INOTECH
Inspiring Technologies and Innovations

Inspiring Technologies and Innovations

https://dergipark.org.tr/tr/pub/inotech

Research Article

Efficiency of Ensemble Learning Algorithms for Trend Analysis in Electricity Consumption in Turkey during Covid-19 Pandemic

Selim BUYRUKOĞLU[a], Ayhan AKBAŞ[b]

[a]Çankırı Karatekin University, Faculty of Engineering, Computer Engineering Department, Çankırı, Turkey
[b]Abdullah Gul University, Faculty of Engineering, Computer Engineering Department, Kayseri, Turkey
ORCID[a]: 0000-0001-7844-3168, ORCID[b]: 0000-0002-6425-104X,

Corresponding Author e-mail: ayhan.akbas@gmail.com

Received	**: 16.03.2022**	**Accepted**	**: 25.05.2022**	**Pages**	**: 9-15**

ABSTRACT:

Aim: We aim to analyze overall electricity consumption in Turkey starting from pre-COVID days until today to illustrate the illustrate efficiency of machine learning algorithms in trend changes.

Design & Methodology: We built machine learning models for the analysis such as AdaBoost (boosting), Random Forest (bagging) and Deep Neural Network (single-based algorithm).

Originality: The originality of this study is the determination of ensemble learning algorithms in the Analysis of Effects of Covid-19 Pandemic on Electricity Consumption in Turkey

Findings: Findings revealed that the proposed boosting (AdaBoost) ensemble algorithm (RMSE: 41848.7, MAE: 18574.3, R2 :0.89) is a significant contributory factor in the analysis of data related to electricity consumption.

Conclusion: As a conclusion, boosting (AdaBoost) ensemble learning algorithm is more preferable in the use of energy-related data than the bagging (random forest) and single-based algorithms (deep neural networks).

KEYWORDS: Ensemble Learning Algorithms, Adaboost, Electricity Consumption, Covid-19 Pandemic

1. INTRODUCTION

Machine Learning (ML) techniques have become increasingly popular in recent years. Machine learning provides a wide spectrum of uses in various industries, such as credit scoring in financial institutions, cancer diagnosis and pharmaceutical development, pattern recognition, audio, video, and image analysis, and electricity forecasting. Machine learning approaches have been used by people from numerous disciplines to address problems more efficiently and quickly (Nallathambi & Ramasamy, 2017). Researchers use machine learning for data processing solutions to build, optimize, and deploy effective prediction models that improve outcomes, and they rely heavily on machine learning for classification or prediction to analyze outcomes and pick up the best classification. With advantages that machine learning provided, ML has become popular in the energy community as well, as a way to forecast future consumption and market prices and has been widely used in different regression and classification problems (Sarker, 2021) (Viloria, Naveda, Palma, Núñez, & Núñez, 2020)(Tapas Ranjan Jena, Swati Sucharita Barik, 2020). A variety of machine learning algorithms have been proposed for the electricity consumption problems (González-Briones, Hernandez, Corchado, Omatu, & Mohamad, 2019) (Shaikh & Namdeo, 2021)(Sen, Tunç, & Günay, 2021).

Despite the fact that the offered research achieves impressive findings in terms of their objectives (Bashawyah & Qaisar, 2021),(Shaikh & Namdeo, 2021) the efficiency of ensemble learning and single-based algorithms for energy consumption problems has never been investigated. The objective of this research is to study and illustrate how powerful ensemble learning and single-based algorithms are for power consumption analysis.

2. MATERIALS AND METHODS

This section provides information about the proposed prediction approach for energy consumption in Turkey. In this study, the data is split into training (80%) and test (20%) sets. Then, 10-fold cross-validation is used in the training process of the AdaBoost, Random Forest, and Deep Neural Network algorithms. Python programming language was used in the implementation of these algorithms. Finally, the performances of the models are compared based on the test set.

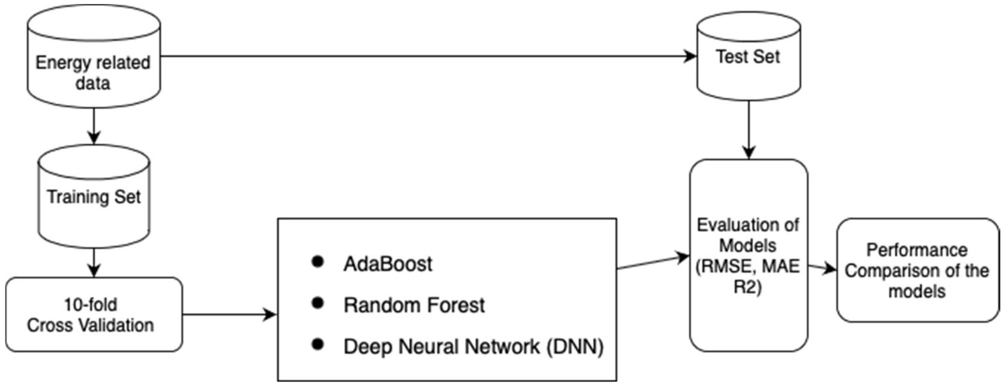

Figure 1. Prediction approach for the electricity consumption

2.1. Dataset

All energy-related data is collected from the Turkish Energy Transparency Platform (EPIAS) in Turkey. This platform provides the data essential for energy markets to operate in a transparent, reliable, fair, and predictable manner. EPIAS requests published data from institutions based on laws enacted by the Energy Market Regulatory Board. The relevant data owner companies are required to disclose the data on the Transparency Platform in a timely fashion, in the prescribed format, and with the correct content. Data is published on the web publicly for all parties to use.

Extracted data for hourly-based electricity consumption in Turkey belongs to all user profiles, including lighting, household, industry, irrigation, and commercial profiles., in MWh from EPIAS (EPIAS, n.d.). The data starts from January 1, 2019, to January 31, 2022, during which the COVID-19 Pandemic was in effect. The extracted data was further processed into daily data and summed up to get the overall electricity consumption of Turkey in MWh. All the analysis carried out has been performed using this EPIAS data.

2.2. Machine Learning Algorithms

Deep Neural Network

Deep neural network (DNN) is one of the popular machine learning algorithms. It can be used for both classification and regression problems (Fu, et al. 2020). The structure of any DNN consists of input, output, and at least one hidden layer (Buyrukoğlu, et al. 2021). In the study, different structures were tried such as DNN structures that were created with one, two, and three hidden layers for energy consumption. In the end, DNN created with one hidden layer performs better compared to the others. Therefore, one input, hidden, and output layers were used in the creation of the DNN as seen in Figure 2. Then, two, four, and six neurons were used in the hidden layer, and the best statistical score was obtained through the 4 neurons. Moreover, a variety number of epochs were used to create a robust and reliable DNN algorithm. In this sense, different epoch values were tried such as 50, 100,150, 200, 250, 300, 350, 400, 450, and 500. The maximum number of epochs was set to 400 because the best performance was obtained in the use of 400 epochs.

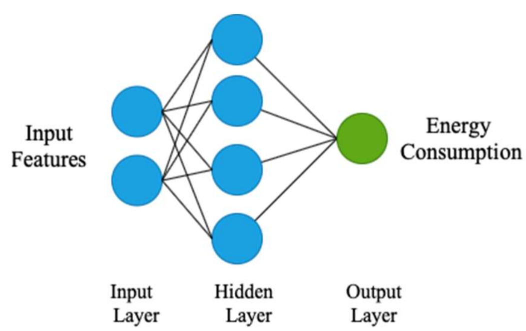

Figure 2. Structure of the Employed Neural Network

AdaBoost

AdaBoost is one of the popular boosting ensemble learning algorithms. Weak (simple) learners are used in the AdaBoost for classification or regression problems. In this sense, each weak learner is used in each estimator, and then a weight coefficient is assigned to this learner. This weighting coefficient is inversely proportional to the weak learner error. Finally, this model provides a solution based on weighted voting (Qu et al., 2021). A case study was also carried out to determine the optimal number of estimators in the energy consumption process such as 10, 20, 30, 40, 50, 60, 70, 80, 90, and 100. In our study, 80 estimators were used in the proposed AdaBoost algorithm.

Random Forest

Bagging ensemble learning algorithm consists of two algorithms including decision tree and random forest (Fayaz, n.d.). The Random Forest algorithm is considered a more effective algorithm than the decision tree. The reason behind this is that the random forest algorithm consists of more than one decision tree. Average estimation of decision trees is used in the random forest for the prediction or classification problems (Nallathambi & Ramasamy, 2017). In our study, different numbers of trees were tried to determine the optimal number of trees such as 50, 100,150, 200, 250, 300, 350, 400, 450, and 500. In the end, the most beneficial statistical score is obtained through the 500 trees.

2.3. Evaluation Scores

In this study, MSE, MAE, and R2 statistical metrics are used to evaluate the employed algorithms in the energy consumption analysis. Equation 1, 2, and 3 gives the mathematical formula of these metrics (Chicco, D., Warrens, M. J., & Jurman, 2021).

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(P_i - \hat{P}_i\right)^2 \tag{1}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|P_i - \hat{P}_i\right| \tag{2}$$

$$R^2 = 1 - \frac{Unexplained\ Variation}{Total\ Variation} \tag{3}$$

where P_i is the actual value, $\hat{P}_i$ is the predicted value from the model and n is the number of observations.

3. RESULTS AND DISCUSSION

3.1. Prediction Results of Machine Learning Algorithms

Table 1 compares the results obtained from the employed machine learning algorithms for energy consumption. From the data in Table 1, it is apparent that the AdaBoost ensemble learning algorithm performs better than the other algorithms. Deep Neural Network (DNN) has the second-best RMSE, MAE, and R2 value for energy consumption. On the other hand, the random forest has the lowest statistical scores. These findings help us to highlight that Boosting ensemble learning algorithm (AdaBoost) provides significantly better results than the DNN and bagging algorithm (Random Forest). It means that boosting ensemble learning can be considered helpful for the prediction of energy consumption.

Table 1. Comparison of machine learning algorithms for the electricity consumption in Turkey

Model	RMSE	MAE	R2
AdaBoost (80 Estimators)	*41848.7*	*18574.3*	*0.890*
Random Forest (500 Trees)	*58253.4*	*23465.8*	*0.807*
Deep Neural Network (400 epochs)	*47456.4*	*19267.3*	*0.872*

3.2. The impact of different epochs in DNN

A case study was carried out to determine the optimal epoch number to be used in DNN for energy consumption. Figure 3 illustrates the MAE scores of DNN based on the use of the different number of epochs. The peak MAE score of DNN for energy consumption is 23858 (MWh) in the use of 50 epochs. In contrast to this, the use of 400 epochs in this DNN enables to reach the lowest MAE score (19267,3). Moreover, the MAE score is not changed during the use of 400, 450, and 500 epochs. Thus, it can be inferred from the results that the use of 400 epochs in this DNN has a significant impact on energy consumption.

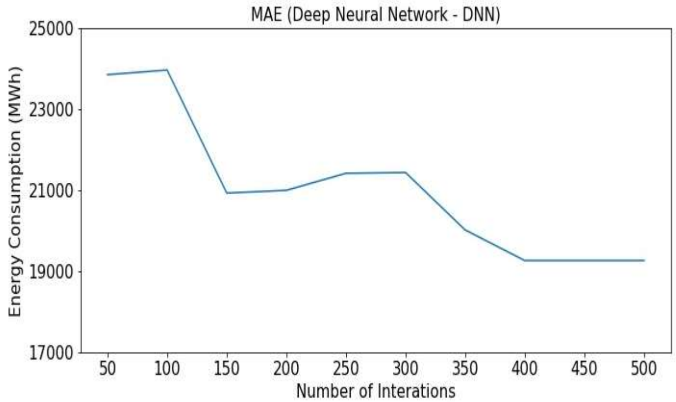

Figure 3. MAE scores of DNN based on different numbers of epochs

3.3. The impact of different iterations in AdaBoost

AdaBoost ensemble learning algorithm is one of the effective boosting algorithms as highlighted in Section 2.3. Figure 4 presents the MAE scores of AdaBoost based on the different number of estimators. What can be clearly seen in this figure is the highest MAE score (39958) of the AdaBoost algorithm is obtained with the use of 10 estimators. The score of MAE dropped sharply until 40 estimators (from 10 to 40 estimators). Then, the MAE score slightly decreased between the use of 40 and 80 estimators. In the end, the AdaBoost algorithm provided the best MAE score when the use of 80 estimators. Even if the MAE score remains the same when the use of 80, 90, and 100 estimators, the optimal estimator number of the AdaBoost algorithm is determined as 80. Thus, this proposed algorithm may provide more benefit in terms of time and cost in the prediction of energy consumption.

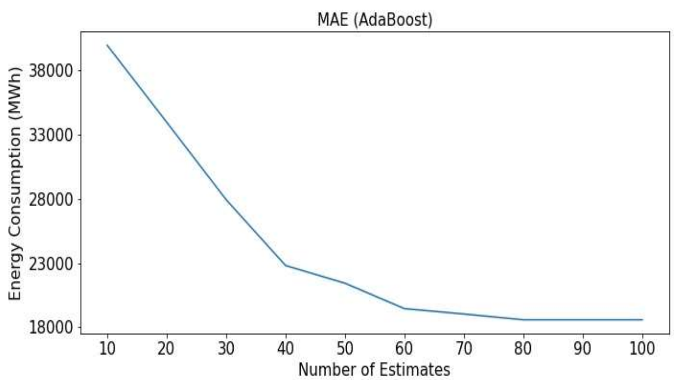

Figure 4. MAE scores of AdaBoost based on different number of estimators

3.4. The impact of different trees in Random Forest

Figure 5 shows the MAE scores of Random Forest based on the different number of trees. The MAE score (24944) rose to the highest point in the use of 200 trees. The MAE score fell to the lowest point of 23465,8 when the use of 500 trees. What is interesting in Figure 4 is the rapid decrease of MAE score between the use of 400 and 500 trees. The reason behind this is that the difference between the MAE scores is the use of 400 and 500 trees is 1049,4. In the end, the number of 500 trees is accepted as the optimal value for the proposed random forest algorithm in the energy consumption prediction.

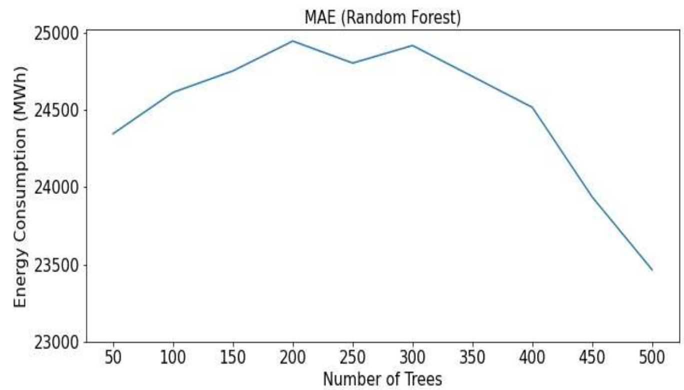

Figure 5. MAE scores of Random Forest based on different numbers of trees

3.5. Discussion of Electricity Consumption before and after Covid-19

In Figure 6, when total monthly electricity consumption was examined, it was discovered that consumption fell sharply during the lockdown period until May 2020, when average monthly electricity consumption peaked. The lowest value recorded was 26399 MWh in May 2020, after which electricity consumption began to rise as restrictions were relaxed, reaching a maximum of 43181 MWh once all industries resumed production.

Starting with the first announcement of the pandemic in March'2019, electricity consumption dropped below 1.000.000 MWh and fluctuated between 950.000 and 1.000.000 MWh due to frequent closures and restricted working hours of businesses. On June 1, 2019, due to the tourism season, restrictions have been loosened that burst electricity demand again. In the spring of 2020, lockdowns were effective in the decrease of consumption. With the application of vaccinations to millions of people, restrictions were largely lifted, which gave rise to a gradual increase in electricity demand.

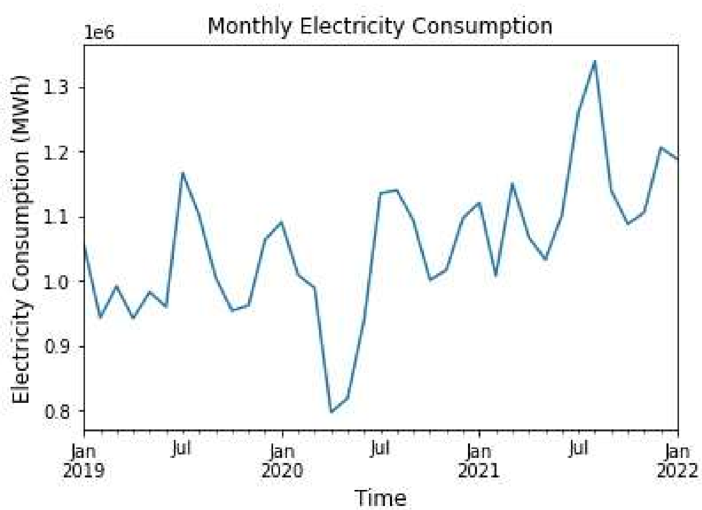

Figure 6. Monthly Average Electricity Consumption (MWh) in Turkey

4. DISCUSSION

This section discusses the performance of the employed algorithms in the electricity consumption and the electricity consumption from pre-COVID days until now, respectively.

As highlighted in Table 1, AdaBoost boosting ensemble learning algorithms achieved to provide the best statistical scores compared to the other employed algorithms including Random Forest and Deep Neural Network. Most of the studies in the literature have focused on comparing the base-learner algorithms with bagging algorithms (González-Briones, et al., 2019). The efficiency of Random Forest and Decision Tree bagging ensemble learning algorithms was compared with k-Nearest Neighbours, Support Vector Regressor, and Linear Regression base learner algorithms using the data belonging to a shoe store located in Salamanca, Spain. The linear regression algorithm provided the best accuracy score (85.7%) while the Random Forest algorithm was providing the lowest accuracy score (79.9%). A different study proposed by Nguyen & Nguyen (2019) predicts energy consumption employing a radial basis function neural network (RBF-NN). Due to the insufficiency of analytical and linear models for energy consumption, RBF-NN was employed in the study. Even if the studies proposed by González-Briones, et al., (2019) and Nguyen & Nguyen (2019) fulfilled their aim and objectives, they do not compare the bagging and boosting ensemble learning algorithms. In contrast to this, in a different study, Pinto et al., (2021) compared the efficiency of three ensemble learning algorithms (AdaBoost, Random Forest, Gradient Boosting) in the energy consumption using a real data from an office building. The AdaBoost ensemble learning algorithm provided the best MAPE (5.34) score than the others. In the presented study, the employed AdaBoost ensemble learning algorithm is also achieved to provide the best statistical score compared to the DNN and RF as shown in Table 1. It can be inferred from the results that the AdaBoost ensemble learning algorithm seems more efficient than the other ensemble and base-learner algorithms in electricity consumption.

In Figures 3, 4, and 5, the MAE scores of the employed algorithms based on the different number of parameters (epochs, estimators, trees) are illustrated. In these Figures, the number of parameter values is set within a certain range. Moreover, a range is fixed for the number of parameters between which the model should be trained to predict the best possible results. In the end, the optimal parameter values were determined for DNN (400 epochs), AdaBoost (80 estimators), and RF (500 trees).

5. CONCLUSION AND FUTURE DIRECTIONS

In this study, the aim is to show the efficiency of ensemble learning (AdaBoost, Random Forest) and single-based (Deep Neural Network) algorithms for the prediction of electricity consumption. The MAE values are 18574.3, 23465.8, 19267.3 for AdaBoost, Random Forest, and Deep Neural Network, respectively. Also, the proposed AdaBoost ensemble model provided less satisfactory RMSE (41848.7) and R2 (0.89) values. These results revealed that ensemble learning algorithms may play a significant role in the prediction problems for electricity consumption. Additionally, the boosting (AdaBoost) ensemble algorithm is a significant contributory factor in the analysis of data related to electricity consumption in comparison to the bagging (random forest) and single-based algorithms (deep neural networks). On the other hand, the electricity consumption rises from July 2021 in the Covid-19 pandemic.

There are still more ML approaches to be investigated for a more efficient way of conducting the analysis of electricity consumption data. Our future research will focus on the combination of statistical feature selection techniques and deep learning algorithms to improve the AdaBoost algorithm for electricity consumption prediction.

Author's Contributions

Data were publicly available. Conceptualization, formal analysis, methodology, and writing – original draft was performed by Selim Buyrukoğlu and Ayhan Akbaş. Resources, software, supervision, writing – review & editing were organized by Selim Buyrukoğlu and Ayhan Akbaş.

Conflict of Interest

The authors claim no conflict of interest to declare.

REFERENCES

Albuquerque, P. C., Cajueiro, D. O., & Rossi, M. D. C. (2022). Machine learning models for forecasting power electricity consumption using a high dimensional dataset. *Expert Systems with Applications*, 187(September 2021). https://doi.org/10.1016/j.eswa.2021.115917

Bashawyah, D. A., & Qaisar, S. M. (2021). Machine Learning Based Short-Term Load Forecasting for Smart Meter Energy Consumption Data in London Households, 99–102. https://doi.org/10.1109/elit53502.2021.9501104

Chicco, D., Warrens, M. J., & Jurman, G. (2021). No Title. *PeerJ Computer Science*, *5*.

Edwards, R. E., New, J., & Parker, L. E. (2012). Predicting future hourly residential electrical consumption: A machine learning case study. *Energy and Buildings*, *49*, 591–603. https://doi.org/10.1016/j.enbuild.2012.03.010

EPIAS. (n.d.). EPIAS. Retrieved February 1, 2022, from https://seffaflik.epias.com.tr/transparency/tuketim/tuketici-bilgisi/tuketim-miktarlari.xhtml

Farahat, A. K., Ghodsi, A., & Kamel, M. S. (2013). Efficient greedy feature selection for unsupervised learning. *Knowledge and Information Systems*, 35(2), 285–310. https://doi.org/10.1007/s10115-012-0538-1

Fayaz, M. (n.d.). Prediction of Energy Consumption in the Buildings Using Multi-Layer Perceptron and Random Forest. *International Journal of Advanced Science and Technology*, 101, 13–22.

Fu, Q., Liu, Q. S., Gao, Z., Wu, H., Fu, B., & Chen, J. (2020). A Building Energy Consumption Prediction Method Based on Integration of a Deep Neural Network and Transfer Reinforcement Learning. *International Journal of Pattern Recognition and Artificial Intelligence*, 34(10), 1–18. https://doi.org/10.1142/S0218001420520059

González-Briones, A., Hernandez, G., Corchado, J. M., Omatu, S., & Mohamad, M. S. (2019). Machine Learning Models for Electricity Consumption Forecasting: A Review. *2nd International Conference on Computer Applications and Information Security, ICCAIS 2019*. https://doi.org/10.1109/CAIS.2019.8769508

Nallathambi, S., & Ramasamy, K. (2017). Prediction of electricity consumption based on DT and RF: An application on USA country power consumption. *Proceedings - 2017 IEEE International Conference on Electrical, Instrumentation and Communication Engineering, ICEICE 2017*, *2017-Decem* (1), 1–7. https://doi.org/10.1109/ICEICE.2017.8191939

Nguyen, D. H., & Tung Nguyen, A. (2019). A Machine Learning-based Approach for the Prediction of Electricity Consumption. *2019 12th Asian Control*

Conference, ASCC 2019, 1301–1306.

Qu, Z., Liu, H., Wang, Z., Xu, J., Zhang, P., & Zeng, H. (2021). Energy & Buildings A combined genetic optimization with AdaBoost ensemble model for anomaly detection in buildings electricity consumption. *Energy & Buildings*, *248*, 111193. https://doi.org/10.1016/j.enbuild.2021.111193

Rajasekar, M., & Geetha, A. (2022). Comparison of Machine Learning Methods for Tamil Morphological Analyzer. *Lecture Notes in Networks and Systems*, *213*, 385–399. https://doi.org/10.1007/978-981-16-2422-3_31

Sarker, I. H. (2021). Machine Learning: Algorithms, Real-World Applications and Research Directions. *SN Computer Science*, 2(3), 160. https://doi.org/10.1007/s42979-021-00592-x

Sen, D., Tunç, K. M. M., & Günay, M. E. (2021). Forecasting electricity consumption of OECD countries: A global machine learning modeling approach. *Utilities Policy*, 70(November 2020), 101222. https://doi.org/10.1016/j.jup.2021.101222

Shaikh, A., & Namdeo, V. (2021). Unerstanding Machine Learning Approach on Various Algorithms: A Case Study Implementation. *2021 6th International Conference for Convergence in Technology, I2CT 2021*, 2–6. https://doi.org/10.1109/I2CT51068.2021.9418166

Tapas Ranjan Jena, Swati Sucharita Barik, S. K. N. (2020). Electricity Consumption & Prediction using Machine Learning Models Tapas Ranjan Jena Regn . No . -180303110001 , Department of Computer Science and Engineering , Centurion University of Technology and Management , Odisha , India Swati Sucharita Barik Assi, IX(180303110001), 2804–2818.

Viloria, A., Naveda, A. S., Palma, H. H., Núñez, W. N., & Núñez, L. N. (2020). Retraction: Electrical consumption patterns through machine learning (Journal of Physics: Conference Series 1432 (012093) DOI: 10.1088/1742-6596/1432/1/012093). *Journal of Physics: Conference Series*, 1432(1). https://doi.org/10.1088/1742-6596/1432/1/012108