Scientific Journal of Astana IT University
ISSN (P): 2707-9031
ISSN (E): 2707-904X

DOI: 10.37943/21AJSD8750

Akylbek Tokhmetov
Candidate of Physical and Mathematical Sciences, Associate Professor
Department of Information Systems
attohmetov@mail.ru, orcid.org/0000-0003-0764-8574
L.N. Gumilyov Eurasian National University, Kazakhstan
Kenzhegali Nurgaliyev
PhD candidate, Department of Information Systems
ken.nurgaliev@yandex.kz, orcid.org/0000-0001-9391-0924
L.N. Gumilyov Eurasian National University, Kazakhstan
Liliya Tanchenko
Master's degree, Postdoctoral, Department of Information Systems
ltanchenko@mail.ru, orcid.org/0000-0002-6811-2303
L.N. Gumilyov Eurasian National University, Kazakhstan

COMPARATIVE ANALYSIS OF VARIOUS FORECAST MODELS OF ELECTRICITY CONSUMPTION IN SMART BUILDINGS

Abstract: The rapidly growing field of smart building technology depends heavily on accurate electricity consumption forecasting. By anticipating energy demands, building managers can optimize resource allocation, minimize waste, and enhance overall efficiency. This study provides a comprehensive comparative analysis of various models used to forecast electricity consumption in smart buildings, highlighting their strengths, limitations, and suitability for different use cases. The investigation focuses on three major categories of forecasting models: statistical methods, machine learning techniques, and hybrid approaches. Statistical models, such as the Moving Average Method, leverage historical data patterns to predict future trends. These models enable analysts to utilize predictive analytics, simulating real-world environments and helping them make more informed decisions. The study offers a detailed comparison of several predictive models applied to Internet of Things (IoT) data, with a particular emphasis on energy consumption in smart buildings. Among the short-term forecasting models examined are gradient-enhanced regressors (XGBoost), random forest (RF), and long short-term memory networks (LSTM). The performance of these models was evaluated based on prediction errors to identify the most accurate one. Time series, machine learning, and hybrid models used to predict energy consumption are considered and analyzed. The focus is on the accuracy of forecasts and their applicability in real-world conditions, taking into account factors such as climate change and data obtained from Internet of Things (IoT) sensors. The analysis shows that hybrid models combining machine learning and time series provide the best prediction accuracy over different time horizons. It also highlights the importance of integrating user behavior data and using IoT technologies to improve model accuracy. The results can be applied to create energy-efficient control systems in smart buildings and optimize energy consumption.

Keywords: smart building; energy consumption forecasting; optimization algorithms; machine learning; computational processes.

Introduction

In recent decades, there has been a significant increase in interest in energy efficiency and sustainability, driven by rising energy consumption and the need for optimization. With urbanization and the rise of smart buildings, the importance of accurately predicting energy consumption has increased.

Predicting energy consumption allows for advanced resource planning and management, optimizing their use and minimizing costs. This is especially important given the ever-growing volumes of data and the increasing complexity of energy resource management. Optimizing energy consumption effectively saves energy, as shown by numerous scientific studies [1], [2].

ISSN (P): 2707-9031
ISSN (E): 2707-904X
Researchers use many methods simultaneously to predict how much electricity people will use. These include mathematical models of fuzzy logic [3], deep and classical machine learning methods [4], [5], and models that take into account how electricity consumption varies depending on the season and the influence of various characteristics [6], [7].

Despite the wide variety of electricity consumption forecasting methods, there is no common approach that allows the presentation of a reliable electricity consumption forecast for each subject area. The main reasons for this include evolving forecasting accuracy standards, the necessity to consider numerous factors that define the subject area's specifics, and advancements in data mining technologies, which enable more efficient handling of vast data sets compared to conventional mathematical statistics methods. Therefore, it was decided to conduct an applied study that involves forecasting electricity consumption using modern data regression tools. Thus, the purpose of this work is to develop an optimal method for forecasting electricity consumption and assess the accuracy of these methods.

Following the introduction, the various types of data analytics techniques and assessment metrics are described. The following section presents the results of an exploratory data analysis experiment for energy consumption model selection.

This work presents a comparative analysis of various forecasting models used for predicting energy consumption in smart buildings. The focus is on machine learning methods, such as neural networks, regression models, and time series, as well as statistical models. The goal of the study is to identify the most accurate and efficient models that can be utilized for developing energy-efficient management systems in smart buildings.

We shall categorize the forecasts employed in this study according to the duration of anticipation before moving on to specific methods. Forecasts are categorized differently depending on how long prediction needed [8]. We shall follow the following guidelines in the context of this investigation. The projections that we will take into consideration are classified as long-term, which spans several years, medium-term, which spans months to years, short-term, which spans days to weeks, and operational, which spans hours and minutes in a single day (Figure 1).

Figure 1. Power consumption prediction categories

By separating the methodologies based on the expected lead time, we were able to look more closely at the approaches that work well for corresponding lead time forecasting. It is believed that this division of approaches makes it easier for us to explore the related literature and helps us to tag the characteristics of the application conditions for every forecasting period.

Short-term forecasting.

Operational forecasts are essential for managing electricity consumption with its peak loads which can be decreased by having a forecast that is accurate one day or several weeks in advance. This type of forecasting is relevant for planning power demand supported by an analysis of small power grids [9].

In [10], the authors present an approach that allows optimal management of the energy consumption of a building (hotel). The method made it possible to reduce operating costs for electricity consumption by more than 10%. The researchers explain that creating such method allows for the nonlinearity of power consumption data for some types of equipment to be considered.

The author [11] presents the results of using machine learning to predict cooling energy consumption in office buildings considering the human behavior. When developing the model, several machine learning algorithms were tested and compared. The simulation results

ISSN (P): 2707-9031
ISSN (E): 2707-904X

demonstrated the great influence of the variables considered in the study on the target result – electricity consumption. The activities of office building workers affected energy consumption by more than 7 times.

In [12], researchers argue that an accurate understanding of energy load curves is the key to efficient management of plant power systems and the basis for anomaly detection. Scientists also note that load curve analysis is an important addition in the absence of methods for assessing various time transitions between energy states.

The paper [13] presents the results of forecasting sales (generation) of electricity based on a deep spatio-temporal residual neural network (ST-ResNet). The use of ST-ResNet is a way to reduce the average absolute percentage forecast error by more than 2.5% for a short-term (1 day) and a medium-term (1 week) forecasting.

On the one hand, the development and application of data mining methods, in particular for the task of predicting electrical loads, helps to reduce and rationalize the use of resources. However, performing machine learning procedures requires significant computing resources and, consequently, an increase in electrical energy consumption.

Long-term forecasting

Continued predictions are used to plan building or repairs of key production and infrastructure facilities, as well as to elaborate strategies for the growth of energy systems at the state level and independently within a certain industrial field. A variety of long-term scenarios are typically employed to get the projection numbers for total energy consumption.

Forecasting allows for optimal management of the operating modes of energy storage devices, which contributes to its more rational use. The study [14] presents the results of forecasting electricity consumption using a decision tree model. The obtained results are used in the process of determining the optimal capacity of an electricity storage device.

The results of a comparative analysis of machine learning methods and traditional methods of forecasting electricity consumption, carried out in the article [15], confirm the significant superiority of machine learning methods in terms of forecast accuracy, which indicates the relevance of developing predictive models based on neural networks and classical machine learning algorithms. However, as noted in [16], the main disadvantage of using these methods is the computational complexity, which makes the task of increasing the efficiency of data mining algorithms especially relevant.

In [17], a model was developed that combines the use of singular spectrum analysis for partitioning the time series of electricity consumption and a fully connected neural network. Such application is a promising direction, since it uses a combination of several methods of intellectual analysis. However, even with the most modern approaches to forecasting electricity consumption, scientists note the limitations in the use of some methods [18]. This only confirms the need to conduct research to modernize and expand the methodological arsenal when solving the problem of forecasting electricity consumption.

Thus, it is advisable to conduct systematic reviews, such as in [19], reflecting the current state of research in the field of forecasting electricity consumption. Given the undying interest of the scientific community in energy saving issues, progress in the means and methods of intelligent data processing and the high practical significance of the development of energy-saving technologies, it is necessary to conduct hybrid forecasting to develop recommendations on the possibility of using crucial building facilities.

Hybrid forecasting

Nowadays, achieving maximum energy efficiency is anticipated to be the main emphasis of IoT innovation in the future. This issue may be met by incorporating AI-powered technologies such as machine learning (ML) and deep learning [20]. AI research fields are advanced as a result of ML systems' constant self-improvement. ML employs algorithms that enable them to react to environmental inputs and recognize nonlinear relationships in complex or uncertain systems.

Predicting how much energy a CPS will use in a various time intervals is a useful approach to optimize energy use, forecast future energy requirements, and spot possible inefficiencies in energy use, among other agents in a smart grid. The performance of ML algorithms is significantly influenced by the relevance and quality of the data they utilize. Electricity usage was divided into several categories in [21]: raw measurements and records of private loads (air conditioners, refrigerators) at certain periods as it enables precise energy management, load disaggregation, and appliance identification.

The main drawbacks of hybrid prediction techniques have been addressed by ML algorithms. For example, authors in [22] created a building energy consumption forecasting model, by combining an enhanced sine cosine optimization algorithm with extended short-term memory networks to perform in real-time prediction. Subsequently, [23] concentrated on forecasting dining room energy usage. Authors employed an LSTM model and its features were extracted using principal component analysis (PCA).

Furthermore, authors in [24] used cloud-based machine learning framework (Microsoft Azure) to offer following methodology. Using three strategies (k-nearest neighbors, artificial neural networks, and support vector machines), the study employs two tenants from an industrial building in Malaysia. According to the experimental results, each renter's energy consumption follows a certain distribution pattern, and the proposed model is able to calculate each renter's energy consumption with accuracy.

Hybrid network is a term used to describe some interconnected systems with different forecasting issues. In [25], the main attention will be paid to the development of forecasting models based on two methods of machine learning (ML). The first is the branch of RNN method, which is named as Long-Short Term Memory (LSTM). The second one is Support Vector Machine. Both methods solve the problems related with prediction and show high accuracy and precision in time-series application. In [26], the Forecaster Autoreg and Neural Network time series forecasting models for forecasting electricity consumption in Astana city are presented. The study covers data for 2020, including temperature and wind speed. The authors highlight the importance of accurate forecasting to improve energy efficiency and reduce environmental impact. The accuracy and reliability of forecasts were evaluated by error indicators such as Average Absolute Error (MAE) and Average Absolute Percentage Error (MAPE), and the results showed that the models can provide accurate forecasts with low errors.

In [27], authors concentrate on the use of an intelligent control algorithm in HVAC (Heating, Ventilation, & Air Conditioning) systems to improve thermal comfort and energy efficiency. The authors suggest optimizing heat transfer coefficients and air temperature values by combining SCADA (Supervisory Control and Data Acquisition) systems with an intelligent building management system. Then, genetic algorithms are utilized to minimize energy usage and preserve user comfort. Regarding power usage, authors in [28] designed a hybrid network that combines long short-term memory (LSTM) with improved complete ensemble empirical mode decomposition with adaptive noise (iCEEMDAN). Using it, researchers separated the original power consumption data into patterns, and then they forecasted each mode separately using a Bayesian-optimized LSTM.

As follows from the analysis of the works, to improve the accuracy of forecasts, it is recommended to use hybrid models that combine machine learning and time series methods. In addition, it is noted that it is important to take into account exogenous factors, such as climate and social changes, which can reduce the error of forecasts.

Methods

This section describes the research methods used to predict energy consumption in cyber-physical systems. Forecasting energy consumption in CPS is a complex task that requires the use of various techniques and approaches to ensure the accuracy and reliability of predictions. A multitude of sources, including sensor data streams, environmental data, and other entities, are used as input for hybrid predicion method. Cutting edge processing tools and algorithms are needed to process this data accurately and quickly. Several categories of data analytics exist according to the needs of specific applications.

Time series of data is used for forecasting purposes and tasks have a number of features. Since a time series is a sequence of values in which each subsequent value contains the past for

ISSN (P): 2707-9031
ISSN (E): 2707-904X

the subsequent ones, any attempt to forecast the future without studying the time series of the past is unscientific and erroneous. Therefore, to obtain sufficiently accurate and reliable forecasts, it is necessary to study in detail the current state of this process. For example, to decompose the series into its constituent components and eliminate the influence of systematic components on the change of random ones, check the series for the presence of a main trend and, if there is one, to isolate it, identify the trend and its direction.

Regarding short-term forecasting following methods might be used:

- *Least Squares Method (LS)*

The LS method is widely used due to its simplicity and ability to fit a trend line to data by minimizing the sum of squared differences between actual and predicted values [29]. It is useful for short-term forecasting where trends are linear and stable.

The disadvantage of this method is that the trend model is rigidly fixed, which makes it possible to use it only for short lead times, i.e., for short-term forecasting. It is also sensitive to outliers, which can distort the accuracy of the forecast since squaring increases the impact of larger deviations.

The method describes the nature of changes in the level of the time series (slope, shift).

$$slope = \frac{(E_a \times t) - \frac{(E_a \times t)}{n}}{t^2 - \frac{t^2}{n}} \quad (1) \qquad\qquad shift = \frac{E_a}{n} - \frac{\text{slope } \times t}{n} \quad (2)$$

where E_a – actual consumption value, n – number of time series, t – time of observation period.

This affects the minimization of the sum of squares of the deviations of the readings:

$$Pred.Consumption = slope \times t + shift \quad (3)$$

- *Moving Averages Method*

The moving averages method calculates the average level of a series by taking the average of a certain number of initial levels, then continuing to average the same number of levels starting from the next level, and so on. This creates the impression that the average is "sliding" along the series, discarding one level and adding the next each time [30]. This method is simple and useful for smoothing out short-term fluctuations and highlighting longer-term trends in data. It can provide a quick overview of consumption patterns without requiring complex computations.

However, it cannot capture nonlinear dependencies or sudden changes in trends, which makes it less effective for complex datasets or in scenarios where significant changes in data values occur. It is also limited in its application for long-term forecasting.

Fluctuations in averaged values are replaced by the arithmetic mean in the selected time series:

$$Pred.Consumption = \ Eav_{n-1} \times \frac{1}{n} \times (E_n - E_{n-1}) \quad (4)$$

where E_{av} – average consumption except the last value, E_n – last value, n – number of time series

- *Exponential Smoothing (ES)*

Exponential Smoothing is a method used to forecast future values based on past observations. It is especially useful for data that does not contain significant seasonal fluctuations. The method offers a simple way to smooth data and account for recent observations by assigning more weight to them. The larger the time series interval width, the smoother the trend will be [31]. If the data noisy, the next ES model can help identify common patterns:

$$E_{n+1} = E_{aver} + \alpha\ (E_n - E_{aver}) \quad (5)$$

where E_{aver} – average consumption, E_n – last value before forecasting, α - smoothing parameter.

If $\alpha \rightarrow 1$, then the influence of past values is reduced, only the last value is considered for the forecast.

A drawback of the method is its limitation in handling long-term trends or data with significant fluctuations or noise. It works well only with stable data and may provide poor predictions when dealing with complex models or non-stationary data.

ISSN (P): 2707-9031
ISSN (E): 2707-904X

Table 1. Analysis of short-term prediction algorithms

	Limitations
Least Squares Method	Trend model is rigidly fixed. Final forecast considers an influence of the latest values.
	High sensitivity to detecting unusual points in the data. Additional outliers can seriously skew the forecast results since squaring increases the number of images and
Moving Averages Method	Consumption values (time series) are heterogeneous
	Identification of large number of model parameters is resource-intensive
Exponential Smoothing	Trend model was formed at the end of the chosen period and does not extrapolate the current dependencies into long future
	Provides a much tighter value for close outputs than for long-term extreme.

Considering limitations in Table 1, it is clear that this research objectives should not be covered by short-time forecasting only but a complex of suitable methods to overcome all possible restrictions and tune forecasting results according to the needed time period whether its long- or short-time. It means that for this study a hybrid prediction should be considered, so that three machine learning (ML) methods are used: **gradient boosting regressor** (**XGBoosting**), **random forest** (**RF**), and **long short-term memory** (**LSTM**).

XGBoost is a powerful and effective machine learning method for regression and classification tasks [32]. It is based on the concept of gradient boosting, which combines multiple weak models (typically decision trees) into a single strong predictive model. In the context of energy consumption forecasting, XGBoost enables you to consider complex nonlinear dependencies and interactions between various factors that influence energy usage, thereby enhancing the accuracy of predictions. Its ability to prevent overfitting through regularization makes it a versatile choice for both short-term and long-term forecasting.

At the same time, XGBoost is more demanding in terms of computational resources than simpler models and requires careful hyperparameter tuning to avoid overfitting or underfitting. This increases the time and computational resources needed for training and optimizing the model.

RF is a machine learning method that uses multiple decision trees to increase the accuracy and reliability of predictions [33]. This technique allows for the assessment of the significance of each feature, helping to interpret the model and identify key factors influencing energy consumption. RF is well-suited for handling large amounts of data and can effectively process high-dimensional datasets.

Although RF works well with large datasets, it can become computationally expensive for large-scale tasks. Additionally, it is not particularly well-suited for handling sequential dependencies in time series data, which may limit its effectiveness in forecasting energy consumption, where historical data plays a crucial role.

LSTM is a type of recurrent neural network (RNN) specifically designed for time series processing and is able to take into account long-term dependencies [34]. This makes it highly effective for time series forecasting, where historical data needs to be taken into account (for example, forecasting electricity consumption). LSTM can adapt to changes in data and dynamically update its forecasts based on new information.

ISSN (P): 2707-9031
ISSN (E): 2707-904X

At the same time, LSTM models are more complex and require significant computational resources and time for training. They can also suffer from overfitting if the training data is not representative of the test set, and they need large datasets to perform optimally.

Considering that each of the aforementioned forecasting methods has its strengths and weaknesses depending on the complexity and nature of the data, it is advisable to use **hybrid methods** that combine the strengths of different techniques to provide better accuracy and adaptability [35]. For example, combining machine learning algorithms with time series methods can overcome the limitations of each method when used separately [36]. Hybrid models can capture both long-term trends and short-term fluctuations in energy consumption, making them quite versatile for real-world applications [37]. At the same time, hybrid models are often more complex and require a deeper understanding of each method for their proper integration. The computational cost of hybrid methods is higher, and they are more difficult to interpret due to the combination of different algorithms.

Before the data is utilized to train and test the models, it will be prepared and examined. Case study was conducted using Jupyter Notebook programming tool and common libraries to analyze selected ML algorithms.

Results

The dataset consists of information collected over a period of 3 years from IoT sensors that were installed in the building of Cornell University [38]. Every sensor node is set up to record and operate every 24 hours on average as well as energy data was recorded too. The appliances energy consumption (MW) is to be measured, then it was chosen as the target variable.

Data preprocessing is necessary to identify trends and features in electrical power consumption statistics. Due to the inconsistent value scales between features and missing data, preprocessing is time-consuming issue for standardization and imputation. It is evident that the data must be split into two sets: training and test samples. Object selection for each sample was performed using the *GridSearchCV* method.

Several key data preparation strategies were employed to enhance forecasting efficiency. One such strategy is feature engineering, aiming to explore not only the relationships between features but also how these features relate to the target variable. Feature engineering involved extracting new features from existing data. Correlation analysis was conducted to establish relationships between various variables.

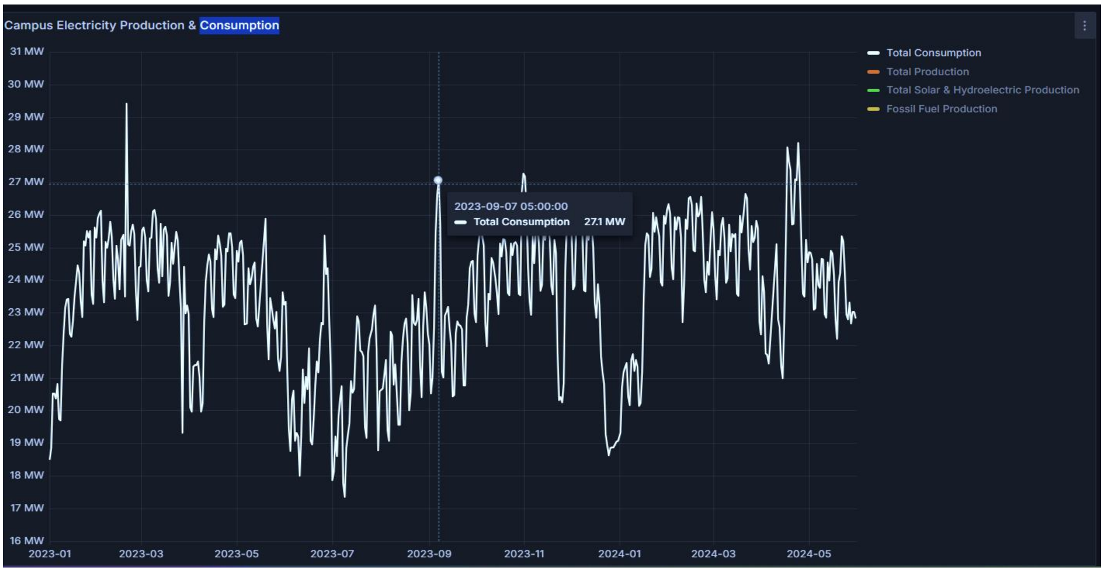

Figure 2. Energy consumption for 1-year period.

ISSN (P): 2707-9031
ISSN (E): 2707-904X

Additionally, values related to time parameters (hour, day, month, year) were consolidated into a single format, as unique time series over several years are required for forecasting. To ensure the data is ready for modeling, we checked for missing or empty values, duplicates, and addressed outliers using other statistical methods. This process aids in providing more relevant and valuable information.

Identifying correlation between variables is beneficial for various purposes, including detecting and subsequently removing irrelevant variables, as well as uncovering unique correlations that might not be evident during direct forecasting. In some cases, certain attributes can be modified and normalized within a specific range in order to reduce data imbalance. Figure 2 illustrates the energy consumption dynamics of a building over a period of one year.

As was mentioned before, proposed model consists of three ML algorithms to analyze patterns of hybrid forecasting. Considering time series data with different lengths, ***LSTM*** works appropriate. It is helpful for predicting energy use because it can handle variable-length sequences, recall prior data, and capture long-term associations. Three layers make up the LSTM model structure: input, unit, and output. Long sequences of data may be effectively collected and sent using LSTM.

A machine learning technique called ***Random Forest Regressor*** combines many decision trees to provide a prediction model for regression problems. A portion of the training data and features is chosen at random to build each tree. The final output is produced during prediction by the regressor combining predictions from each tree (Figure 3). Due to fact that it can learn complex behavior, this approach is frequently used for pattern identification and prediction.

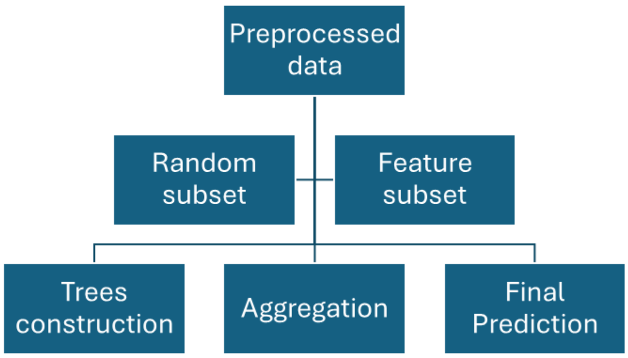

Figure 3. Random Forest Regressor algorithm.

Gradient Boosting technique is an iterative process that prioritizes failures at each stage. By matching weak training sets to the loss function, this technique reduces prediction error and enhances prediction performance. Gradient boosting was used in this work because of its strong predictive performance, capacity to recognize intricate data relationships and nonlinear patterns, and adaptability and customization features.

To avoid overfitting, a straightforward data partitioning technique was used throughout two sets of data. As was mentioned above, dataset was divided into training (20% of given data) and testing (80% of data) sets. ML algorithms and prediction models for consumption recordings were trained on training set. The effectiveness of selected models was assessed using a testing set.

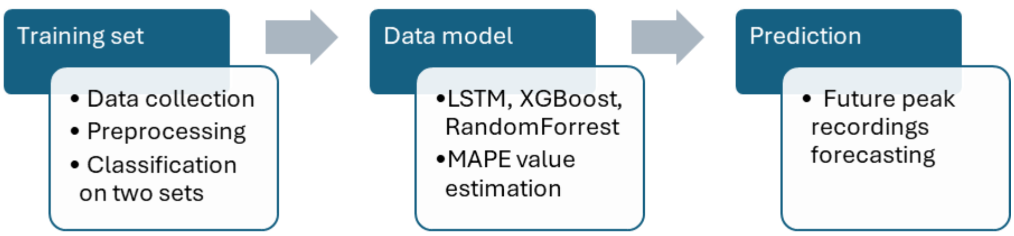

Figure 4. Procedure graph for prediction model based on three data modelling techniques.

ISSN (P): 2707-9031
ISSN (E): 2707-904X

Metrics as MAPE (mean absolute percentage error) were used to assess the accuracy and effectiveness of prediction models generated by ML algorithms. When comparing a model's forecast accuracy to the range of the actual values or different time series, MAPE can be considered as helpful tool due to fact that MAPE is able to scale the error measure to the actual value. The accuracy increases as the MAPE decreases [39].

This is a widely used technique for calculating prediction error, which is easier to grasp because of its scaled units. It calculates the average absolute percent inaccuracy for each time period fewer actual values divided by actual values.

The quality metrics were obtained by MAPE, showing the level of errors made by the algorithm in percentage:

$$\text{MAPE} = \frac{1}{number\ of\ samples} \times \left(\frac{E_{average} - E_{predicted}}{E_{average}} \times 100\%\right) \quad (6)$$

Table 2. Prediction performance using three presented building power metrics

Testing set	**Method**	**MAPE (%)**
Building 1	Random Forest	76.9
	LSTM	56.0
	XGBoosting	40.2
Building 2	Random Forest	38.4
	LSTM	48.1
	XGBoosting	52.5
Building 3	Random Forest	56.5
	LSTM	64.8
	XGBoosting	36.3

There are 2 scenarios of analytics:

- energy index during given period of time (actual values)
- energy consumption after optimization (predicted values)

Table 3. Prediction performance using three presented ML methods

Scenario 1 (Actual metrics)	**Scenario 2 (Random Forest)**	**Scenario 2 (LSTM)**	**Scenario 2 (XGBoosting)**
5602	5626	5637	5527
6842	6893	6884	6417
6173	6178	6366	6354
5789	5730	5596	6355

Discussion

Table 2 indicates that lower MAPE values suggest a better model fit since they identify the difference between anticipated and actual data. As it follows from Table 2, in the case of model on actual power consumption data, the least value was shown by the model based on the extreme gradient boosting regression algorithm (XGBoost) - the error was 36.3% which is less than other models. The GBR approach performed exceptionally well in every building. Nonetheless, there were differences in performance between buildings when comparing LSTM with Random Forrest technique. Furthermore, LSTM produced less mistakes than Random Forest, which showed a significant difference. This discovery implies that LSTM outperformed RF in terms of mistakes and generated fewer errors overall. After this analysis, it was evident that XGBoosting approach worked the best in all buildings.

With the data shown in Table 3, it is clear that Random Forest approach, which exhibits high accuracy, is the one that comes the closest to the actual testing results. In terms of accuracy in estimating average consumption, XGBoosting algorithm comes in bottom, followed by LSTM technique in second place. Nonetheless, there is a crucial performance disparity between LSTM and XGBoosting, indicating that both methods are not appropriate for the same types of data. While LSTM's recurrent nature makes it useful for addressing nonlinear, time-dependent data, XGBoosting performs well in all scenarios.

Table 4. Comparative analysis of power consumption forecasting methods

Method	**Forecast period**	**MAPE (%)**
Proposed model	Min 1 month	36.3 – 76.9
Hybrid (Differential evolution (DE) + LSTM) [40]	Min 1 month	21.8 - 35.2
Hybrid (RF + Linear regression) [41]	Min 1 month	11.4 – 51.0

Comparing to other articles, scientists and researchers are increasingly used hybrid models that combine more than two methodologies. The studies presented in Table 4 describe various approaches used to predict power consumption in comparison to the proposed model. Authors [40] used DE-LSTM algorithm to predict electricity price through accuracy estimation in market of Germany, France and Austria. In comparison with the proposed model, it achieves better forecasting performance in most cases. In [41] authors addressed the household energy load forecasting using hybrid models of RF (long-term) and Linear Regression (LR) for short-term predictions. The dataset contained measurements taken from a house near Paris from 2006 to 2010.

Table 5. MAPE values interpretation

MAPE (%)	**Forecast level**
More than 51%	Inaccurate
21% to 50%	Reasonable
11% to 20%	Good
Less than 10%	Highly Accurate

By taking into consideration proposed interpretation of MAPE [42] as was given in Table 5, it is possible to state that proposed model has reasonable accuracy in comparison to mentioned above related articles.

However, in compared articles and proposed model, exogenous factors that nonlinearly influence the amount of electricity consumption, such as meteorological, social, economic, etc., have not been studied, thus, it is clear that without the assumption made in this study regarding

ISSN (P): 2707-9031
ISSN (E): 2707-904X

weather factors, the forecast error may change. On the one hand, when using more detailed climate data, for example, several settlements with different climate conditions, MAPE should probably decrease. Therefore, one of the promising areas is further research into more detailed meteorological factors and their impact on the target result.

In particular, it is important to assess the risks of unexpected events related to accidents. Given that, forecasting of electricity consumption occurs on the basis of past events and established correlations between factors, new predictions can be taken into account based on the detection of a large mismatch error at the moment when those accidents appear. In this case, the error will occur immediately and it is possible to study the influence of the new factor on the forecast of electricity consumption.

Since one of the challenges in forecasting energy consumption is the lack of universal models suited for all topic areas and forecast lead periods, a potential field of study is the quest for universal ways to forecast model creation. Thus, the findings of this investigation, particularly the structure of input predictor factors, may be applied in comparable studies. The study's suggestions include the usefulness of utilizing gradient boosting models and neural networks to anticipate power usage.

Conclusion

To study smart building energy consumption forecasting, we applied long-term forecasting and used three ML methods: XGBoosting, RF, and LSTM. XGBoosting and RF are based on ensemble learning, which combines the results of several weak models to produce a more accurate forecast. XGBoosting uses gradient boosting to successively improve models, minimizing the error of each subsequent tree. RF builds multiple decision trees in parallel and averages their forecasts to improve reliability and accuracy. LSTM, as a type of recurrent neural network, is specifically designed to work with sequences of data and account for long-term dependencies. This makes LSTM especially useful for time series forecasting, where it is important to take historical data into account.

Each of these methods was tested on real energy consumption data, allowing us to evaluate their efficiency and accuracy in real cyber-physical systems. The results of the study showed that XGBoosting consistently achieves the best results. The LSTM method, which is specifically designed to handle sequential data, makes it ideal for activities such as time series forecasting.

Through rigorous testing and validation using real-world data on smart building energy consumption, the study revealed the strengths and weaknesses of each model category. While statistical models prove reliable for short-term forecasts with consistent patterns, machine learning algorithms demonstrate superior performance in scenarios involving complex dependencies and dynamic environmental influences. Hybrid models are a promising compromise that provides both accuracy and adaptability.

The work systematizes existing forecasting methods, including their strengths and weaknesses, and demonstrates the advantages of hybrid approaches for improving forecasting accuracy. The study utilizes data from real IoT sensors, making the results applicable to real-world conditions. At the same time, the study identified certain challenges: the high computational complexity of some methods, such as LSTM and hybrid models, the dependence of forecast accuracy on the quality of the input data, and insufficient attention to exogenous factors (e.g., climatic and social changes), which can significantly affect forecast accuracy.

The findings provide valuable insights for building managers and energy stakeholders seeking to optimize energy consumption in smart buildings. By understanding the capabilities and limitations of different predictive models, decision makers can select the most appropriate approach based on their specific needs and constraints. This study makes a significant contribution to the development of sustainable building practices by facilitating accurate energy forecasting and efficient use of resources.

ISSN (P): 2707-9031
ISSN (E): 2707-904X

References

[1] Iftikhar H, Bibi N, Canas Rodrigues P, López-Gonzales JL. Multiple Novel Decomposition Techniques for Time Series Forecasting: Application to Monthly Forecasting of Electricity Consumption in Pakistan. Energies. 2023; 16(6):2579. doi:10.3390/en16062579

[2] Almuhaini, S. H., & Sultana, N. (2023). Forecasting Long-Term Electricity Consumption in Saudi Arabia Based on Statistical and Machine Learning Algorithms to Enhance Electric Power Supply Management. *Energies*, 16(4), 2035.

[3] Poczeta, K., & Papageorgiou, E. I. (2022). Energy use forecasting with the use of a nested structure based on fuzzy cognitive maps and artificial neural networks. *Energies*, 15(20), 7542.

[4] Sayed, H. A., William, A., & Said, A. M. (2023). Smart Electricity Meter Load Prediction in Dubai Using MLR, ANN, RF, and ARIMA. *Electronics*, 12(2), 389.

[5] Pan, L., Wang, S., Wang, J., Xiao, M., & Tan, Z. (2022). Research on central air conditioning systems and an intelligent prediction model of building energy load. *Energies*, *15*(24), 9295.

[6] Hu, X., Li, K., Li, J., Zhong, T., Wu, W., Zhang, X., & Feng, W. (2022). Load forecasting model consisting of data mining based orthogonal greedy algorithm and long short-term memory network. *Energy Reports*, *8*, 235-242.

[7] Chaganti, R.; Rustam, F.; Daghriri, T.; De La Torre Díez, I.; Mazón, J.L.V.; Rodríguez, C.L.; Ashraf, I. (2022) Building Heating and Cooling Load Prediction Using Ensemble Machine Learning Model. Sensors, 22, 7692

[8] Hong, T., & Fan, S. (2016). Probabilistic electric load forecasting: A tutorial review. International Journal of Forecasting, 32(3), 914-938.

[9] Zhang, Y., Ma, R., Liu, J., Liu, X., Petrosian, O., & Krinkin, K. (2021). Comparison and explanation of forecasting algorithms for energy time series. Mathematics, 9(21), 2794.

[10] Ikeda, S., & Nagai, T. (2021). A novel optimization method combining metaheuristics and machine learning for daily optimal operations in building energy and storage systems. Applied Energy, 289, 116716.

[11] Amasyali, K., & El-Gohary, N. (2021). Machine learning for occupant-behavior-sensitive cooling energy consumption prediction in office buildings. Renewable and Sustainable Energy Reviews, 142, 110714.

[12] Flick, D., Keck, C., Herrmann, C., & Thiede, S. (2020). Machine learning based analysis of factory energy load curves with focus on transition times for anomaly detection. Procedia CIRP, 93, 461-466.

[13] Cao, M., Wang, J., Sun, X., Ren, Z., Chai, H., Yan, J., & Li, N. (2022). Short-Term and Medium-Term Electricity Sales Forecasting Method Based on Deep Spatio-Temporal Residual Network. Energies, 15(23), 8844.

[14] Senchilo, N. D., & Ustinov, D. A. (2021). Method for determining the optimal capacity of energy storage systems with a long-term forecast of power consumption. Energies, 14(21), 7098.

[15] Habbak, H., Mahmoud, M., Metwally, K., Fouda, M. M., & Ibrahem, M. I. (2023). Load forecasting techniques and their applications in smart grids. Energies, 16(3), 1480.

[16] Szczepaniuk, H., & Szczepaniuk, E. K. (2022). Applications of artificial intelligence algorithms in the energy sector. Energies, 16(1), 347.

[17] Xie, Y., Yang, Y., & Wu, L. (2022). Power consumption forecast of three major industries in China based on fractional grey model. Axioms, 11(8), 407.

[18] Frikha, M., Taouil, K., Fakhfakh, A., & Derbel, F. (2022). Limitation of deep-learning algorithm for prediction of power consumption. Engineering Proceedings, 18(1), 26.

[19] Aguiar-Pérez, J. M., & Pérez-Juárez, M. Á. (2023). An insight of deep-learning based demand forecasting in smart grids. Sensors, 23(3), 1467.

[20] IEA (2022), World Energy Outlook 2022, IEA. https://www.iea.org/reports/world-energy-outlook-2022.

[21] Ahajjam, M. A., Bonilla Licea, D., Essayeh, C., Ghogho, M., & Kobbane, A. (2020). MORED: A moroccan buildings' electricity consumption dataset. Energies, 13(24), 6737.
[22] Somu, N., MR, G. R., & Ramamritham, K. (2020). A hybrid model for building energy consumption forecasting using long short term memory networks. Applied Energy, 261, 114131.
[23] Suranata, I. W. A., Wardana, I. N. K., Jawas, N., & Aryanto, I. K. A. A. (2021). Feature engineering and long short-term memory for energy use of appliances prediction. TELKOMNIKA (Telecommunication Computing Electronics and Control), 19(3), 920-930.
[24] Shapi, M. K. M., Ramli, N. A., & Awalin, L. J. (2021). Energy consumption prediction by using machine learning for smart building: Case study in Malaysia. Developments in the Built Environment, 5, 100037.
[25] A. Satan, A. Khamzina, D. Toktarbayev, Z. Sotsial, I. Bapiyev and N. Zhakiyev, Comparative LSTM and SVM Machine Learning Approaches for Energy Consumption Prediction: Case Study in Akmola, 2022 International Conference on Smart Information Systems and Technologies (SIST), Nur-Sultan, Kazakhstan, 2022, pp. 1-7, doi: 10.1109/SIST54437.2022.9945776.
[26] Sagadatova, N., Talas, B., Nugumanova, A., & Zhakiyev, N. (2023). Forecasting electricity consumption: case study in Astana. Scientific Journal of Astana IT University, 14(14), 16–25. https://doi.org/10.37943/14TRMF1662.
[27] Mohammed, S. A., Awad, O. A., & Radhi, A. M. (2020). Optimization of energy consumption and thermal comfort for intelligent building management system using genetic algorithm. Indonesian Journal of Electrical Engineering and Computer Science, 20(3), 1613-1625.
[28] Y He, Y., & Tsang, K. F. (2021). Universities power energy management: A novel hybrid model based on iCEEMDAN and Bayesian optimized LSTM. Energy Reports, 7, 6473-6488.
[29] Ahmad, T., & Chen, H. (2020). A review on machine learning forecasting growth trends and their real-time applications in different energy systems. Sustainable Cities and Society, 54, 102010.
[30] Al-Shaikh, H., Rahman, M. A., & Zubair, A. (2019, November). Short-term electric demand forecasting for power systems using similar months approach based SARIMA. In 2019 IEEE International Conference on Power, Electrical, and Electronics and Industrial Applications (PEEIACON), pp. 122-126.
[31] Megat Muainuddin, M. M. A., & Abu, N. (2023). Forecasting Electricity Consumption Using Exponential Smoothing Methods. European Proceedings of Finance and Economics.
[32] Semmelmann, L., Henni, S., & Weinhardt, C. (2022). Load forecasting for energy communities: A novel LSTM-XGBoost hybrid model based on smart meter data. Energy Informatics, 5(Suppl 1), 24.
[33] Rathore, H., Meena, H. K., & Jain, P. (2023, January). Prediction of EV energy consumption using random forest and XGboost. In 2023 International Conference on Power Electronics and Energy (ICPEE) (pp. 1-6). IEEE.
[34] Ribeiro, M. H. D. M., da Silva, R. G., Ribeiro, G. T., Mariani, V. C., & dos Santos Coelho, L. (2023). Cooperative ensemble learning model improves electric short-term load forecasting. Chaos, Solitons & Fractals, 166, 112982.
[35] Arslan S. (2022) A hybrid forecasting model using LSTM and Prophet for energy consumption with decomposition of time series data. PeerJ Computer Science 8:e1001. https://doi.org/10.7717/peerj-cs.1001
[36] Z. Zheng, Y. Yang, J. Zhou and F. Gu, "Research on Time Series Data Prediction Based on Machine Learning Algorithms," *2024 IEEE 2nd International Conference on Control, Electronics and Computer Technology (ICCECT)*, Jilin, China, 2024, pp. 680-686, doi: 10.1109/ICCECT60629.2024.10546215

[37] AL-Ghamdi M, AL-Ghamdi AA-M, Ragab M. A Hybrid DNN Multilayered LSTM Model for Energy Consumption Prediction. *Applied Sciences*. 2023; 13(20):11408. https://doi.org/10.3390/app132011408
[38] Cornell University Energy Management Control System (EMCS). Real-time electricity consumption. https://portal.emcs.cornell.edu/d/2/dashboard-list?orgId=2
[39] Vivas, Eliana, Héctor Allende-Cid, and Rodrigo Salas. 2020. "A Systematic Review of Statistical and Machine Learning Methods for Electrical Power Forecasting with Reported MAPE Score" Entropy 22, no. 12: 1412.
[40] Chou, J.-S., & Tran, D.-S. (2018). Forecasting Energy Consumption Time Series using Machine Learning Techniques based on Usage Patterns of Residential Householders. Energy.
[41] Massidda, L., & Marrocu, M. (2018) Smart Meter Forecasting from One Minute to One Year Horizons. Energies, 11, 3520.
[42] Lewis C.D. (1982), Industrial and business forecasting methods. London: Butterworths.