represent observations classified by the model as anomalies.

The Confuzion Matrix graph shows how well the Isolation Forest model matches real labels. The matrix reflects the following results (Fig. 11).

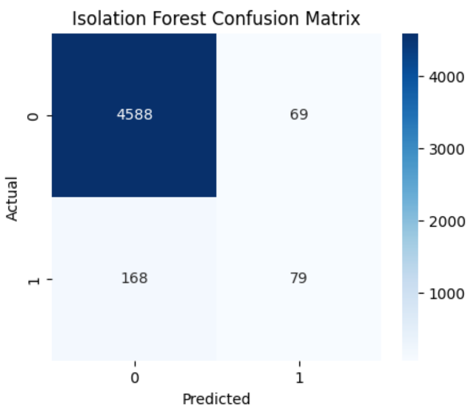


FIGURE 11. Indicators of classification effectiveness of the model.

The evaluation results show that the Isolation Forest distinguishes normal conditions very well, but cannot fully cover them due to the fact that the anomalies are more complex.

Nevertheless, considering that the total accuracy of the Isolation Forest model is 95%, we can consider it an effective approach that gives an initial signal about unusual situations in the system.

XGBOOST REGRESSION MODEL

The purpose of applying this model is to forecast electricity consumption for the next 30-minute interval (one-step ahead forecasting). As input characteristics to the model, we choose external weather parameters (Temperature, Humidity, Wind_Speed), historical average consumption (Avg_Past_Consumption), and lag-characteristics (lag1, lag48, lag96). We divide the data into train-set [:500] and test-set [500:] by time.

We configure the XGBoost with the following hyperparameters: n_estimators=300; learning_rate=0.05; max_depth=6; subsample=0.8; colsample_bytree=0.8.

We trained the model based on historical data, obtained predictive values based on a set of tests, and calculated the following evaluation metrics:

RMSE = 0.15838 and MAE = 0.12810

The graphs constructed for the XGBoost model accurately describe how the forecast results are formed over time. In the graph viewed over the entire period of the test set, the forecast line of the model is located very close to the actual consumption values, which indicates that XGBoost was able to successfully reproduce the overall dynamics (Fig. 12).

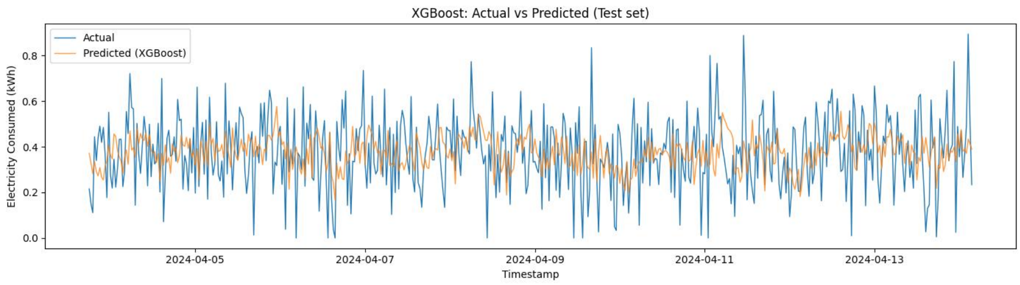


FIGURE 12. Forecast results obtained using the XGBoost model

Although the model cannot fully reflect sharp signal fluctuations, it maintains the general trend of energy consumption in a smooth and stable form. This situation indicates the advantage of XGBoost as a noise-filtering regression model in time series with high noise levels, such as electricity consumption.

The zoom, i.e., analysis of the first 200 points of the test set, more clearly demonstrated the model's reaction to high-frequency oscillations (Fig. 13).

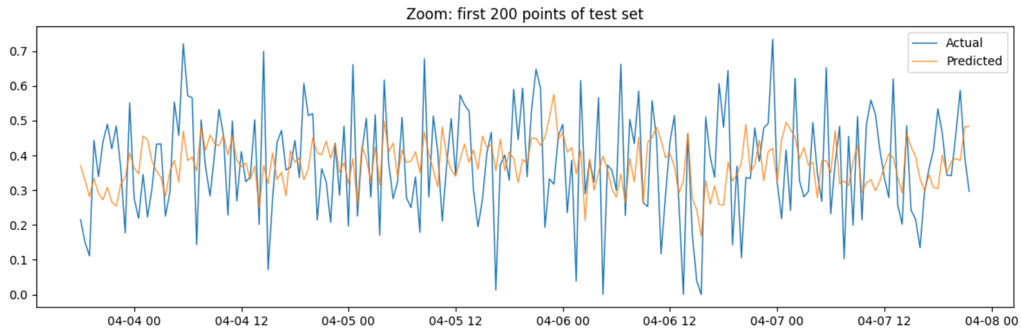


FIGURE 13. The result of the analysis of the first 200 points of the test set.

The graphs show that XGBoost operates with a small error in the low-to-medium consumption range, accurately reproducing the main trend. However, during high-amplitude sharp changes (peaks), the model's prediction reacted with a slight delay. This “lag effect” is a common phenomenon in the process of predicting time series with gradient-based models, especially in energy systems where sharp oscillations have a stochastic nature.

It has been shown that the XGBoost algorithm can better study lag signals and demonstrates high stability in short-term (one step ahead) forecasting.

PREDICTING THE TIME SERIES USING THE LSTM MODEL

Within the framework of the study, we use the LSTM model only for the Electricity_Consumed column. We first normalized the data in the interval [0.1] using MinMaxScaler; then converted it into a controlled learning format (lag=48) over 48 30-minute intervals (24 hours) and divided the data into train/test sets with an 80/20 ratio.

We built the proposed model of in-depth learning based on the two-layer LSTM architecture. Using 128 neurons in the first LSTM layer, we used the return_sequences=True parameter to fully preserve the time relationships in the sequence. In the second LSTM layer, 64 neurons were located. Applying the Dropout (0.3) mechanism after both LSTM layers, we reduced the model's tendency to overfitting.

Table 3. Results of reducing the model's tendency to overfitting

Layer type	Output shape	Number of parameters
LSTM (128 units)	(None, 48, 128)	66 560
Dropout (0.3)	(None, 48, 128)	0
LSTM (64 units)	(None, 64)	49 408
Dropout (0.3)	(None, 64)	0
Dense (ReLU, 16 units)	(None, 16)	1 040
Dense (Output, 1 unit)	(None, 1)	17
Total	—	**117 025**

To enhance the model's ability to study high-level zero-linear relationships, we add a Dense (16) intermediate layer with ReLU activation after the LSTM layers and finally use the Dense (1) output layer. The model was trained for 30 epochs with the Adam optimizer and the elimination function of the root mean square error (MSE).

In the training process, the value of the training loss function remained stable in the range of ≈0.026-0.028, and the validation loss function - in the range of ≈0.028-0.029. The small difference between training and validation losses indicates that no serious overfitting cases were observed in the model. At the same time, the fact that the loss values do not fall below a certain level is explained by the high noise and weak periodicity of the energy consumption signal.

According to the test results, the RMSE value of the LSTM model was equal to ≈ 0.162. This result showed that the model could not adequately study deep time dependencies due to the absence of strong seasonality or autocorrelation in the internal structure of the time series (Figures 14-15).

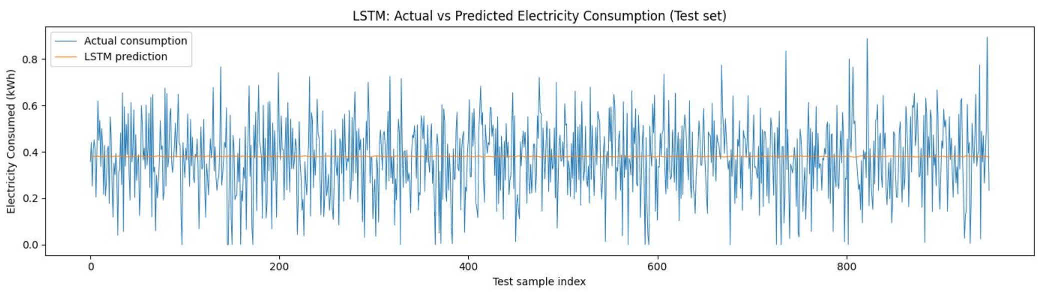


FIGURE 14. Prediction graph obtained using the LSTM model

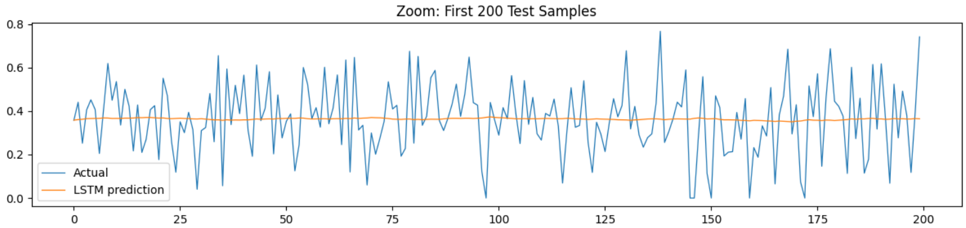


FIGURE 15. The first 200 points of the test set.

As a result, the LSTM model generated predictions in the test set, mainly around average values. This phenomenon is described in the scientific literature as "mode collapse" or "mean predictor behavior" and is frequently encountered in time-series signals that are highly noisy and do not have a clear structural diagram [28-29].

Due to the fact that the model was trained on the basis of only one feature - electricity consumption, as well as insufficient expression of stable and recurring seasonal patterns in the signal structure, the LSTM model could not effectively reproduce sharp oscillations and local peaks [30-33]. At the same time, the RMSE of the LSTM model was slightly higher than that of the XGBoost model, which showed that under the conditions of this dataset, the classical gradient boosting approach gave a relatively more stable result in time series predictions.

To improve the obtained results, first of all, it is advisable to provide multidimensional input data to the LSTM model. In particular, the introduction of additional characteristics, such as temperature, humidity, wind speed, and historical average consumption, without limiting the consumption column, can help the model to study complex relationships more deeply.

Also, optimizing the length of the lag window, deepening the model architecture (with drop-out and regularization mechanisms), and using signal smoothing or denoising methods serve to reduce the “mean predictor” effect.

This situation is explained by the weak expression of long-term and specific seasonal structures in energy consumption data, as well as the predominant stochastic nature of the signal. Therefore, although the LSTM model demonstrates limited efficiency for the dataset within this study, it is more suitable for use in large energy systems with strong seasonality and complex time dependencies.

COMPARATIVE ANALYSIS OF THE RESEARCH RESULTS

The indicators of the comparative analysis of the results obtained on the basis of three different models constructed within the framework of the study are summarized in Table 4.

The obtained results made it possible to comprehensively assess the capabilities of the models used in this study in the analysis and forecasting of energy consumption.

The **Isolation Forest model** was able to distinguish normal consumption cases with high accuracy, however, due to the small number of anomalies and their random distribution in the data space, the F1-score value was recorded relatively low. Nevertheless, the model's light architecture and fast operation make it an optimal solution for use as a preliminary warning mechanism in real-time monitoring systems.

Table 4. Indicators of comparative analysis of the research results

Model Type	Function of the model	Evaluation Criteria	Result
Isolation Forest	Detection of abnormality	F1-score	0.40
XGBoost	Prediction	RMSE	0.1584
LSTM	Deep learning time series prediction	RMSE	0.1622

The **XGBoost model** showed the lowest RMSE value when enriched with time lag properties. This model was distinguished by high stability in short-term energy consumption forecasting, quick learning, and ease of adaptation of hyperparameters. The results confirm that XGBoost is the most effective and reliable approach for practical forecasting tasks under these dataset conditions.

The **LSTM model**, on the other hand, showed forecasts that were mainly smoothed around average values in this dataset. This is explained by the fact that the energy consumption signal has high noise and low structural patterns, and the model is trained based on only one feature - the value of consumption. As a result, the LSTM could not fully recover the sharp oscillations, and its RMSE was slightly higher than the XGBoost model.

At the same time, the LSTM model can fully demonstrate its advantages when applied to time series with pronounced long-term time dependencies and seasonal structures, low noise, and multiple characteristics. In this dataset, the use of XGBoost for short-term forecasting tasks, Isolation Forest for anomaly detection, and LSTM in cases of complex and deep time patterns is recommended as the most optimal approach.

CONCLUSION

In general, processing data from smart meters based on ML and DL methods significantly expands the possibilities of early detection of abnormal energy consumption situations, effective management of the electrical network load, rational use of resources, and optimization of energy management systems. The combined application of the models tested in this study can serve to further improve the stability, security, and digital management systems of the energy infrastructure.

REFERENCES

1. Himeur, Y., Alsalemi, A., Bensaali, F., & Amira, A. (2020). A novel approach for detecting anomalous energy consumption based on micro-moments and deep neural networks. *Cognitive Computation*, 12, 1381–1401. https://doi.org/10.1007/s12559-020-09764-y
2. Himeur, Y., Ghanem, K., Alsalemi, A., Bensaali, F., & Amira, A. (2021). Artificial intelligence based anomaly detection of energy consumption in buildings: A review, current trends and new perspectives. *Applied Energy*. https://doi.org/10.1016/j.apenergy.2021.116601
3. Chiosa, R., Piscitelli, M. S., & Capozzoli, A. (2021). A data analytics-based energy information system (EIS) tool to perform meter-level anomaly detection and diagnosis in buildings. *Energies*, 14(1), 237. https://doi.org/10.3390/en14010237
4. Pan, H., Yin, Z., & Jiang, X. (2022). High-dimensional energy consumption anomaly detection: A deep learning-based method for detecting anomalies. *Energies*, 15(17), 6139. https://doi.org/10.3390/en15176139
5. Gulyamov, Shukhrat, Avazov, Yusuf, Samadov, Elyor, Nasiba, Khojiyeva, Qabilbek, Bazarbaev. Some issues of increasing energy efficiency of advanced process and production control systems // AIP Conference ProceedingsOpen source preview, 2025, 3331(1), 060009. DOI: 10.1063/5.0305712.
6. Lee, S., Jin, H., Nengroo, S. H., Doh, Y., Lee, C., Heo, T., & Har, D. (2023). Anomaly detection of smart metering system for power management with battery storage system/electric vehicle. *ETRI Journal*. https://doi.org/10.4218/etrij.2022-0135
7. Shrestha, R., Mohammadi, M., Sinaei, S., Salcines, A., Pampliega, D., Clemente, R., Sanz, A. L., Nowroozi, E., & Lindgren, A. (2024). Anomaly detection based on LSTM and autoencoders using federated learning in smart electric grid. *Journal of Parallel and Distributed Computing*, 193, 104951. https://doi.org/10.1016/j.jpdc.2024.104951
8. Akagic, A., & Džafić, I. (2024). Enhancing smart grid resilience with deep learning anomaly detection prior to state estimation. *Engineering Applications of Artificial Intelligence*, 127, 107368. https://doi.org/10.1016/j.engappai.2023.107368
9. Noorchenarboo, M., & Grolinger, K. (2024). Explaining deep learning-based anomaly detection in energy consumption data by focusing on contextually relevant data. *Energy and Buildings*, 115177. https://doi.org/10.1016/j.enbuild.2024.115177
10. Chen, Y., Li, J., Huang, Q., Li, K., Zhao, Z., & Ren, X. (2023). Non-technical losses detection with Gramian angular field and deep residual network. *Energy Reports*. https://doi.org/10.1016/j.egyr.2023.05.183
11. Kabir, B., Qasim, U., Javaid, N., Aldegheishem, A., Alrajeh, N., & Mohammed, E. A. (2022). Detecting nontechnical losses in smart meters using a MLP-GRU deep model and augmenting data via theft attacks. *Sustainability*, 14(22), 15001. https://doi.org/10.3390/su142215001
12. (MDPI Sensors) Detection of anomalies in daily activities using data from smart meters. *Sensors*, 24(2), 515. https://doi.org/10.3390/s24020515

13. Śmiałkowski, T., et al. (2025). Anomaly detection in urban lighting systems using autoencoder and transformer algorithms ARIMA + RNN). *Scientific Reports*.
14. Massarani, A. H., Badr, M. M., Baza, M., Alshahrani, H., & Alshehri, A. (2025). Efficient and accurate zero-day electricity theft detection from smart meter sensor data using prototype and ensemble learning. *Sensors*, 25(13), 4111. https://doi.org/10.3390/s25134111
15. Lee, S., Jin, H., Nengroo, S. H., Doh, Y., Lee, C., Heo, T., & Har, D. (2022). Smart metering system capable of anomaly detection by bi-directional LSTM autoencoder. *IEEE International Conference on Consumer Electronics (ICCE)*.
16. Lim, J. Y., et al. (2022). Anomalous energy consumption detection using a Naïve Bayes approach (energy consumption data). *F1000Research*, 11, 64.
17. Aghazadeh Ardebili, A., et al. (2024). Enhancing resilience in complex energy systems through real-time anomaly detection (edge + security aspects). *Energy Informatics*.
18. Pamir, Javaid, N., Akbar, M., Aldegheishem, A., Alrajeh, N., & Mohammed, E. A. (2022). Employing a machine learning boosting classifiers based stacking ensemble model for detecting non-technical losses in smart grids. *IEEE Access*, 10, 121886–121899. https://doi.org/10.1109/ACCESS.2022.3222883
19. Ullah, A., Javaid, N., Javed, M. U., Pamir, Bahaj, S. A., et al. (2022). Adaptive data balancing method using stacking ensemble model and its application to non-technical loss detection in smart grids. *IEEE Access*, 10, 133244–133255. https://doi.org/10.1109/ACCESS.2022.3230952
20. Asif, M., Nazeer, O., Javaid, N., Alkhammash, E. H., & Hadjouni, M. (2022). Data augmentation using BiWGAN, feature extraction and classification by hybrid 2DCNN and BiLSTM to detect non-technical losses in smart grids. *IEEE Access*, 10, 27467–27483. https://doi.org/10.1109/ACCESS.2022.3150047
21. Ullah, A., Javaid, N., Asif, M., Javed, M. U., & Yahaya, A. S. (2022). AlexNet, AdaBoost and artificial bee colony based hybrid model for electricity theft detection in smart grids. *IEEE Access*, 10, 18681–18694. https://doi.org/10.1109/ACCESS.2022.3150016
22. Shehzad, F., Javaid, N., Almogren, A., Ahmed, A., Gulfam, S. M., & Radwan, A. (2021). A robust hybrid deep learning model for detection of non-technical losses to secure smart grids. *IEEE Access*, 9, 128663–128678. https://doi.org/10.1109/ACCESS.2021.3113592
23. Javaid, N. (2021). A PLSTM, AlexNet and ESNN based ensemble learning model for detecting electricity theft in smart grids. *IEEE Access*, 9, 162935–162950. https://doi.org/10.1109/ACCESS.2021.3134754
24. Pamir, Javaid, N., Qasim, U., Yahaya, A. S., Alkhammash, E. H., & Hadjouni, M. (2022). Non-technical losses detection using autoencoder and bidirectional gated recurrent unit to secure smart grids. *IEEE Access*, 10, 56863–56875. https://doi.org/10.1109/ACCESS.2022.3171229
25. Peng, Y., et al. (2021). Electricity theft detection in AMI based on clustering and local outlier factor. *IEEE Access*, 9, 107250–107259.
26. Firoozi, H., et al. (2022). Non-technical loss detection in limited-data low-voltage distribution networks (smart meter based). *International Journal of Electrical Power & Energy Systems*.
27. https://www.kaggle.com/datasets/ziya07/smart-meter-electricity-consumption-dataset/code
28. Goodfellow, I., Bengio, Y., Courville, A. Deep Learning. MIT Press, 2016, pp. 321–323.;
29. Lai, G., Chang, W. C., Yang, Y., Liu, H. Modeling Long- and Short-Term Temporal Patterns with Deep Neural Networks. *Proceedings of the 41st International ACM SIGIR Conference*, 2018, pp. 95–104.
30. Khan, N., et al. (2024). A novel deep learning technique to detect electricity theft in smart grids using AlexNet. *IET Renewable Power Generation*.
31. Saqib, S. M., et al. (2024). Deep learning-based electricity theft prediction in non-smart grid environments (relates to smart meter theft literature). *Heliyon*. https://doi.org/10.1016/j.heliyon.2024.e35167
32. Chen, J., et al. (2025). Improved DBSCAN-based electricity theft detection using smart meter/AMI features. *Applied Sciences*.
33. Elshennawy, N. M., et al. (2025). An efficient electricity theft detection based on deep learning (smart meter based context). *Scientific Reports*.