KPU International Journal of Engineering & Technology
Volume 2, Issue2, Dec. 2022
e-ISSN:2790-0819, p-ISSN:2790-0800

Date Received: 2022-08-21
Date Accepted: 2022-12-25
Date Published: 2022-12-30

# Forecasting Monthly Electricity Consumption Using Machine Learning and Statistical-Based Models: A Case Study

ABDUL JALIL NIAZAI [1*], MOHAMMAD AKBAR SHAHPOOR [2], ABDULLAH ZAHIRZADA [3], ABDUL RAHMAN SAFI [4], NOOR AHMAD NOORI [5]

[1]Assistant professor, Department of Software Engineering, Computer science Faculty, Kunduz University,1th District, Kunduz, Afghanistan. Email: ab.niazai@gmail.com

[2]Assistant professor, Department of Software Engineering, Computer science Faculty, Kunduz University, 1th District, Kunduz, Afghanistan. Email: akbarshshpoor@gmail.com

[3]Assistant professor, Department of Information Systems, Computer science Faculty, Kunduz University, 1th District, Kunduz, Afghanistan. Email: ab.zahirzada95@gmail.com

[4]Assistant professor, Department of Information Systems, Computer science Faculty, Kabul University, 3th District, Kabul, Afghanistan. Email: abrahman.safi@gmail.com

[5]Assistant professor, Department of Information Technology, Computer science Faculty, Kandahar University, 9th District, Kandahar, Afghanistan. Email: noorah.kdr@gmail.com

## Abstract

*Forecasting electricity consumption demand plays a vital role in short-term load balancing and long-term planning for having secure power systems, new energy generation, and transmission infrastructures. Knowing the consumption demand is also essential in preparing the power system according to the anticipated load status. In this research paper, we conducted a time series forecasting analysis on the monthly electricity consumption data of Kunduz province gathered from April 2011 to July 2022 to predict the consumption demand in this province. For this purpose, we employed two well-known forecasting methods: Multilayer perceptron (MLP) from machine learning and ARIMA from statistics. The MAPE and MAE metrics were used to measure and compare the predictive performance of developed MLP and ARIMA models. The comparison result found that MLP outperforms ARIMA in model forecasting accuracy. We believe that the outcomes of this research work significantly help Kunduz's regional electricity authorities in better energy management and allow them to make effective decisions regarding electrical energy cost and efficiency.*[*]

**Keywords:** Forecasting, Electricity consumption demand, MLP, ARIMA, Kunduz province.

[*] Corresponding Author

## 1. Introduction

The rapid increase in human population, insistence on large-scale industrialization, technology adoption, rise in energy cost, and the need to keep the country's economic growth positive have caused to drive the global demand for electricity consumption [1]. Indeed, electrical energy nowadays is considered one of the critical factors that can highly affect a country's economic growth and human living standards. The global demand for this energy increased by 4.5%, with over 1000 TWH in 2021. This upsurge was almost five times greater than the downfall in demand for electrical energy in 2020. Nearly 80 percent of the increase in demand was in emerging markets and developing economies [2]. Fulfilling the demand for electrical energy can be achieved from different sources, including wind, solar, hydro, coal, gas, and others. In 2021, the world's electrical energy production sources comprised 36% coal, 22% gas, and 38% clean electricity, including solar, wind, hydro, nuclear, and bioenergy [3].

Electricity consumption demand is steadily growing, and we are required to have statistical knowledge about this growth, which involves a predictive mathematical model of electricity consumption to foresee future demands. Forecasting electricity consumption demand plays a vital role in short-term load balancing and long-term planning for having secure power systems, new energy generation, and transmission infrastructures. Knowing the energy demand is also essential in preparing the power system according to the anticipated load status. Moreover, an accurate forecasting result enables authorities to better energy management and allows them to make effective decisions in terms of energy cost and efficiency. Hence, this paper proposes a mathematical model that predicts the electricity consumption demand in Kunduz. Kunduz province, located in northeast Afghanistan, shares a long international border with Tajikistan. With around 1,136,677 population, including 34% living in rural areas and 66% in urban areas, it is considered one of Afghanistan's largest provinces. Approximately 70% of the Kunduz population, including 95% of households in the city and 60% in a rural areas, have access to electricity. The total acquired electricity of Kunduz is supplied by imported power, and the consumption demand increased from 121,261,205 kWh equivalent in 2010 to 184,772,152 kWh in 2021, as per Kunduz regional electricity organization report.

Although the accessibility of households to the electricity grid has significantly improved, load shedding is common in most areas of Kunduz province. The reasons are weak transmission infrastructure, lack of generation capacity, unreliable power systems, and no precise knowledge about electricity demand growth in the upcoming years. Hence, we believe the outcome of this research significantly helps Kunduz's regional electricity organization make an effective energy strategy for solving the prevenient electricity supply shortage in the next three years. The rest of this research paper is organized as follows. Section 2 discusses the related literature review. Section 3 presents the data and research methodology used. The model development and discussions on results are given in section 4. Finally, section 5 concludes this research paper.

## 2. Related Work

Increases in electricity costs caused energy modeling and forecasting to become a hot topic for academic and research scholars [4]. Researchers employed different techniques to predict the electrical energy consumption demand. For instance, S. L. Lai *et al.*[5] adopted artificial neural network (ANN) and Auto-regressive moving average (ARIMA) models to predict the electricity consumption demand in Hong Kon. As a comparison result, ANN has been found acceptable regarding forecasting accuracy when considering a single predictor. J. Miao used the ARIMA model to forecast China's energy consumption demand for formulating economic policies [6]. C. Hamzacebi and H. A. Es [7] employed an optimized Gray model to predict Turkey's electric energy demand from 2013 to 2025. A. Bagnasco *et al.*[8] implemented ANN-based multilayer perceptron algorithm to forecast the electrical consumption of the Cellini medical hospital in Turin. The model was proved with satisfactory forecasting accuracy. S. Mishra and V. K. Singh [9] applied the ANN-based Windowed Momentum algorithm to three years of monthly consecutive consumed energy data and weather information to predict monthly energy consumption. S. G. Yoo and H.-Á. Myriam [10] used neural networks and information from other variables such as the overall population and the elderly population to forecast the electricity consumption demand of residential areas in Seoul city. A. Camara *et al.*[11] deployed seasonal ARIMA and Multilayer perceptron approaches to predict U.S. energy consumption. As an empirical evaluation result, they found that the MLP (ANN) model slightly outperforms the seasonal ARIMA model in forecasting accuracy.

## 3. Methodology

Researchers unanimously agree with the forecasting literature that any single model may not always capture the distinctive patterns in the data equally well since data-analytic problems are frequently dynamic and complicated. Hence, based on the literature conducted above, two well-known forecasting approaches, Multilayer Perceptron (ANN) and ARIMA, are employed to predict the electricity consumption demand in Kunduz province for the next three years.

### 3.1 Data Collection

The monthly electricity consumption data of Kunduz province used in this research was obtained from Kunduz regional electricity organization over the period from April 2011 to July 2022. The dataset consists of 124 observations with two features (i.e., consumed electricity and corresponding month). Each observation in the collected data represents the monthly kWh consumed electricity in Kunduz province. In the data preprocessing stage, a total of 6 observations founded as missing values and handled through the Python Skylearn preprocessing library. The data distribution and description of our dataset are given in Table 1. Also, fig. 1 shows the time series plot of Kunduz's consumed electricity amount.

Table 1. The statistical description of the dataset.

| *Observations* | *Mean* | *Std* | *Min* | *25%* | *50%* | *75%* | *Max* |
|---|---|---|---|---|---|---|---|
| 124 | 12894330 | 4127101 | 582394 | 10452600 | 13212750 | 15593910 | 21400540 |

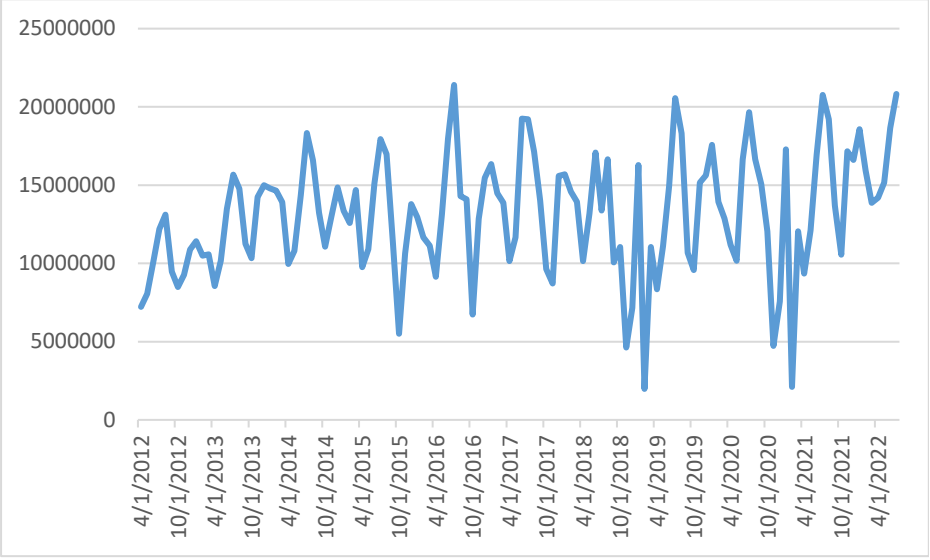



Fig. 1 Time series plot of electricity consumption in Kunduz.

As shown in Table 1, the minimum amount of electricity consumption is 582394 kWh, and the maximum consumption is 21400540 kWh. The average consumption is 12894330 kWh, whereas 50% of this province's monthly consumption is above 13212750 kWh.

Also, in the above plot graph, the x-axis represents the date, and the y-axis indicates the consumed electricity per month. As seen in Fig.1, consumption shows a steadily linear growth. In March and April, the plot marked a decrease in consumption, probably due to the cool weather in these months. However, the plot shows a sharp increase in consumption between July and August, probably due to summer air and intensive agricultural and industrial activities, as presented in Fig.1.

### 3.2 Proposed Models

Time series forecasting models' primary objective is to make future value predictions using data collected over a specific period. This research study employed two famous and most common time series forecasting models, Multilayer Perceptron (MLP) and Autoregressive Integrated Moving Average (ARIMA), to estimate and predict Kunduz's electricity consumption demand. The following phases summarize the development of the proposed models:

1. Developing an ARIMA model that includes the phases of model identification, model parameters estimation, and model diagnosis.
2. Constructing a Multilayer Perceptron model with optimum neural network architecture.
3. Evaluating proposed developed models utilizing various accuracy criteria.

### 3.3 ARIMA Model

Autoregressive Integrated Moving Average (ARIMA), also known as Box & Jenkins method, is one of the most popular models to forecast future values [12]. It predicts the future value of time series data as a linear combination of its previous values and forecast errors [13]. Unlike ordinary predictive models that involve explanatory variables, the ARIMA model is an utterly univariate method that uses its previous information in the series to forecast future value. Hence, the model is beneficial when little information, such as predictors, can be retrieved from the underlying data generation process and the variable to be forecasted is unknown. Three key components make up the ARIMA model: the autoregressive (AR) term, the Integrated (I) part, and the Moving Average (MA) part. For simplicity, each of these components is symbolized as parameters p,d, and q in the ARIMA model. Here, p represents the order of autoregressive term, d indicates the order of differencing to make the series stationary, and parameter q represents the size of the moving average process [14]. When parameter d=0, the series is stationary, and the ARIMA model is transformed into the ARMA model. Mathematically it can be written as follow:

$$y_t = c + \varepsilon_t + \sum_{i=1}^{p} \alpha_i \, y_{t-i} + \sum_{j=1}^{q} \beta_j \, \varepsilon_{t-j} \qquad (1)$$

Where $yt$ is the actual value of series $y$ at period $t$, $c$ is the constant value in the series, and $p$ and $q$ denote the order of autoregressive and moving average terms, respectively. Building the ARIMA model with time series data is detailed: (1) Determining the stationarity of the original sequence. Apply differencing transformation on the original series to stabilize the mean and variance before applying the ARIMA model to the data. (2) Estimate order for parameters p and q using statistical functions, i.e., calculating auto-correlation and partial auto-correlation of the data series. (3) perform a diagnostic check for the developed model to confirm that the model is well-fitted to the data and that the residuals are random errors. The three phases are repeated until an ARIMA model with the fewest generated errors is developed.

### 3.4 MLP model

Artificial Neural Network (ANN) is a category of learning algorithms that mimic the human brain system to discover patterns in the data. Multilayer Perceptron (MLP), also known as a feed-forward neural network, is one of the ANN algorithms used most

frequently in time series forecasting [15]. The MLP typically comprises three layers an input layer, a hidden layer or core processing layer of the network, and an output layer, as illustrated in Fig. 2.

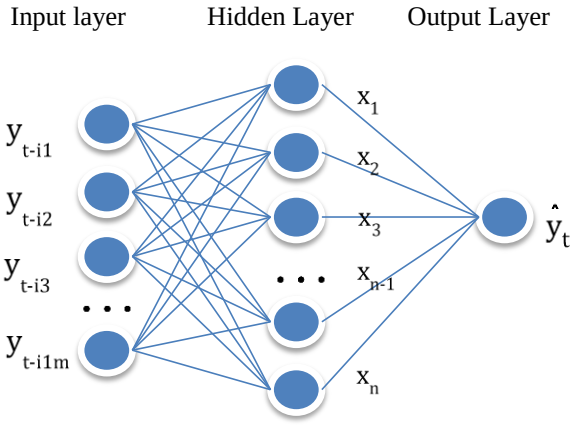



Fig. 2 Multilayer perceptron (MLP) neural network

The neural network of MLP iteratively learns by adjusting the weights of the neurons' connections based on the discrepancy between the actual and expected output. Initially, weights are assigned to neurons' connections with a random value between 0 and 1. The error estimated at the output layer is propagated back to the hidden layer, and weights are updated for subsequent iterations. This process continues until a minimum amount of errors is produced [9]. The equation below describes how MLP is trained for time series prediction.

$$y_t = w_0 + \sum_{i=1}^{n} w_i f\left(\beta_{0i} + \sum_{j=1}^{m} \beta_{ji}\, y_{t-j}\right) \qquad (2)$$

In the equation, $w_0$ ( $i$ = 0, 1, 2,3,…,$n$) and $B_{ji}$ ( $j$ =0,1,2,3,…,$n$) are coefficients also referred to as the network's connection weights. Where $n$ and $m$ represent the number of input and hidden nodes, respectively.

## 4. Proposed predictive model development

This study employed Multilayer Perceptron (ANN) and ARIMA, two well-known time series forecasting techniques, to model and predict electricity consumption demand in Kunduz province. All model development has been carried out by splitting the original electricity consumption data into two sets of training and testing, each containing 90% (112 data points) and 10% (12 data points) of the data, respectively. Python's data analysis library, called Pandas, was employed to split the dataset. The training set was used for

model building, and the test set was employed to assess the predictive accuracy and effectiveness of the developed models. Python's existing predictive modules are utilized for all modeling and forecasting tasks.

### 4.1 Developing ARIMA Model

Data series stationarity is a precondition for building an ARIMA model. In the ARIMA model development, as a first step, we are required to determine the stationarity of our data series. The Augmented Dick-Fuller (ADF) method was employed for this purpose. Table 2 contains the ADF test result for our data series before differencing.

Table 2. ADF test result before differencing

| *ADF value* | *P-value* | *Critical Values* | | *Result* |
|---|---|---|---|---|
| -2.3572 | 0.1541 | -3.491245 | @ 1% | Non-stationary |
| -2.3572 | 0.1541 | -2.888195 | @ 5% | Non-stationary |
| -2.3572 | 0.1541 | -2.580987 | @ 10% | Non-stationary |

As we can see, the calculated p-value is lower than the significant level ( i.e., p-value << 0.05), and the ADF value is higher than any critical value at the level of 10%, 5%, and 1%, respectively. Our series, therefore, is non-stationary. To make our series stationary, we must differentiate it at various time lags and then draw the autocorrelation function (ACF) to ascertain the proper differencing order for our ARIMA model.

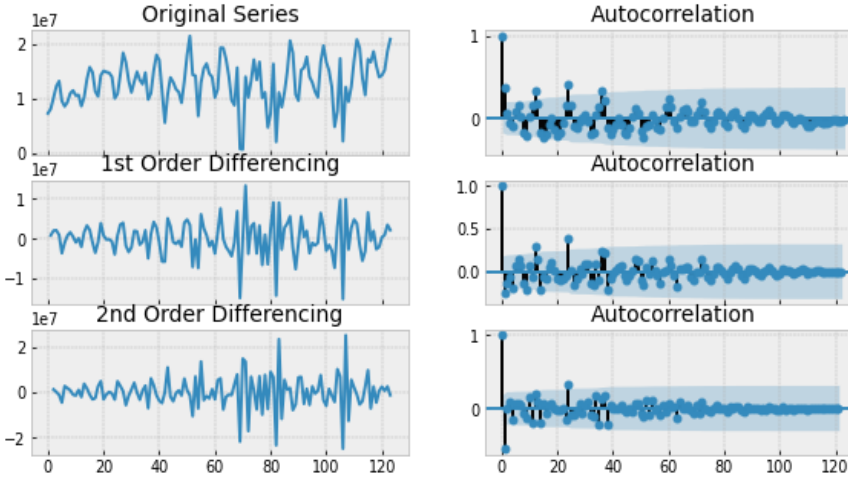



Fig. 3 The ACF plots various order of differencing.

The autocorrelation function plot steadily reached zero in the first order of difference. Also, some of the dotted horizontal lines occurred outside the significant (blue region) area, showing the series might not have been appropriately differenced see Fig. 3. However, in the second order of difference, the ACF plot quickly converged to zero, and the dotted horizontal lines did not cross the significant zone, showing that our series

perfectly transformed to stationary in the second order of difference. To verify this, let us apply the ADF test on our second-order differenced series again.

Table 3. ADF test result after second-order of difference.

| *ADF Statistics* | *P-value* | *Critical Values* | | *Result* |
|---|---|---|---|---|
| -6.952662 | 0.000000 | -3.492401 | @ 1% | Stationary |
| -6.952662 | 0.000000 | -2.888696 | @ 5% | Stationary |
| -6.952662 | 0.000000 | -2.581255 | @ 10% | Stationary |

The result displayed in Table 3 shows that the p-value is highly significant [<<0.05], and the ADF Statistics value is lower than all the critical values. Hence, we can set the differencing order as 2, depicting parameter d's value in the ARIMA (p,d,q) model.

Next, we determine the value of autoregressive (AR) and moving average (MA) terms, alternatively referred to as parameters p and q in the ARIMA model. The machine learning approach is used for this purpose. Grid search, a popular machine-learning technique for hyperparameter optimization, is used for this purpose. The method tries each combination of p and q from the subsets we provided. Additionally, to determine the ideal combination for parameters p and q, the method requires an objective function that evaluates how perfectly the model fits the data. The Akaike information criterion (AIC) function is commonly used for this purpose. AIC is a statistical measure used to evaluate and compare the ARIMA models in terms of how well they fit the data [14]. The lower the AIC value, the better the model is at forecasting error. Mathematically it can be described as:

$$AIC = -2l + 2K \quad (3)$$

Where $l$ is the log-likelihood of the model's prediction of the observed y-values and $k$ is the number of parameters in the model. Hence, we consider AIC values for optimal ARIMA model selection. So let us use the stats model module in Python to applying the ARIMA model to our consumed electricity data. Initially, we fixed ARIMA's AR and MA terms, each with a value of four. As a result, numerous ARIMA models were developed with different p and q parameter combinations. Table 4 holds the summary of developed ARIMA models having the least AIC value on our data.

Table 4. Development summary of candidate ARIMA models.

| *Model* | *Terms* | *Coefficient* | *P > \|z\|* | *AIC values* |
|---|---|---|---|---|
| ARIMA(0,2,3) | ma.L1.D2.kWh | -1.6362 | 0.000 | 3672 |
| | ma.L2.D2.kWh | 0.2818 | 0.124 | |
| | ma.L3.D2.kWh | 0.3545 | 0.000 | |
| ARIMA(0,2,4) | ma.L1.D2.kWh | -1.5796 | 0.000 | 3576 |
| | ma.L2.D2.kWh | 0.2972 | 0.146 | |
| | ma.L3.D2.kWh | 0.1784 | 0.396 | |
| | ma.L4.D2.kWh | 0.1170 | 0.250 | |

| | | | | |
|---|---|---|---|---|
| ARIMA(4,2,2) | ar.L1.D2.kWh | 0.4172 | 0.000 | 3678 |
| | ar.L2.D2.kWh | -0.0344 | 0.738 | |
| | ar.L3.D2.kWh | -0.0073 | 0.943 | |
| | ar.L4.D2.kWh | -0.0783 | 0.414 | |
| | ma.L1.D2.kWh | -1.9917 | 0.000 | |
| | ma.L2.D2.kWh | 1.0000 | 0.000 | |
| ARIMA(4,2,2) | ar.L1.D2.kWh | -0.4085 | 0.000 | 3682 |
| | ar.L2.D2.kWh | -0.3566 | 0.003 | |
| | ar.L3.D2.kWh | -0.2664 | 0.006 | |
| | ar.L4.D2.kWh | -0.3124 | 0.001 | |
| | ma.L1.D2.kWh | -1.0000 | 0.000 | |
| ARIMA(0,2,2) | ma.L1.D2.kWh | -1.8857 | 0.000 | 3683 |
| | ma.L2.D2.kWh | -0.8858 | 0.009 | |

According to the model development summary in Table 4, the ARIMA(0,2,3) model has the lowest AIC values compared to all other developed ARIMA models. The model's parameters p,d, and q have been determined to have values 0,2, and 3, respectively. Also, all MA terms' coefficient sum is less than one. Hence, we can conclude that the ARIMA(0,2,3) model has a high degree of fit compared to all other models constructed on our consumed electricity data (see Table 5).

Table 5. Development summary of the selected ARIMA model

| | | | |
|---|---|---|---|
| **Dep. Variable:** | D2.kWh | **No. Observations:** | 110 |
| **Model:** | ARIMA(0, 2, 3) | **Log Likelihood** | -1831.408 |
| **Method:** | CSS-mle | SD **of innovation** | 3853493.588 |
| **Date:** | Mon, 08 Aug 2022 | **AIC** | 3672.815 |
| **Time:** | 10:22:57 | **BIC** | 3686.318 |
| **Sample:** | 2 | **HQIC** | 3678.292 |

| | *coef* | *std err* | *z* | *P>\|z\|* | *[0.025* | *0.975]* |
|---|---|---|---|---|---|---|
| **const** | -872.4962 | 1025.473 | -0.851 | 0.395 | -2882 | 1137 |
| **ar.L1.D2.Value** | -1.6362 | 0.140 | -11.706 | 0.000 | -1.910 | -1.36 |
| **ar.L2.D2.Value** | 0.2818 | 0.183 | 1.538 | 0.124 | -0.077 | 0.64 |
| **ar.L3.D2.Value** | 0.3545 | 0.099 | 3.571 | 0.000 | 0.160 | 0.54 |

Next, the ARIMA (3,1,0) model is put through diagnostic testing to ensure it has provided the best fit for the data and that no more pertinent information in our data series is left for extraction. This can be achieved by performing the residual analysis of our best-fitted model. Fig 4. depicts the outcome of the residual analysis of the selected ARIMA(0,2,3) model.

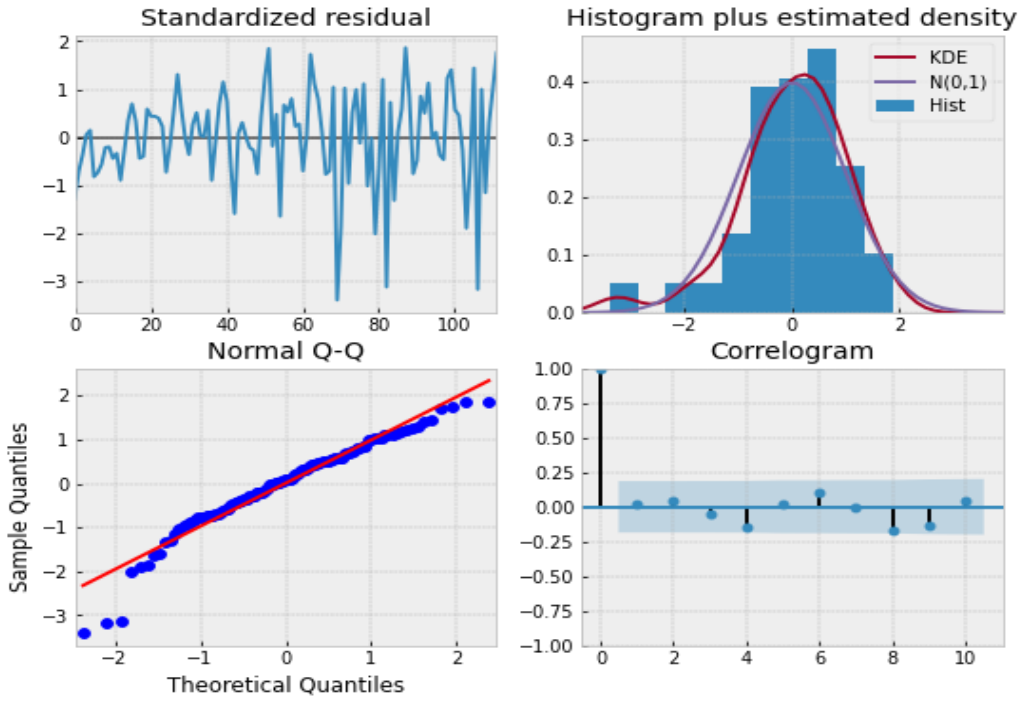



Fig. 4 Illustrates the residual errors, density plot, Q-Q plot of the residual, and ACF plot of the residual.

The model's residual errors, shown in Fig. 4, lack any discernible trend and retain a constant variance along the zero means. Similarly, the density plot illustrated in Fig. 4 implies that the residual errors have a normal distribution with near-zero means. This is verified by the Q-Q plot shown in Fig. 4, which suggests a linear trend for the residual errors of our proposed model. Likewise, the dotted vertical lines in Fig. 4 do not cross the significant threshold, showing no autocorrelation between our model's residual errors and its lagged version. So, from the statistics illustrated in the above figures, we can infer that our proposed ARIMA (0,2,3) model has successfully extracted all the essential knowledge from our consumed electricity data series and residual errors are white noise.

### 4.2 Developing MLP model

The fundamental structure of the MLP neural network must be defined before we apply the multilayer perceptron algorithm to our consumed electricity data series. We chose a single node in the input layer, eight neurons in a single hidden layer, and a node in the output layer as a preliminary network structure. The Stochastic Gradient Decent (SGD) optimization method is used to get optimal network architecture. Rectified Linear Unit (Relu) activation function is considered a transformation function, and Root Mean Square Error (RMSE) is used as a cost function. For model implementation and model evaluation, Python's open-source 'Keras' deep learning package is employed. The neural

network is trained for 400 epochs, for which we obtain the minimum amount of training and test errors.

After performing an empirical experiment, we found the MLP model with a single input node, two hidden layers each with 12 and 4 neurons, and a node in the output layer to be the best fit for our data series. This indicates that the model predicts electricity consumption demand for time $t_4$ using the consumed electricity amount of time $t_0$, $t_1$, and $t_3$. Hence, the 1-8-12-1 is the best neural network structure to train the MLP model on our data series. The forecasting accuracy of this proposed MLP model is assessed in the model evaluation and comparison section.

### 4.3 Model Evaluation and comparison

This research work has considered Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE) to measure and compare the predictive accuracy of developed models (i.e., ARIMA (0,2,3) and MLP). Researchers recommend these metrics when assessing the forecasting accuracy of a model on a single data series [16]. MAE metric measures a model's predictive performance by taking the absolute difference between the forecasted and actual values. Mathematically it can be formulated as:

$$MAE = \frac{\sum_{i=1}^{n} |Y_i - \hat{Y}_i|}{n} \quad (4)$$

In the formula, $\hat{Y}_i$ is the forecasted value, $Y_i$ is the actual value, and $n$ is the overall number of observations in the test set. A model in terms of performance is deemed superior to other models if it has the lowest MAE value. Likewise, the MAPE metric determines the accuracy of a model by taking the mean absolute difference between a forecasted value and an actual value divided by the actual value. The final result is expressed in percentages. A model is considered superior to others if it has a low MAPE. Mathematically it can be described as:

$$\text{MAPE} = \frac{1}{n} \left| \frac{X_i - \hat{X}_i}{X} \times 100 \right| \quad (5)$$

Where $\hat{X}_i$ indicates the predicted value, $X_i$ denotes the actual value, and the letter $n$ indicates the total number of observations in the test set. We can show a model's MAPE value in percentage of accuracy with the equation below:

$$\text{Acc\%} = 100 - \text{MAPE} \times 100 \quad (6)$$

Table 6 shows the predictive performance of MLP and ARIMA models in terms of mean absolute percentage error and mean absolute error on the test set. Likewise, Fig. 5 and Fig. 6 illustrate the ratio of the forecasted and actual value in predicting ARIMA and MLP models.

Table 6. Predictive performance of MLP and ARIMA models.

| *Model* | *MAPE* | *MAE* | *Accuracy (MAPE)* |
|---|---|---|---|
| Multilayer perceptron (MLP) | 18.00 | 11126380000000.0 | 82.0% |
| ARIMA | 21.50 | 11724960107369.0 | 78.5% |

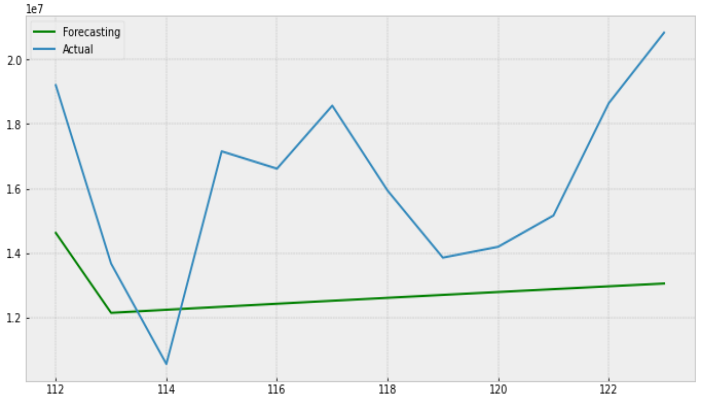



Fig. 5 Actual electricity consumption vs forecasted consumption by ARIMA model.

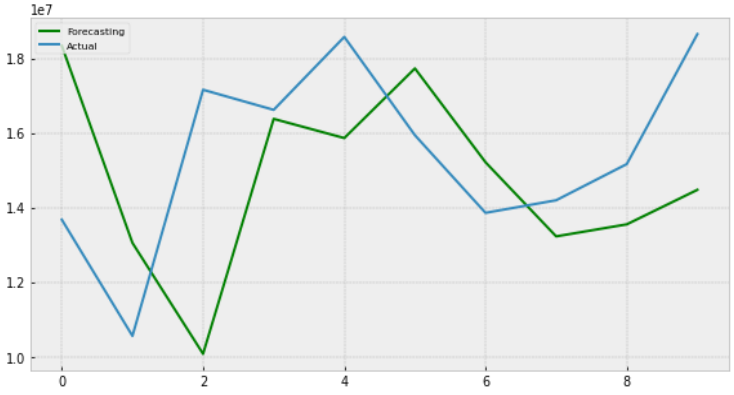



Fig. 6 Actual electricity consumption vs forecasted consumption by MLP model.

In Fig. 5 and Fig. 6, it can be observed that the electricity consumption values predicted by the MLP model are closer to actual consumption values than the prediction made by the ARIMA model. Also, from Table 6, it can be noticed that the MLP model has a lower rate of errors in terms of MSE and MAPE compared to ARIMA Model. In the case of forecasting accuracy, the MLP model has obtained 82%, with the ARIMA being modeled with lower accuracy of 78.5%. In general, from Table 6, Fig. 5, and Fig.

6, we can conclude that the MLP algorithm best fits our data and can be used to predict the electricity consumption demand of Kunduz province for the next three years.

## 5. Result and Discussion

As discussed previously, the main goal of this research study is to build a model that precisely forecasts the electricity consumption demand in Kunduz province for the next three years. For this purpose, we conducted a time series forecasting analysis on the monthly electricity consumption data of Kunduz province gathered from April 2011 to July 2022. ANN-based multilayer perceptron technique from machine learning and ARIMA model from statistics were employed. First, we implemented the ARIMA model on stationary series, and by considering the AIC criterion for model specification, we found the ARIMA (0,2,3) model best fitted our data. Subsequently, we applied multilayer perceptron on data series. As a result of an empirical analysis, we found a neural network structure with a single input node, two hidden layers each with 12 and 4 neurons, and a node in the output layer as the best MLP model. Next, considering the MAPE and MAE performance metrics, evaluation and comparison were performed between MLP and ARIMA-developed models. As a result, we found that the MLP model outperforms the ARIMA (0,2,3) model in forecasting electricity consumption demand. According to results shown in Table 6 and Figs. 4-5, we can claim that our MLP 1-8-12-1 model has best fitted our consumed electricity data and can be used to forecast the monthly electricity consumption demand of Kunduz province for the next three years. Table 7 and Fig. 7 depict the electricity consumption demand predicted by the MLP model from

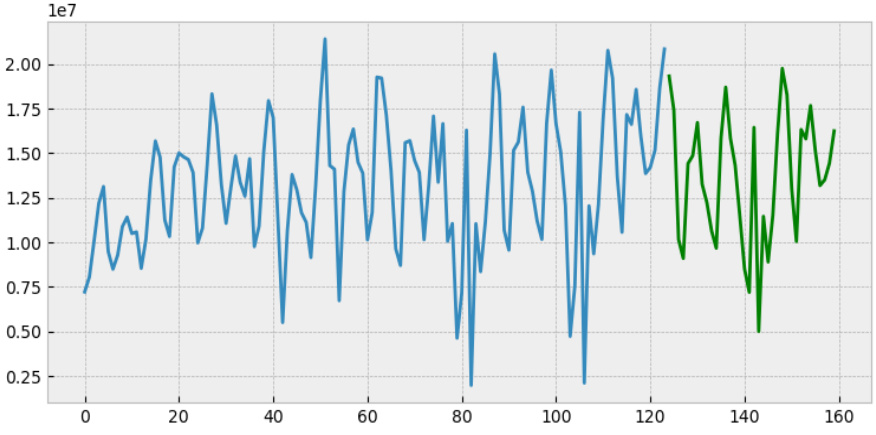



Fig. 7 Forecasted electricity consumption demand in Kunduz Province.

August 2022 to July 2025 in Kunduz province.

In the above plot graph, the x-axis represents the month, and the y-axis indicates the consumed electricity per month. The green curve in Fig. 8 represents the electricity consumption demand for the next three years. The blue curve depicts the previously consumed electricity amounts gathered from April 2011 to July 2022.

Table 7. Predicted monthly consumption demand from 2022 to 2025.

| *Date* | *Consumption demand (kWh)* | *Date* | *Consumption demand (kWh)* |
|---|---|---|---|
| 8/31/2022 | 19314937 | 2/28/2024 | 16439771 |
| 9/30/2022 | 17414938 | 3/31/2024 | 5005506 |
| 10/31/2022 | 10155868 | 4/30/2024 | 11450871 |
| 11/30/2022 | 9097788 | 5/31/2024 | 8899191 |
| 12/31/2022 | 14416275 | 6/30/2024 | 11523678 |
| 1/31/2023 | 14855235 | 7/31/2024 | 16109855 |
| 2/28/2023 | 16713509 | 8/31/2024 | 19746326 |
| 3/31/2023 | 13257330 | 9/30/2024 | 18268200 |
| 4/30/2023 | 12232078 | 10/31/2024 | 13009195 |
| 5/31/2023 | 10640783 | 11/30/2024 | 10048901 |
| 6/30/2023 | 9671043 | 12/31/2024 | 16318495 |
| 7/31/2023 | 15834149 | 1/31/2025 | 15806349 |
| 8/31/2023 | 18693112 | 2/28/2025 | 17664622 |
| 9/30/2023 | 15837168 | 3/30/2023 | 15160499 |
| 10/31/2023 | 14323535 | 4/30/2025 | 13183281 |
| 11/30/2023 | 11458181 | 5/31/2025 | 13504001 |
| 12/31/2023 | 8496194 | 6/30/2025 | 14425635 |
| 1/31/2024 | 7197085 | 7/31/2025 | 16239756 |

The prediction result reported by our MLP model shows that the electricity consumption demand will tend to have a slightly increasing trend over the proposed period in Kunduz province, as illustrated in Fig. 7 and Table 7. According to the forecasting result, the peak demand for electricity consumption will be in July and August. However, the model marked a decrease in consumption in March and April, probably due to the cool weather we had in these months. Overall, we can conclude that the demand for electricity consumption will somehow have an upward trend in Kunduz province for the upcoming three years. In this contest, we suggest to Kunduz regional electricity organization:

- Adopt more practical policies for electricity energy saving and efficiency.
- Make an effective energy generation strategy for solving the possible electricity supply shortage in the next three years.
- Facilitate investments by establishing a competitive market for generating electricity from other resources.

## 6. Conclusion

In this research paper, we conducted a time series forecasting analysis on the monthly electricity consumption data of Kunduz province to predict the consumption demand for the next three years. Two well-known forecasting methods, MLP from machine learning and ARIMA from statistics, were employed. Various MLP and ARIMA models were developed on our data series. By considering the AIC criterion for ARIMA and RMSE metrics as a cost function for MLP, we found the ARIMA (0,2,3) and MLP

with 1-8-12-1 neural network structure as the best-fitted models. Next, we assessed and compared the predictive performance of our established MLP and ARIMA models using the MAE and MAPE accuracy criteria. As for the evaluation result, we noted the MLP model with satisfactory forecasting results and can be used effectively to predict the consumption demand for the next three years. The prediction result from our MLP model showed that the electricity consumption demand would tend to have a slightly upward trend in Kunduz province for the next three years. The consumption might increase in July and August and decrease in March and April.

To the best of our literature knowledge, this is the first research work to apply time series forecasting approaches to predict the electricity consumption demand in Kunduz province. We believe that the outcomes of this research work significantly help Kunduz's regional electricity authorities in better energy management and allow them to make effective decisions regarding energy cost and efficiency. Also, the ministry of energy and water of Afghanistan and Kunduz's regional electricity organization can utilize the result of this research work in making effective decisions for short-term load balancing and strategic plans for new energy generation, transmission infrastructure, and preparation of secure power systems. For future research, we recommend that the forecasting models consider info from other predictors such as GDP, population, and indexations of industrial activity in Kunduz province. Also, in this research study, we implemented the conventional time series forecasting techniques. Future research can adopt other practical forecasting approaches, including the Gray model, Long Short-Time Memory (LSTM), and Support Vector Regression (SVR), which might improve the accuracy of consumption demand prediction.

## References


[1] V. Bianco, O. Manca, and S. Nardini, "Electricity consumption forecasting in Italy using linear regression models," *Energy (Oxf.)*, vol. 34, no. 9, pp. 1413–1421, 2009.

[2] International Energy Agency, "Global Energy Review 2012." [online] Available: https://iea.blob.core.windows.net/assets/d0031107-401d-4a2f-a48b-9eed19457335/GlobalEnergyReview2021.pdf. [Accessed: 24-Jul-2022].

[3] Statistical Review of World Energy. [online] Available at: <https://www.bp.com/content/dam/bp/business-sites/en/global/corporate/pdfs/energy-economics/statistical-review/bp-stats-review-2021-full-report.pdf> [Accessed 24 July 2022].

[4] C. Nichiforov, I. Stamatescu, I. Fagarasan, and G. Stamatescu, "Energy consumption forecasting using ARIMA and neural network models," in *2017 5th International Symposium on Electrical and Electronics Engineering (ISEEE)*, 2017.

[5] S. L. Lai, M. Liu, K. C. Kuo, and R. Chang, "Energy consumption forecasting in Hong Kong using ARIMA and artificial neural networks models," *Appl. Mech. Mater.*, vol. 672–674, pp. 2085–2097, 2014.

[6] J. Miao, "The energy consumption forecasting in China based on ARIMA model," in *Proceedings of the 2015 International Conference on Materials Engineering and Information Technology Applications*, 2015.

[7] C. Hamzacebi and H. A. Es, "Forecasting the annual electricity consumption of Turkey using an optimized grey model," *Energy (Oxf.)*, vol. 70, pp. 165–171, 2014.

[8] A. Bagnasco, F. Fresi, M. Saviozzi, F. Silvestro, and A. Vinci, "Electrical consumption forecasting in hospital facilities: An application case," *Energy Build.*, vol. 103, pp. 261–270, 2015.

[9] S. Mishra and V. K. Singh, "Monthly energy consumption forecasting based on windowed momentum neural network," *IFAC-PapersOnLine*, vol. 48, no. 30, pp. 433–438, 2015.

[10] S. G. Yoo and H.-Á. Myriam, "Predicting residential electricity consumption using neural networks: A case study," *J. Phys. Conf. Ser.*, vol. 1072, p. 012005, 2018.

[11] A. Camara, W. Feixing, and L. Xiuqin, "Energy consumption forecasting using seasonal ARIMA with artificial Neural Networks models," *Int. J. Bus. Manag.*, vol. 11, no. 5, p. 231, 2016.

[12] S. Ozturk and F. Ozturk, "Forecasting energy consumption of turkey by Arima model," *J. Asian Sci. Res.*, vol. 8, no. 2, pp. 52–60, 2018.

[13] A. Jalil Niazai, A. Zahirzada, M. Akbar Shahpoor, and A. Rahman Safi, "Time Series Forecasting of Registered, Recovered, and Death Cases of COVID-19 for

the Next Sixty Days in Afghanistan," *2020 IEEE International Conference on Advent Trends in Multidisciplinary Research and Innovation (ICATMRI)*, 2020, pp. 1-6, doi: 10.1109/ICATMRI51801.2020.9398481.

[14] R. Adhikari and R. K. Agrawal, "An introductory study on time series modeling and forecasting," arXiv [cs.LG], 2013.

[15] X. Fan, L. Wang, and S. Li, "Predicting chaotic coal prices using a multi-layer perceptron network model," *Resour. Policy*, vol. 50, pp. 86–92, 2016.

[16] R. J. Hyndman and A. B. Koehler, "Another look at measures of forecast accuracy," *Int. J. Forecast.*, vol. 22, no. 4, pp. 679–688, 2006.

**Authors Profile:**



**Abdul Jalil Niazai** is an assistant professor at Kunduz University's software engineering department. He got a master's degree in information management from the Asian Institute of Technology (AIT), Thailand, in 2019, and BSc in computer science from Peshawar University in 2015. His research interest lies in data science, machine learning, and computer vision. Abdul Jalil Niazai has been acting as leader of the research committee of the software engineering department since 2020 at Kunduz University.



**Mohammad Akbar Shahpoor** received a BCS degree in computer science from Nangarhar university in 2012 and an MSC degree in computer science from the Indian Academy Degree College of Bangalore, India, in 2018. Presently, he is working as an assistant professor in the department of software engineering at Kunduz university. His research areas include software engineering, data science, and machine learning.



**Abdullah Zahirzada** received a B.Sc. degree in computer science from Kunduz University, Afghanistan, in 2016 and an M.Sc. in Information Technology at the School of Information Technology (SIT) from King Mongkut's University of Technology Thonburi, Thailand, in 2022. Presently, he is working as an assistant professor in the department of Information Systems at Kunduz university. His research interests are Data Science, Data Mining, Image Processing, Machine Learning, and Cognitive Computing.



**Abdul Rahman Safi** received a BCS degree in Information Systems from Kabul University in 2013 and an MS degree in Information Management from the Asian Institute of Technology, Thailand, in 2019. His research areas include software engineering, data science, and artificial intelligence. Presently, he is working as an assistant professor in the department of Information Systems at Kabul University.



**Noor Ahmad Noori** is currently a Ph.D. scholar at the National University of Science & Technology, Pakistan. He received an MS in Information Management from the Asian Institute of Technology, Thailand, in 2018 and a BSc degree in Computer Science from Peshawar University, Pakistan, in 2014. His research interests include Data Science, Big Data Analysis, Machine Learning, and Computer Networking. Noor Ahmad has eight years of administrative, teaching, and research experience.