Scientific Journal of Astana IT University
ISSN (P): 2707-9031 ISSN (E): 2707-904X
VOLUME 14, JUNE 2023

DOI: 10.37943/14TRMF1662

Nadezhda Sagadatova
Bachelor of Technical Sciences, Student of Master of Applied Data Analytics
sagadatovanadezhda@gmail.com, orcid.org/0009-0000-2349-6465
Astana IT University, Astana, Kazakhstan

Balnur Talas
Bachelor of Information Systems, Student of Master of Applied Data Analytics
222237@astanait.edu.kz, orcid.org/0009-0002-0461-6868
Astana IT University, Astana, Kazakhstan

Aliya Nugumanova
Research and Innovation Center "Big Data and Blockchain Technologies"
a.nugumanova@astanait.edu.kz, orcid.org/0000-0001-5522-4421
Astana IT University, Astana, Kazakhstan

Nurkhat Zhakiyev*
Department of Science and Innovations
nurkhat.zhakiyev@astanait.edu.kz, orcid.org/0000-0002-4904-2047
Astana IT University, Astana, Kazakhstan

FORECASTING ELECTRICITY CONSUMPTION: CASE STUDY IN ASTANA

Abstract: This paper presents the application of time series forecasting models, namely Forecaster Autoreg and Neural Network, for predicting electricity consumption in the city of Astana. Given the pressing need to address energy efficiency and reduce environmental impact in light of global climate change, employing scientifically sound and effective solutions is crucial. Machine learning methods offer a promising approach to tackle this challenge by predicting electricity demand. In this study, a comprehensive analysis of a time series dataset was conducted, encompassing data on electricity consumption (measured in MW) in Astana from January 1, 2020, to December 31, 2020, at an hourly interval. The dataset was utilized to accurately forecast electricity demand for each hour of the following day. To enhance the precision of predictions, additional factors such as air temperature and wind speed were incorporated. These factors were deemed significant due to Astana's distinct sharply continental climate and high windiness, making their inclusion essential for accounting for their influence on electricity demand and achieving more accurate forecasts. The neural network model was selected as the chosen methodology, as it has the capability to reveal intricate dependencies and patterns in the data, thereby facilitating more precise predictions of electricity consumption. To evaluate the accuracy and reliability of the forecasts, error indicators such as Average Absolute Error (MAE) and Average Absolute Percentage Error (MAPE) were employed. The results demonstrated that the proposed models could provide accurate forecasts with low errors.

Keywords: Machine learning, time series, prediction, electricity, consumption

Copyright © 2023, Authors. This is an open access article under the Creative Commons CC BY license

DOI: 10.37943/14TRMF1662
© Nadezhda Sagadatova, Balnur Talas,
Aliya Nugumanova, Nurkhat Zhakiyev

Introduction

The global population growth has led to an escalating demand for electricity, resulting in higher prices and the need for more efficient resource utilization. Projections indicate that by 2050, electricity may become the dominant energy carrier, accounting for approximately 50% of the global energy mix, up from 20% in 2019 [1]. This surge in energy demand is also observed in Astana, the capital city of Kazakhstan, which is undergoing rapid urbanization and facing infrastructure deterioration. To address these challenges, Astana is striving to transform into a smart city by implementing digitalization strategies for the heat supply system, control and optimization systems for Combined Heat and Power Plants (CHPP), and intelligent thermal and electricity grid systems powered by renewable energy sources [2]. Machine learning algorithms applied to resource forecasting can enhance management decision-making, reduce costs, and ensure adequate distribution during peak hours [3, 4]. Accurate forecasting of electricity consumption is crucial for the efficient allocation of resources, cost reduction, and promoting intelligent usage, particularly considering that energy consumption constitutes a significant portion of total societal expenditure [5].

Data-driven analysis and prediction of energy usage have become increasingly relevant in recent research. In this article [6], a comprehensive review of heat energy consumption is presented, with a particular focus on the utilization of machine learning algorithms for forecasting future heat energy consumption. The authors employed various machine learning methods, including Linear Regression, K-neighbors Regressor, and Random Forest Regressor, to predict the consumption of thermal energy in different city zones and its correlation with ambient temperature and wind. The findings of this study enable the identification of inefficient high-loss zones and overheated zones, which can potentially contribute to the implementation of Smart City concepts. By leveraging machine learning techniques, this research offers valuable insights and predictive capabilities for optimizing energy consumption and enhancing the efficiency of thermal energy systems.

Siti et al. conducted a study focusing on the impact of weather conditions on electricity demand in Bali Island, Indonesia, using data from 2018-2019 [7]. Two models, namely the Generalized Regression Neural Network (GRNN) and Support Vector Machine (SVM), were employed in the study. By systematically adding weather parameters as features, the study revealed the relative importance of different weather factors in relation to electricity load. The analysis indicated that temperature exhibited the highest correlation with electricity load, followed by sun radiation and wind speed. These findings emphasize the crucial role of weather conditions in electricity load forecasting. The GRNN and SVR models demonstrated superior performance in predicting electricity load, with correlation coefficient values of 0.95 and 0.965, respectively.

Another commonly used method for energy forecasting is the LSTM (Long Short-Term Memory) method [8]. The paper [9] proposes the utilization of two machine learning techniques, namely Long Short-Term Memory (LSTM) and Support Vector Machine (SVM), for energy consumption forecasting within the framework of the Smart City concept. The research focuses on addressing the long-term load forecasting problem specifically in the Aqmola region of Kazakhstan and aims to compare the accuracy of the two models in this context. The study presents the results of the comparison, revealing that the LSTM model outperforms the SVM model slightly in terms of Root Mean Square Error (RMSE) values. Both models, however, demonstrate the capability to predict energy consumption patterns effectively. The authors highlight that the accuracy of the models could be further improved by expanding the size of the training dataset and incorporating additional weather-related variables.

The existing literature demonstrates the use of various methodologies in leveraging machine learning techniques for energy system prediction. In this study, the authors assess the

Scientific Journal of Astana IT University
ISSN (P): 2707-9031 ISSN (E): 2707-904X
VOLUME 14, JUNE 2023

effectiveness of two forecasting approaches: the ForecasterAutoreg model, which employs recursive multi-step forecasting, and the Neural Network Model [10], [11]. By comparing the performance of these two models, the investigation aims to evaluate their predictive capabilities in the context of energy system forecasting. The research aims to accomplish the following tasks:

a) Investigate and analyze the distribution and patterns of electricity consumption, considering various factors such as seasonality, days of the week, and weather conditions, specifically air temperature and wind speed.

b) Train a recursive multi-step forecasting model using the available data to make accurate predictions.

c) Develop and train a neural network model that incorporates additional factors, such as weather conditions, to enhance the accuracy of the forecasts.

d) Evaluate the accuracy of the generated forecasts and draw conclusions based on the results obtained from the forecasting models.

1. Materials and methods

1.1 The Data

This study utilizes a dataset encompassing hourly electricity usage in Astana, along with corresponding data on air temperature and wind speed. The electricity usage data is provided by the Kazakhstan Electricity Grid Operating Company (KEGOC), while the air temperature data is obtained from the Kazgidromed Agency. The dataset covers the time period from January 1st, 2020 to December 31st, 2020, comprising 8,762 observations and 4 variables. Before conducting the analysis, the dataset underwent preprocessing procedures to address certain issues. These procedures aimed to handle missing values and irregular time intervals present in the data. Specifically, while the energy data was available at hourly intervals, the temperature and time data were recorded at unordered intervals and contained some missing values.

1.2 Data analysis

The time series plot spanning from September 8th, 2020, to September 9th, 2020, provides a view of the scaled power usage behavior for a single day. The plot exhibits a distinct peak in power consumption during mid-morning hours, around 10-11 am, and in the evening at approximately 9 pm. Following this, there is a gradual decline in consumption from midnight to the early morning hours. This pattern indicates certain time periods of heightened electricity demand throughout the day (Fig. 1). An analysis of the annual seasonality using a boxplot reveals higher median demand values for November and December compared to other months (Fig. 2a). This suggests a seasonal pattern in electricity consumption, with increased usage during the late autumn and winter months. Furthermore, another boxplot analysis (Fig. 2b) demonstrates that the median values for each day of the week are similar. However, it is observed that energy consumption tends to be slightly higher on Sundays compared to the other days of the week.

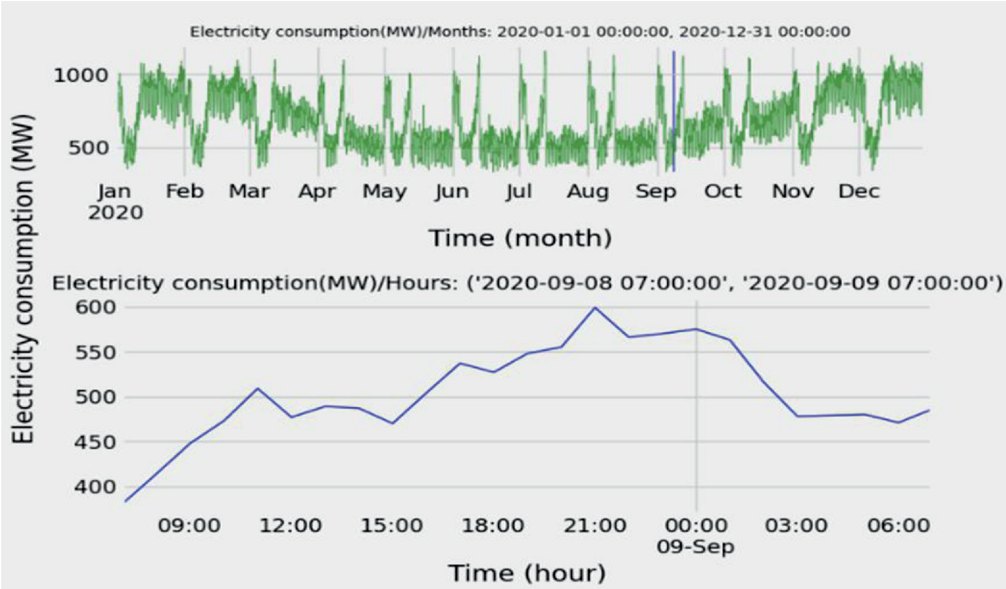
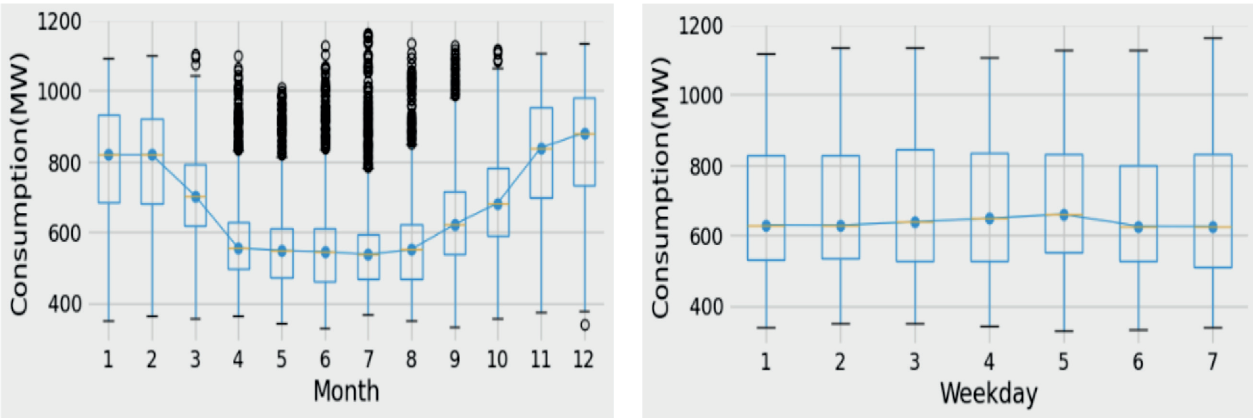


Figure 1. Zooming time series chart



a) b)

Figure 2. Electricity consumption distribution by a) months, b) weekdays

As part of the exploratory data analysis, the distribution of electricity demand across different categories of wind speeds was visualized. The box plot analysis reveals a distinct pattern wherein higher electricity demand is observed during periods of moderate breeze, while lower demand is observed during periods of calm and light air wind. This finding suggests that the availability of wind energy may play a crucial role in influencing electricity demand patterns in Astana.

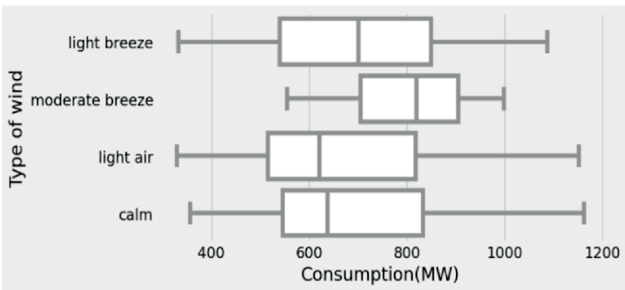


Figure 3. Distribution of consumption between types of wind

Scientific Journal of Astana IT University
ISSN (P): 2707-9031 ISSN (E): 2707-904X
VOLUME 14, JUNE 2023

1.2.1 Descriptive Statistics

Descriptive statistics were computed for the energy data in Astana, and the results are presented in Table 1. These statistics offer a summary of the central tendency, variability, and range of electricity consumption in Astana. The mean and median values illustrate that the typical hourly electricity usage in Astana is approximately 4 units. The range of consumption spans from a minimum of 0 units to a maximum of 20 units. These statistics provide insights into the general level of energy consumption and the variability in usage patterns within the dataset.

Table 1

	Consumption	Temp	Wind
count	8761.00	8620.00	8620.00
mean	677.51	4.324	4.96
std	186.77	14.15	2.934
min	330.00	-30.00	0.00
25%	530.00	-6.00	3.00
50%	636.00	5.00	4.00
75%	828.00	16.00	6.00
max	1162.00	35.00	20.00

1.3 ForecasterAutoreg (the recursive multi-step forecasting) model

A recursive autoregressive model, known as ForecasterAutoreg, is employed in this study. The model is trained using a linear regression approach with a Ridge penalty. A time window consisting of 24 lagged values is utilized, where the demand values from the preceding 24 hours are used as predictors for each prediction made by the model. The recursive multi-step-ahead forecasting technique is employed, wherein a single time series model is estimated, and each forecast is computed using the previously generated forecasts [10]. By adopting this methodology, the study aims to leverage the autoregressive nature of the time series data and exploit the relationship between past demand values to make accurate predictions for future electricity consumption.

$$y_t = m(x_{t-1}; \theta) + e_t, \text{ where } x_t = [y_{t_3} \dots y_{t-p+1}]' \text{ and } E[e_t] = 0 \tag{1}$$

More precisely, the forecasts are computed as follows:

$$\widehat{m}^{(h)}(x_t) = \begin{cases} \widehat{m}\left(\left[\widehat{m}^{(h-1)}(x_t), \dots, \widehat{m}^{(h-p)}(x_t)\right]'\right), & if\ h > 0; \\ x'_t \omega_h, & if\ 1 - p \leq h \leq 0; \end{cases} \tag{2}$$

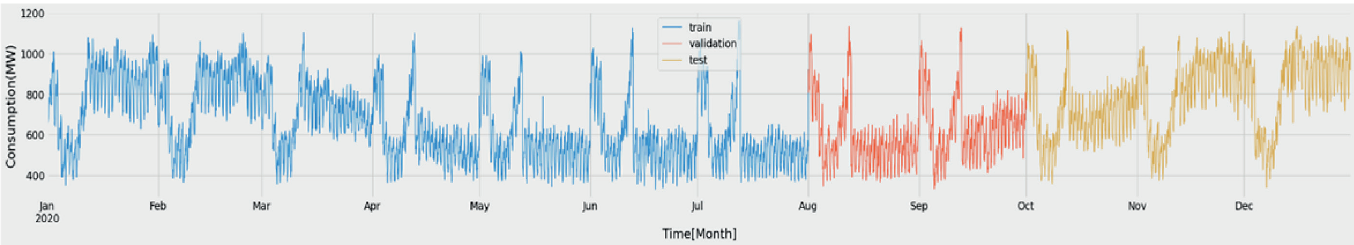


Figure 4. Electricity consumption training data

In this study, the training dataset comprised approximately 70% of the total data, covering the period from January 1, 2020, to August 1, 2020. The remaining 30% of the data was evenly divided between the validation and test datasets, with each accounting for 15% of the total records. The validation dataset concluded on October 1, 2020. Figure 4 visualizes the electricity consumption training data, where the blue line represents the training data, the red line represents the validation data, and the yellow line represents the test data.

1.4 Neural Network Model

The neural network model provides a powerful tool for capturing complex patterns and dependencies in time series data, making it well-suited for accurate and reliable electricity consumption forecasting. Inspired by the structure and functioning of the human brain, a neural network is a computational model composed of interconnected nodes, referred to as artificial neurons or units, which collaborate to process and analyze data [11]. Among the various neural network architectures, the multilayer perceptron (MLP) is commonly used for time series analysis. The MLP employs the backpropagation algorithm for training [12]. It is characterized by multiple layers of interconnected computational units arranged in a feed-forward manner. Typically, an MLP consists of an input layer, one or more hidden layers, and an output layer. The output of each unit serves as input to units in the subsequent layer. The connections between units in consecutive layers are defined by learned weights, which are updated during training using the backpropagation algorithm.

$$E = \frac{1}{2}\sum_{i=1}^{N}(X_i - X_i^*)^2 , \tag{3}$$

where, X is the observed value of i sample and X is the i-th predicted value for i sample.

When predicting electricity consumption, a neural network can be trained to learn from historical patterns and the relationships between various factors that influence electricity usage. These factors may include variables such as time of day, wind speed, temperature, seasonality, and other relevant variables. In this research paper, a neural network training model was employed to incorporate temperature and wind speed as input factors. The dataset was divided into training and test samples to assess the model's performance. The average absolute error (MAE) was chosen as the loss function, which quantifies the average absolute difference between the predicted and true values. The Adam optimizer, known for its efficiency in optimizing neural networks, was selected to minimize the MAE losses. The training process was conducted over 100 epochs, with each epoch representing a complete pass through the entire training dataset. A batch size of 32 was utilized, meaning that the model updates its weights and biases after processing 32 data samples.

1.5 Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE)

In time series analysis, it is common to use various error metrics to evaluate the accuracy of forecasting models. Two popular error metrics are the Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE).

Mean Absolute Error (MAE):

The MAE is a commonly used metric for evaluating the accuracy of a forecasting model. It measures the average absolute difference between the actual and predicted values. The formula for calculating MAE is:

$$MAE = \frac{\sum_{i=1}^{n}|actual_i - forecast_i|}{n} , \tag{4}$$

where n is the number of observations, Σ is the sum of the absolute differences between the actual and forecasted values, and $|actual_i - forecast_i|$ represents the absolute value function.

Scientific Journal of Astana IT University
ISSN (P): 2707-9031 ISSN (E): 2707-904X
VOLUME 14, JUNE 2023

Mean Absolute Percentage Error (MAPE):

The MAPE is another popular error metric used in time series analysis. It measures the average percentage difference between the actual and predicted values. The formula for calculating MAPE is:

$$MAPE = \frac{\sum_{i=1}^{n} \left| \frac{actual_i - forecast_i}{actual_i} \right|}{n} * 100, \tag{5}$$

where n is the number of observations, Σ is the sum of the absolute percentage differences between the actual and forecasted values, and $\left| \frac{actual_i - forecast_i}{actual_i} \right|$ represents the absolute value function.

One of the advantages of using MAPE is that it scales the error metric to the magnitude of the actual value, which makes it easier to compare the accuracy of the forecasting model across different time series.

2. Results and Discussion

To assess the performance of the models, the authors plotted the predicted electricity demand against the actual demand. This comparison allows for the examination of any underestimation or overestimation by the models and the identification of potential biases or anomalies in the data that could impact the model's accuracy. The resulting graphs depict the predicted values (represented by the blue line) and the actual values (represented by the orange line) for the corresponding days and months. Figures 5, 7, and 9 illustrate the performance of the ForecasterAutoreg model, while figures 6, 8, and 10 correspond to the Neural Network Model.

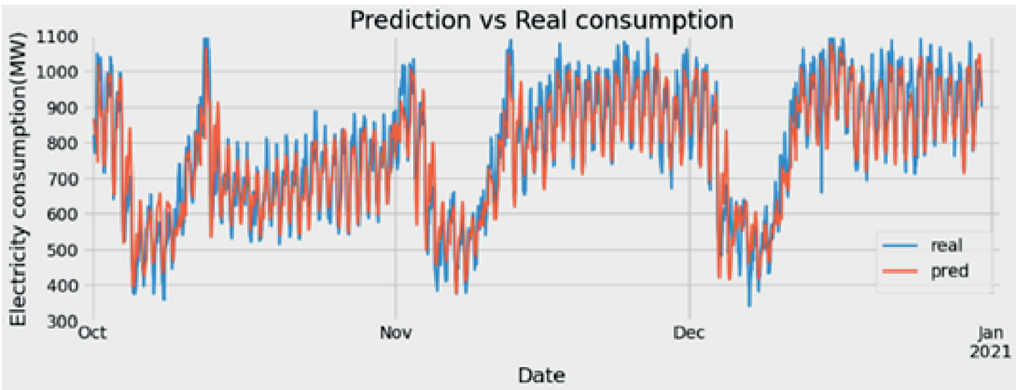


Figure 5. Prediction vs real consumption (Oct-Dec. 2020y.) by ForecasterAutoreg model (93.72%)

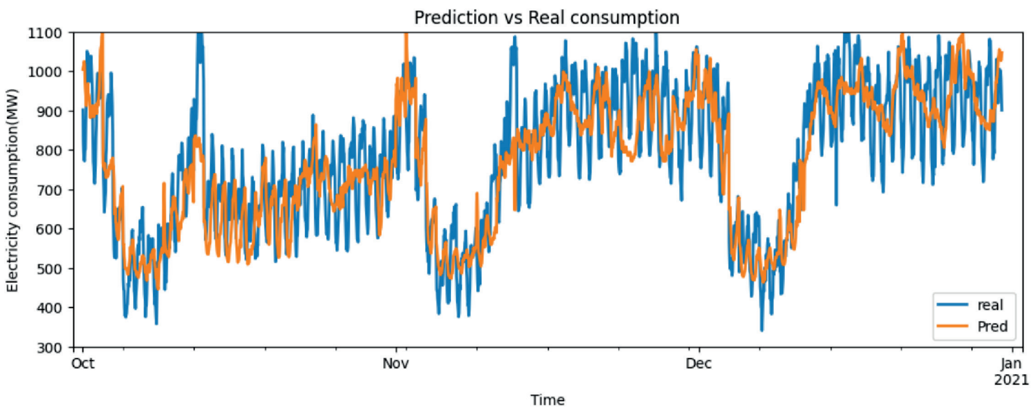


Figure 6. Prediction vs real consumption (Oct-Dec. 2020y.) by Neural Network model (88,57%)

DOI: 10.37943/14TRMF1662
© Nadezhda Sagadatova, Balnur Talas,
Aliya Nugumanova, Nurkhat Zhakiyev

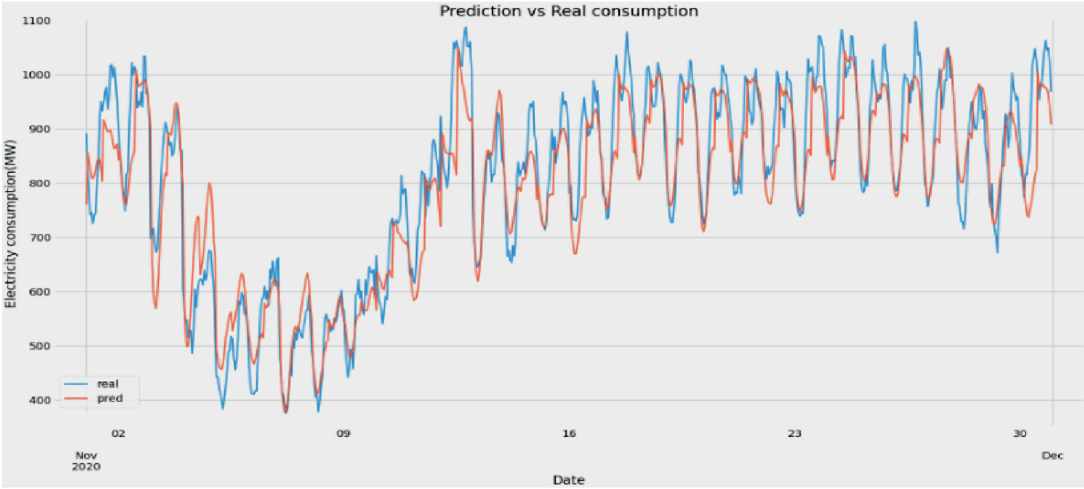


Figure 7. Prediction vs real consumption (Nov. 2020y.) by ForecasterAutoreg model (93.75%)

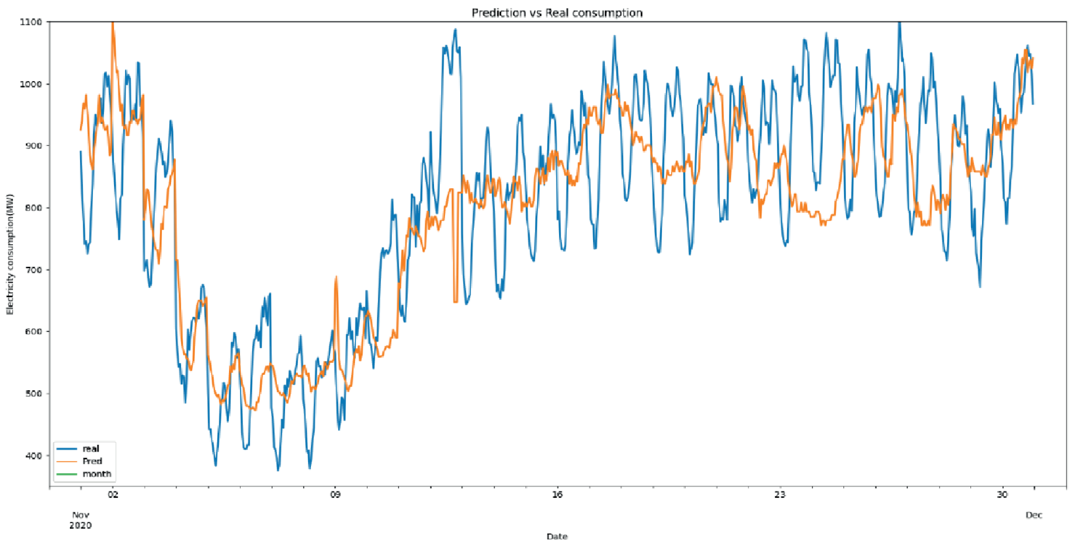


Figure 8. Prediction vs real consumption (Nov. 2020y.) by Neural Network Model (89.07%)

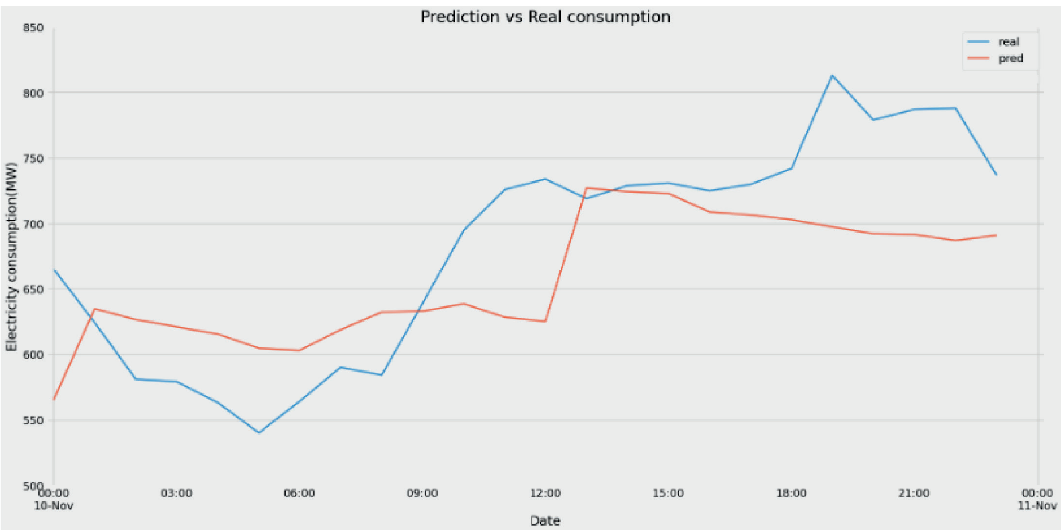


Figure 9. Prediction vs real consumption (10-Nov. 2020y.) by ForecasterAutoreg model 92.48%

Scientific Journal of Astana IT University
ISSN (P): 2707-9031 ISSN (E): 2707-904X
VOLUME 14, JUNE 2023

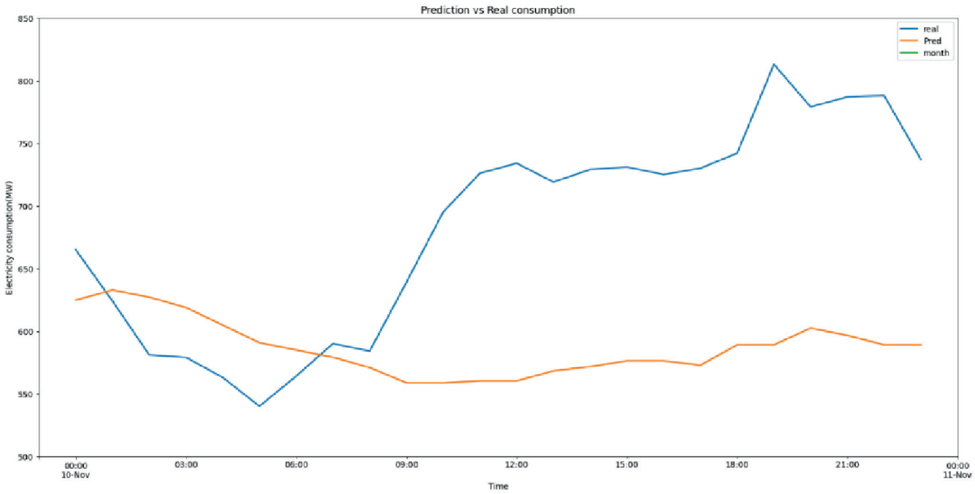


Figure 10. Prediction vs real consumption (10-Nov. 2020y.) by Neural Network Model (86.06%)

Table 2

	ForecasterAutoreg		Neural Network	
Prediction	MAE	MAPE	MAE	MAPE
Monthly	47.16	6.28	87.51	11.43
Each day	49.51	6.25	87.27	10.93
One day ahead	51.81	7.52	112.04	13.94

Table 2 presents the Mean Absolute Percentage Error (MAPE) values for the recursive multi-step approach and the Neural Network Model. It is observed that the MAPE values for the recursive multi-step approach are lower compared to the Neural Network Model, indicating its potential for improved accuracy through retraining. Particularly, both models perform most accurately in the day-ahead prediction, with MAPE values of 7.52 and 13.94, respectively.

Conclusion

The dynamics of electricity demand in Astana exhibit distinct patterns, with peak consumption observed during mid-morning and evening hours. Seasonal variation is evident, with higher demand occurring in November and December. Sundays show slightly higher consumption compared to other weekdays. Furthermore, wind speed influences electricity demand, as moderate breeze wind speeds correlate with higher consumption, while calm and light air winds result in lower consumption. The study demonstrates that the ForecasterAutoReg model provides accurate and reliable forecasts for electricity demand in Astana. The Neural Network Model is trained to identify historical patterns and dependencies among factors such as temperature and wind speed, which impact electricity consumption. Evaluation of the models' performance using error metrics like Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE) indicates relatively low errors in the forecasts. However, it is crucial to acknowledge that forecasting electricity demand is a complex task influenced by multiple factors, including weather conditions, economic factors, and social factors. Future research could focus on developing more sophisticated models that consider a broader range of factors, leading to more precise predictions of electricity demand. These advanced models can facilitate the optimization of production processes and enhance resource efficiency in the energy sector.

Acknowledgements

This research has been funded by the Science Committee of the Ministry of Science and Higher Education of the Republic of Kazakhstan (BR10965311).

References

1. International Renewable Energy Agency. (2019). Global energy transformation: A roadmap to 2050 (2019 edition). Abu Dhabi: Author.
2. Zhakiyev, N., Kalenova A., & Khamzina A. (2022). The energy sector of the capital of Kazakhstan: Status quo and policy towards smart city. *International Journal of Energy Economics and Policy, 12*(4), 414–423. https://doi.org/10.32479/ijeep.13126
3. Lai, J.-P., Chang, Y.-M., Chen, C.-H., & Pai, P.-F. (2020). A survey of machine learning models in renewable energy predictions. *Applied Sciences (2076-3417), 10*(17), 5975. https://doi.org/10.3390/app10175975
4. Mosavi, A., & Bahmani, A. (2019). Energy consumption prediction using machine learning; A review. https://doi.org/10.20944/preprints201903.0131.v1
5. Runge, J., & Zmeureanu, R. (2021). A review of deep learning techniques for forecasting energy use in buildings. *Energies, 14*(3), 608. https://doi.org/10.3390/en14030608
6. Omirgaliyev, R., Zhakiyev, N., Aitbayeva, N., & Akhmetbekov, Y. (2022). Application of machine learning methods for the analysis of heat energy consumption by zones with a change in outdoor temperature: Case study for Nur-Sultan city. *International Journal of Sustainable Development and Planning, 17*(4), 1247-1257. https://doi.org/10.18280/ijsdp.170423
7. Siti Aisyah, Simaremare, A. A., Adytia, D., Aditya, I. A., & Alamsyah, A. (2022). Exploratory weather data analysis for electricity load forecasting using SVM and GRNN, case study in Bali, Indonesia. *Energies, MDPI, 15*(10), 1-17. https://doi.org/10.3390/en15103566
8. Chou, S.-Y., Dewabharata, A., Zulvia, F. E., & Fadil, M. (2022). Forecasting building energy consumption using ensemble empirical mode decomposition, wavelet transformation, and long short-term memory algorithms. *Energies, 15*(3), 1035.https://doi.org/10.3390/en15031035
9. Satan, A., Khamzina, A., Toktarbayev, D., Sotsial, Z., Bapiyev, I., & Zhakiyev, N. (2022). Comparative LSTM and SVM machine learning approaches for energy consumption prediction: Case study in Akmola. In *2022 International Conference on Smart Information Systems and Technologies (SIST)* (pp. 1-7). https://doi.org/10.1109/SIST54437.2022.9945776
10. Rodrigo, J. A., & Ortiz, J. E. Forecasting electricity demand with Python. Retrieved from https://www.cienciadedatos.net/py29-forecasting-electricity-power-demand-python.html
11. Dubey, A. D. (2015). Artificial neural network models for rainfall prediction in Pondicherry. *International Journal of Computer Applications, 120*(3), 30-35.
12. Duda, R., Hart, P., & Stork, D. G. (2001). Pattern Classification.