ЗАХИРИДДИН
МУҲАММАД БОБУР
НОМИДАГИ АНДИЖОН
ДАВЛАТ УНИВЕРСИТЕТИ

ANDIJAN STATE
UNIVERSITY NAMED
AFTER ZAKHIRIDDIN
MUKHAMMAD BABUR

ИЛМИЙ ХАБАРНОМА

Физика-математика
Тадқиқотлари
(Махсус сон)

SCIENTIFIC BULLETIN

Physical and
Mathematical Research
(Special Issue)

Андижон
2023 йил

Муассис

Захириддин Мухаммад Бобур номидаги Андижон давлат университети

ИЛМИЙ ХАБАРНОМА.
ФИЗИКА-МАТЕМАТИКА
ТАДҚИҚОТЛАРИ

Журнал бир йилда 2 марта чоп этилади.

Андижон вилояти ахборот ва оммавий коммуникациялар бошқармаси
томонидан 2019-йил 26 декабрда
0452 рақам билан рўйхатга олинган.

Нашр индекси: 344

Нашр учун масъул:
А.Й.Бобоев

Босишга рухсат этилди:
27.12.2019.
Қоғоз бичими: 60x81 1/8
Босма табоғи: 13,5
Офсет босма. Офсет қоғози.
Адади: 110 дона.
Баҳоси келишилган нархда.
Буюртьа №: 165.
“Муҳаррир” нашриёти манбаа бўлимида чоп этилди.
Тошкент шаҳри, Сўгалли ота кўчаси 7-уй

Таҳририят манзили:

170100, Андижон шаҳри, Университет кўчаси, 129. Телефон: +998911602043.
Факс: (374) 223-88-30

E-mail: adu_xabarnoma@mail.ru Расмий сайт: uzjournals.edu.uz/adu

Сборник статей международной научно-практической конференции по «Полупроводниковая опто- и наноэлектроника, альтернативные источники энергии и их перспективы» Андижан, 12-13 октября 2023 года

ОГЛАВЛЕНИЕ

Power load forecasting using linear regression method of machine learning: andijan regional case

N.N.Abdurazokov, R.U. Aliyev,

Andijan State University

Abstract: Integration of renewable energy sources (RES) into the grid has a huge effect on the supply behavior of the system. The output of the RES is very dynamic, so the reliability of the system strongly relies on the accurate balancing of generation capacity to the demand of the consumers. Generated energy is lost in the transmission networks during the off-peak time of the day, in contrast, the lack of energy during the peak time leads to the disconnection of certain areas from the network. This situation causes serious damage to the power grid system and economic losses for the operators. In this work, daily power consumption data of the Andijan region of Uzbekistan has been analyzed and the future amount of regional power consumption is forecasted by linear regression model. The performance of the proposed method has been evaluated based on real-world data. The proposed linear model has shown mean absolute percentage error in the prediction of 5.8% on the train data and 4.75 % on the test data.

Keywords: Renewable energy sources, power demand, load forecasting, machine learning, linear regression, model, performance evaluation.

1.Introduction. In recent years, the demand for electricity has increased several times in all regions of our country. The increase in industrial, customer service, and residential usage has created the need to increase the generation capacities. To solve this problem, experts in the field plan to increase production capacity and widely integrate renewable energy sources (RES) into the power grid. It should be taken into account that electricity in Uzbekistan is mainly produced by burning fossil fuel resources. The limitation of natural resources, their non-renewability, and the large amount of harmful gases released into the atmosphere make it necessary to switch from the usual production method to the production of electricity using renewable sources. Storing the generated electricity in batteries is a very expensive process, so it is necessary to continuously deliver it to the consumer.

Electricity consumption has an unstable nature, and during the peak hours of the day, the consumption is higher than the maximum supply amount, as a result, it creates stress in the power system, often creating "brownout" situations in some areas. At off-peak time of the day, mainly at night, the decrease in consumption to a minimum amount leads to the loss of the produced power in the network. Both situations are a source of great economic loss for utilities.

Currently, in the world experience, great attention is paid to increasing the efficiency of consumption by accurately predicting the future dynamics of consumption and managing it. Achieving an accurate estimate allows for optimal adjustment of the amount of dynamic power received by RES, which is increasingly used, to the amount of consumption. A lot of scientific research is being conducted on forecasting the future values of consumption. Traditional statistical methods, autoregressive (AR), moving average MA, and ARIMA methods, a generalized form of AR and MA methods, are widely used [1]. ARIMA was introduced by Box and Jenkins in the early 1970s as a model for forecasting and analyzing time series data, hence it was called the Box-Jenkins model. Later, an adapted form of ARIMA for short-term consumption forecasting was presented by Lee and Ko [2]. ARIMA algorithm based on "sliding window" was used in forecasting the consumption for the next short term according to the data obtained with the help of smart meters [3]. In general, short-term consumption characteristics have many random factors and strong subtle characteristics. Traditional methods are theoretically linear models that cannot perfectly represent consumer data's random and non-linear factors. That is why consistent research activities continue on new theoretical solutions and algorithms for short-term load forecasting [4].

2. Data preprocessing. The electricity consumption dataset is a time-series data. Time series data is used to refer to the results of measurements or observations taken at a certain time interval. For example, the information reflecting the daily values of electricity consumption for a year consists of 365-time series values. If measurements are made every 6 hours a day, 1460 time series data will be generated for a year. Today, it is possible to easily get information about electricity consumption every day or several times a day through smart meters from a distance. Due to a small number of failures in the communication channel, the information received does not always consist of accurate values, and in some cases, as a result of interruptions in the communication channel, the presence of anomalous values in the information or a drop in the value are also observed. Passing such information directly to the prediction algorithm will cause the algorithm to be misconfigured. Therefore, it is necessary to perform preprocessing operations on the data set (Fig. 2). Initial processing operations include replacing anomalous values or missing values with the mean value or removing rows with these values, as well as normalization and standardization. Scaling of data features or data normalization in multivariable data is an important step to quickly train machine learning algorithms and to avoid incorrect predictions of the trained model due to outliers in the input data [5,6,7 ,8].

EMA - Exponential moving average, a mathematical method is used to fill the missing values in the data set. EMA is based on giving a larger weight to the values of recent dates. The formula of the EMA method is defined as follows [7]:

$$EMA = (X_i - EMA_{i-1}) * K + EMA_{i-1}, \qquad (1)$$

where X_i – is the value of current timestep, K – smoothing coefficient and it is equal to $\frac{2}{n+1}$, n – is length of the period under consideration.

The value of the EMA depends on the value of the EMA of the previous day. So, the initial value of EMA starts from the particular previous day, it is equal to simple moving (SMA) and is defined as follows:

$$SMA = \frac{X_1 + X_2 + \ldots + X_n}{n}, \qquad (2)$$

The analysis of the characteristics of the load data allows to determine the indicators of short-term and long-term fluctuations in consumption and to choose the appropriate algorithm for their modeling.

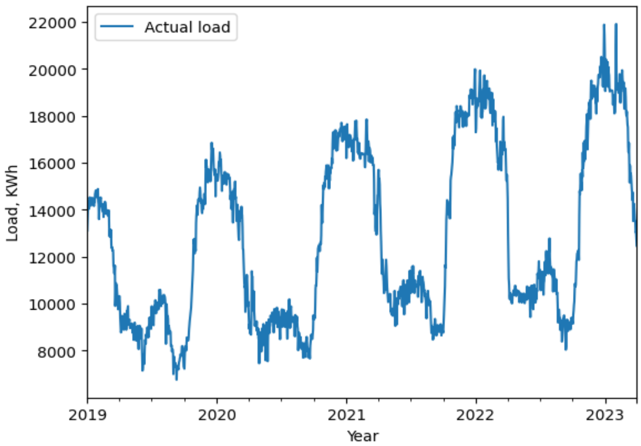

Fig. 1. Daily aggregate load of Andijan region during 2019-2022.

In this work, seasonal consumption characteristics and the trend in the consumption has been analyzed based on real data about the daily aggregate electricity consumption of Andijan region for 4 years. Also, based on these features, future consumption amounts were forecasted using the linear regression method.

3. Methodology. An overview of the proposed system for analyzing electricity consumption data and forecasting future consumption amounts is presented in Figure 1. It starts with data collection, followed by data cleaning and preprocessing steps. Based on the preprocessed data, short-term, seasonal and trend characteristics of consumption has been extracted. According to these features, the training of the linear regression algorithm to model the consumption profile is performed. Hence the model is used to predict future amounts of consumption.

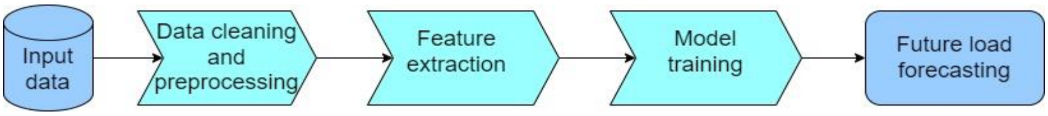

Fig. 2. Overview of load data processing and forecasting system

The slowest change in consumption is the trend, which shows the overall rate of increase or decrease over the entire time period observed. Depending on the characteristics of the data, the trend will have a linear, quadratic, or polynomial graph.

To determine the trend indicator for the considered period, the moving average (MA) method is widely used [9]. Fig. 3 shows the trend in consumption using MA. A linear regression algorithm is trained using available consumption data. The prediction of the trained model represents the growth of consumption for the entire period, and it can be seen from fig.4 that it is close to the trend graph obtained on the basis of MA.

The linear regression method used to determine the trend in consumption is determined based on the following formula:

$$f_a(x) = a_1 x + a_0 , \tag{3}$$

where a_1 and a_0 are the coefficients of the regression function, x is the trend feature of consumption, $f_a(x)$ is the amount of consumption on an appropriate day. The trend feature is extracted from the general consumption data and is transmitted to the algorithm as input data. The calculation result of the algorithm $f_a(x)$ is compared to the actual amount of daily consumption, that is. The motivation of the algorithm is to minimize the difference of these two values as much as possible. Thus the main purpose of linear regression is to optimize indicators a_1 and a_0. For this, it minimizes cost function, that is the mean squared error between the actual value of consumption y_i and the estimated value $f_i(x)$ of the model on each timestep by updating a_1 and a_0 in each iteration.

$$J = \frac{1}{n}\sum_{i=1}^{n}(f_i(x) - y_i)^2 , \tag{4}$$

$$minimize \ \frac{1}{n}\sum_{i=1}^{n}(f_i(x) - y_i)^2 , \tag{5}$$

After each iteration, the algorithm sets new values of a_1 and a_0 to minimize the error. The Gradient Descent method is widely used to determine the new value of variables [10].

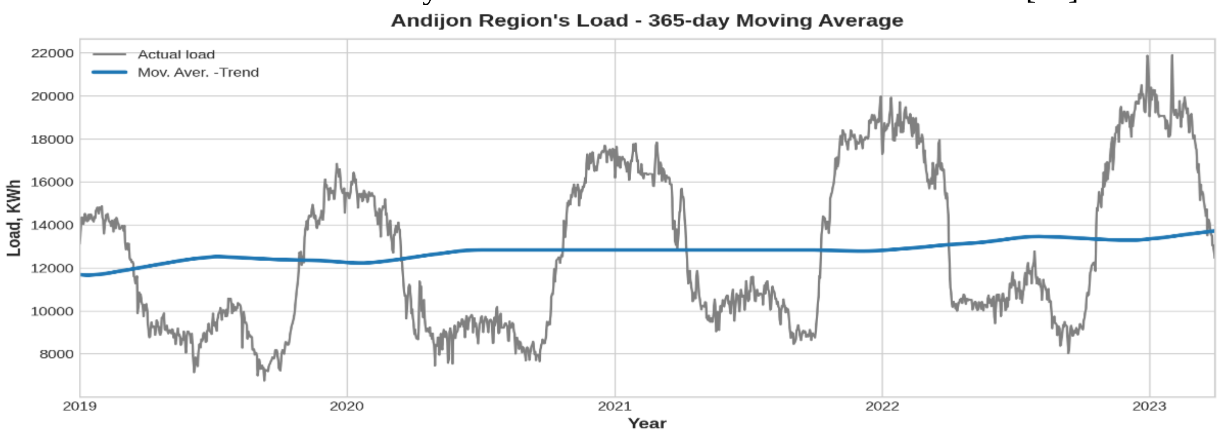

Fig. 3. A graph of trend estimated using the moving average method.

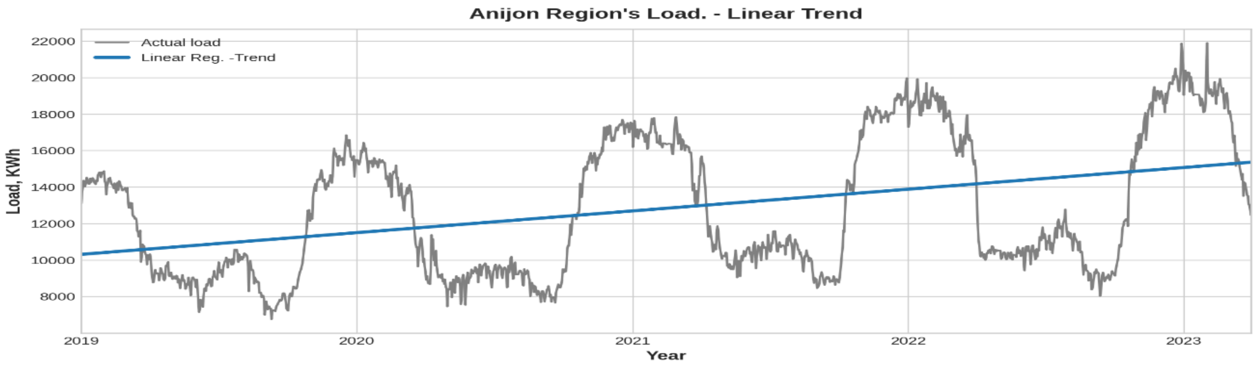

Fig. 4. A graph of trend estimated using a model of linear regression.

In order to fully generalize consumption characteristics, it is necessary to extract the characteristics of annual seasonal fluctuations and model them using a multivariable regression algorithm. The Fourier features are used as the input variable to the algorithm which allow us to construct a seasonal consumption graph fig.5. Fourier features are a set of pairs of sine and cosine functions, with the help of their combination, a complex consumption graph can be made [11].

4. Performance analysis. The increase in consumption for the observed total time is determined using MA in Figure 3, respectively, its model using linear regression is represented in Figure 4. From both graphs, it can be seen that the volume of consumption increased linearly during the entire 4 years.

Figure 5 shows the seasonality of the consumption graph. The graph basically has two seasonal growth characteristics during the year, winter and summer. These seasonal fluctuations were modeled by linear regression based on a combination of Fourier features. As a result, a prediction model that summarizes the seasonal characteristics of consumption was achieved in Fig. 6. The data set is divided into train and test sets to train and test the model. The training set consists of daily consumption data from 01.01.2019 to 31.01.2022. The data from 01.01.2023 to 03.31.2023 which is unseen to the algorithm during the training phase is used to test the performance of the model. In Fig. 6, the actual values of consumption were represented with black dots, where predicted values the model based on the training data is shown with the blue curve, and the forecasted values of the model based on the test data with the red curve. It can be concluded from the graph that the linear regression model achieves a good generalization of consumption characteristics. To check the prediction accuracy of the model, the widely used Mean Absolute Percentage Error (MAPE) statistical indicator was used. According to [12,13], MAPE is the average value of daily percentage errors for the full observed period. The average overall percent error of the proposed model was 5.8 % for the 48 months daily data , which is used to train the model, while the lower error 4.75% was achieved for the test data set of three months.

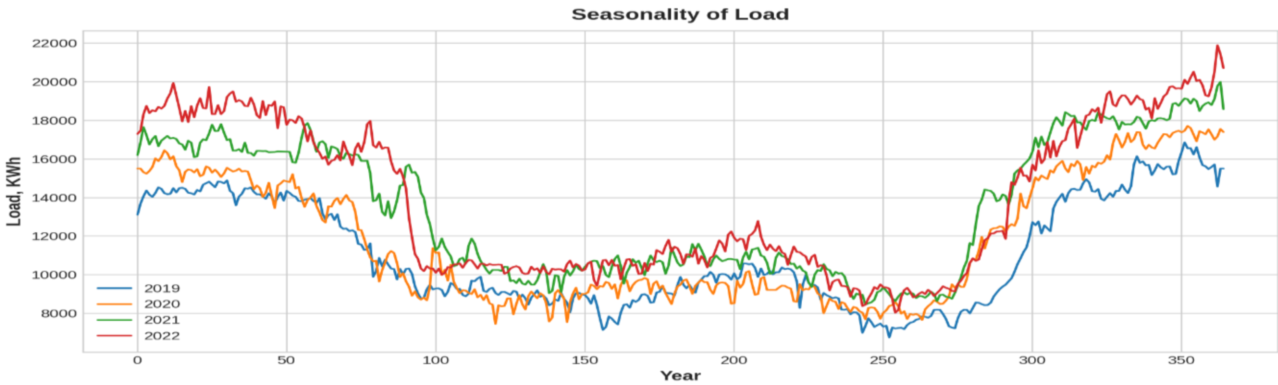

Fig.5. Yearly seasonal fluctuations of the load

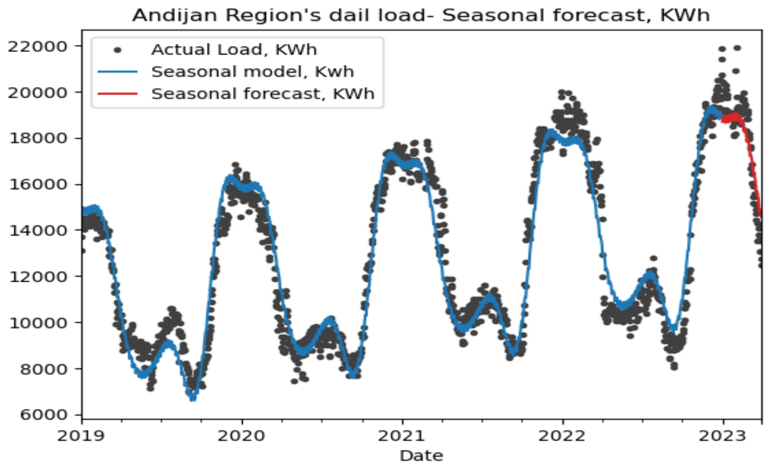

Fig. 6. Forecasted amount of load according to seasonal features.

5. Conclusion. In this work, the information about the daily electricity consumption of the Andijan region, Uzbekistan for 4 years was analyzed using a machine learning algorithm, and the trend in consumption for the overall period, as well as the characteristics of seasonality, were studied. Using the linear regression method of machine learning, seasonal fluctuations in consumption were modeled as a combination of Fourier features. The available real data was divided into two parts, i.e. training and testing set. The training of the algorithm was performed using the train set, and the performance of the resulting model was estimated based on the MAPE statistical indicator on the test data set. As a result, the average percentage error rate of the model was 4.75%. In conclusion, it can be said that the linear regression method can be successfully used to predict future amounts of electricity load of the consumers.

References

1. Zhang D., Xin M., Lui L., et al.: ‘Research on development strategy for smart grid big data’, Proc. CSEE, 2015, 35, (1), 2-12
2. Lee, C. M., Ko, C. N.: ‘Short-term load forecasting using lifting scheme and ARIMA models’, Expert Syst. Appl., 2011, 38, (5), pp. 5902– 5911.
3. Alberg, D., Last, M.: ‘Short-term load forecasting in smart meters with sliding window-based ARIMA algorithms’, Intell. Inf. Database Syst., 2017, pp. 299– 307.
4. Tang, Xianlun, et al. "Short-term power load forecasting based on multi-layer bidirectional recurrent neural network." IET Generation, Transmission & Distribution 13.17 (2019): 3847-3854.
5. Mahjoub, Sameh, et al. "Predicting Energy Consumption Using LSTM, Multi-Layer GRU and Drop-GRU Neural Networks." Sensors 22.11 (2022): 4062.
6. Gal MS, Rubinfeld DL. Data standardization. NYUL Rev.. 2019;94:737.
7. Mahjoub, Sameh, Larbi Chrifi-Alaoui, Bruno Marhic, and Laurent Delahoche. "Predicting Energy Consumption Using LSTM, Multi-Layer GRU and Drop-GRU Neural Networks." Sensors 22, no. 11 (2022): 4062.
8. V. N. G. Raju, K. P. Lakshmi, V. M. Jain, A. Kalidindi and V. Padma, "Study the Influence of Normalization/Transformation process on the Accuracy of Supervised Classification," 2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT), Tirunelveli, India, 2020, pp. 729-735, doi: 10.1109/ICSSIT48917.2020.9214160.
9. Lauren, S., & Harlili, S. D. (2014, August). Stock trend prediction using simple moving average supported by news classification. In 2014 International Conference of Advanced Informatics: Concept, Theory and Application (ICAICTA) (pp. 135-139). IEEE.
10. Ruder, Sebastian. "An overview of gradient descent optimization algorithms." arXiv preprint arXiv:1609.04747 (2016).
11. Tompkins, Anthony, and Fabio Ramos. "Fourier feature approximations for periodic kernels in time-series modelling." Proceedings of the AAAI Conference on Artificial Intelligence. Vol. 32. No. 1. 2018