Electricity Price Forecasting based on XGBooST and ARIMA Algorithms

Kübra Bitirgen[1], Ümmühan Başaran Filik[2]

Abstract—Accurate electricity price information is critical for wholesale electricity markets. Machine learning algorithms are thought to be most efficient methods as they successfully observe the dependencies between electricity price, historical data and other factors. In this study, two machine-learning models are purposed for electricity price forecasting. The historical prices and other important factors are processed. Then, the future values of the electricity prices are forecasted using properly fitted XGBoost and ARIMA models. To validate the results, some statistical error measurement methods are selected. Consequently, when comparing the results in terms of performance in detail, XGBoost model has become more efficient as the computation speed and lowest error.

Index Terms—Forecasting, XGBoost, ARIMA, Electricity Price

I INTRODUCTION

The deregulation of competitive markets arrange traditionally government-controlled and monopolistic power sectors. By using derivative and spot contracts, electricity is traded. On the other hand, electricity is highly particular commodity. Electricity is also an important integral part of economic and technological industry [1]. From its generation to distribution, accurate information is required. From its generation to distribution, accurate information is required. Nevertheless, it depends on multiple volatile factors. For this reason, forecasting is a popular issue in this field. To forecast electricity generation, many studies are conducted such as estimation of PV generation [2], wind speed [3], hydropower generation [4] etc. As well as the generation of electricity, electricity price is non-storable and requires a stability between generation and consumption. Therefore, electricity demand belongs to daily activities (weekends, holidays etc.). Furthermore, weather conditions, historical prices, fuel prices and many other factors are significant for accurate computation of electricity price. Because of these reasons, these special properties cause to unobservable price results, having seasonality at the various time levels as daily, weekly and yearly it has also typically unexpected price spikes.

However, the forecasting studies supply fundamental inputs for both market operators and its participants in wholesale electricity markets. Accurate forecasting results have significant economic advantages. Market participants used outputs of forecasting for bidding decision-making, portfolio allocation, and investment planning. They are benefited from electricity price forecasting outputs to calculate measurements and variable indexes to monitore the markets. Short-term electricity price forecasting is generally used by market participants to maximize their profits. Medium term electricity price forecasting is preferred in reciprocal contracts in this way it enables to avoid from volatility risk of electricity prices. Long term electricity price forecasting is mandatory for investors, decision makers and it is used for ensuring development of generation, transmission and distribution investments planning.

The main techniques of electricity price forecasting are divided into some different categories that are simulation methods [5], equilibrium analysis [6], volatility analysis [7], inteligent systems [8], time series [9], and econometric methods [10]. While some studies mainly address day-ahead [11] or spot [12, 13] electricity price forecasting, some studies focus on the long-term forecasting [14,15] or even observing results between long-term and short term [16].

In this paper, ARIMA and XGBoost models are implemented for day ahead electricity price forecasting. ARIMA model is generally prefered to forecast day ahead electricity prices and load with effective results. This model is the most commonly used traditional statistical model using effective features to forecast the price. It also processes training data to construct a linear relationship. As in [9], it is stated that composit models have the ability of better forecasting comparing with ARIMA. In [17], two ARIMA models are proposed for forecasting hourly eletricity prices for California and Spain electricity markets. Average errors are around 5%

[1]*Corresponding author mail: kubrabitirgen@eskisehir.edu.tr (https://orcid.org/0000-0002-4468-4905)
Electrical and Electronics Engineering, Eskişehir Technical University, Eskişehir
[2] Second author mail: ubasaran@eskisehir.edu.tr (https://orcid.org/0000-0002-0715-821X)
Electrical and Electronics Engineering, Eskişehir Technical University, Eskişehir

and 10% for the Californian and the Spanish market. It is also mentioned that using an additional variable does not effectively improve accuracy results. In [18], ARIMA model is implemented to make a prediction of natural gas consumption. According to the results of this study, the model outperforms other time-series models. ARIMA generally has effective outputs for producers and consumers as mentioned in many studies.

The other selected model is Extreme Gradient Boosting called XGBoost. There is a study inroducing XGBoost for load forecasting [19]. To make an effective prediction, load factor is standardized and a feature map is established. According to the results of the study, the performance of the XGBoost model is validated. In [20], authors worked on short-term load forecasting and found that highly efficient memory usage and computing time. Based on statistical error models measuring performance, the proposed model outperformed other schemes. XGBoost and LSTM are combined to forecast electricity price in [21]. Authors used the error reciprocal method to make a combination of two models results. The error of this combined model is computed as 0.57 % and obviously, it is better than the single model.

However, different prediction approaches cannot outperform each other accurately because of the challenging characteristics of electricity prices, which are seasonality, sharp price spikes and high volatility [22]. The main aim of this study is to make a comparison between two machine-learning algorithms. Furthermore, verification of practicality and feasibility of these algorithms for electricity price forecasting are identified. The novel and fundamental contribution of this paper is to make a performance comparison of ARIMA and XGBoost based on some significant effective factors on electricity price.

The remaing parts of the paper are organized as follows. Section II defines background information and the methods used for forecasting of hourly electricity prices. Section III presents the results of the selected forecasting models based on ARIMA and XGBoost. Finally, Section IV shows the conclusion.

II METHODOLOGY

In this study, electricity price data was collected for one year and it is obtained from NYISO website [23]. The dataset covers hourly historical electricity prices, coal prices, natural gas and radyoactive mineral U prices for one year [24-25]. It also includes weekends and national holidays. Electricity price forecasting has multiple required steps. In addition, this study starts with explaining the aim, obtaining necessary data, analysing the data, applying pre-process stage, performing the selected forecasting models, comparing and evaluating the selected models' results.

However, identifying a single relation between future data and historical data is significant to obtain accurate forecasting results. Thus, it is possible to see fluctuations in the data and it causes error which points out not a good performance of studied forecasting model. For this study, there are some fluctuations in the historical electricity price data. The results of this forecasting have electricity price spikes due to some reasons. According to the study [26], yearly oil future contracts volatility is around 30%; for future contracts of natural gas, it is around 50%. In addition, it is about 60% for future contracts of electricity. Yearly volatility is above 200% in electricity spot markets. Obtaining an accurate result for the electricity market is not an easy issue due to the importance of volatility. By the reason of the fact that, accuracy of load forecasting is higher than that of electricity price forecasting. The main reason for the spikes is that the electricity price is highly volatile. Some other reasons are in the following:

- Volatility in fuel price
- Outages meaning production uncertainty
- Load uncertainty
- Market manipulation (market power, counterparty risk)
- Congestion of transmission
- Hydroelectricity production fluctuations
- Market participant behavior

Moreover, to this, electricity price series includes complex features such as nonlinearity, nonstationarity, and high volatility as explained above. The most significant parts of the original price series are these features and they should be taken into account. The features considered in this study can be seen in figure 1 and figure 2.

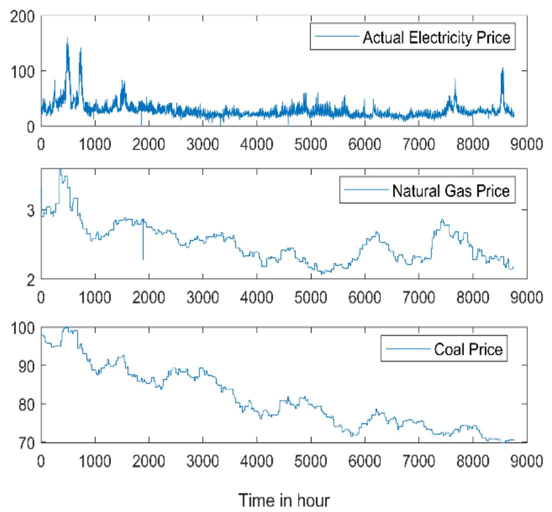

Figure 1 Features incliding prices of actual electricity, natural gas and coal in $

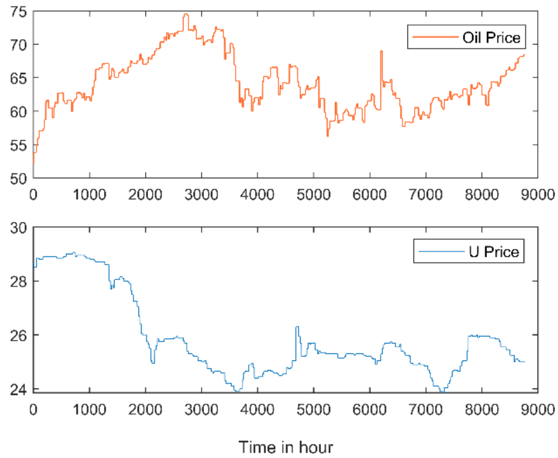

Figure 2 Features including prices of oil and Uranium in $

A XGBoost

In [27], XGBoost is proposed and its algorithm depends on the Regression Tree and Classification. Furthermore, this model redefines the partition attributes, and uses the loss function minimization to indicate the partition attributes.

However, it is necessary to compute the feature importance in selected dataset. It helps for selection the most suitable features. In addition, features, which have the most important values for training and testing, are considered. The data is accurately divided into training data and testing data. In the training process, 75% training data is effectively. Testing process is performed by using 25% testing data. The stages of electricity price forecasting using the XGBoost algorithm can be seen in following Table 1.

TABLE 1

Algorithm of XGBoost for electricity price forecasting [20]

	XGBoost
1.Step	Explore historical data
2.Step	for i ← 1 to size(features) do
3.Step	Computation of feature importance
4.Step	End for
5.Step	Selection of features with importance value greater then the threshold value
6.Step	Separate data into testing and training data
7.Step	Training the model over training data
8.Step	Forecasting hourly electricity price using trained model over testing data

B ARIMA

ARIMA is a suitable method for forecasting future values from historical data. The significant advantage of ARIMA is that time series data is neccessary for forecasting. By virtue of the fact that, this property is advantageous in studies having a large number of time series.

The algorithm starts with the following calculation and steps and is clearly predefined in [28]. It is also defined by (1):

$$x_t = \phi_0 + \phi_1 x_{t-1} + \phi_2 x_{t-2} + \ldots + \phi_p x_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \ldots - \theta_q \varepsilon_{t-q} \quad (1)$$

Where $\phi_i (i = 0,1, \ldots, p)$ and $\theta_j (j = 0,1, \ldots, q)$ are model parameters. p and q are orders of this algorithm. Normal distribution is followed by ε_t called the random error. ARIMA model has four fundamental Steps:

Step 1: Identification of the model

Step 2: Estimation of parameters

Step 3: Recognition of the model

Step 4: Verifying the model and forecasting.

The ARIMA model working principle is shown in the following:

TABLE 2

Algorithm of ARIMA for electricity price forecasting

	ARIMA
1.Step	Explore historical data and model identification
2.Step	Clarification of data
3. Step	Estimation of parameters
4.Step	Validity test
5.Step	Refine the model
6.Step	Forecast the future data

C Evaluation Methods

As defined above, in the first pattern, some careful considerations and preprocessing steps are taken with respect to the models. After that, two different approaches are successfully implemented. Finally, in order to observe the performance of each model, three suitable methods are selected: Mean absolute error (MAE), Mean Absolute Percentage Error

(MAPE), and Root Mean Square Error (RMSE) [29]. The formulations of the selected performance measurement methods are as follows:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - t_i| \tag{1}$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\frac{|y_i - t_i|}{y_i}X100\% \tag{2}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - t_i)^2} \tag{3}$$

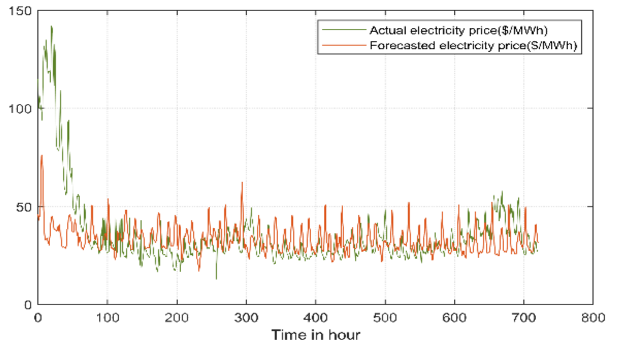

Figure 4 Actual electricity prices and XGBoost forecasting results for a month

III RESULTS

In this Section, the forecasting results of ARIMA and XGBoots are presented. Electricity price forecasting is obtained according to the models described in Section II. Note that each model is generally associated with some features of the electricity price. Therefore, the features play important roles on the electricity price forecasting. To strengthen the models, it is required to obtain more data associated with the electricity price. Figure 3 shows the actual and forecasted value of electricity price.

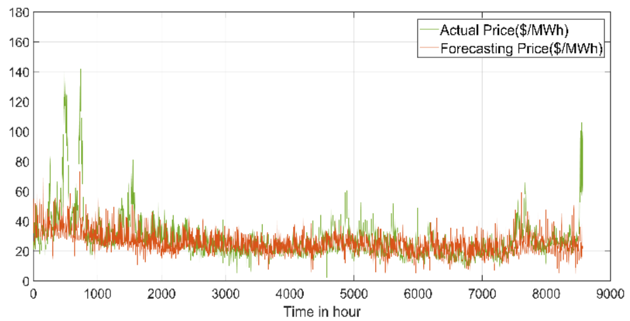

Figure 3 Annual actual electricity prices and XGBoost forecasting results in hour

It is clear from this Figure that the optimal model is XGBoost in the forecasting. Figure 3 shows the actual and forecasted value of electricity price. It shows the forecasted value is very close to the actual value. There is some undesired spikes in the input data and it cannot catch this spikes effectively. This result shows that these spikes depends on some other volatile factors like wheather conditions or load data.

As shown the above Figure, this model performs well during steady conditions. For better determination of the results and understanding the reasons of difference between forecasted and actual data, the result can be observed in more limited time like a month or a week.

It is noticed that less model parameters defined by user are required in the algorithm of XGBoost that is convenient for data analysis that is large scale. The reason of the fact that it gives an advantage for calculation time and memory of the system. Annual actual electricity prices and ARIMA forecasting results in hour is given in Figure 5.

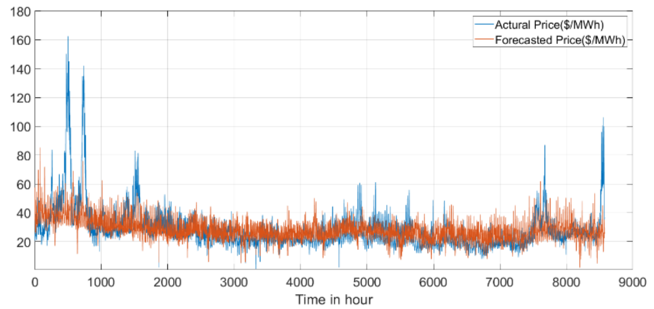

Figure 5 Annual actual electricity prices and ARIMA forecasting results in hour

It is anticipated that the ARIMA model can be used as an effective electricity price-forecasting model as well as the XGBoost model. However, ARIMA does not fully follows the actual electricity price as can be seen in figure 5, so there still can improve the accuracy of this algorithm for its forecasting. ARIMA model can be combined with another model to solve this above problem. Actual electricity prices and ARIMA forecasting results for a month is presented in Figure 6.

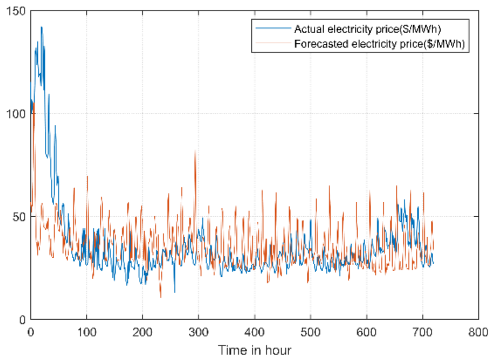

Figure 6 Actual electricity prices and ARIMA forecasting results for a month

The difference of the performance of both models is recognized by observing the results graphically. It is found that ARIMA is higher results. From figure 6, we can conclude that ARIMA model have a strength when tracking of the actual value in stationary conditions but it has also a weakness due to high difference from the actual price.

Both models results show the forecasted and actual value with large spike in electricity prices. These machine-learning models are unable to track the sudden and large spikes in electricity price like the small spike. However, the main reason of unabling the spikes is as explained in Section II. Normally, machine-learning models perform well but there is an exception at the peak duration. The main problem of electricity price forecasting is that the forecasting error during spikes.
Based on the performance of these selected models for the electricity price forecasting, XGBoost gives more efficient than output in comparison with ARIMA. Graphically analysing of these models gives an idea for the forcasting factors importance but to assess the performance of each model in terms of hourly forecasting, results of the three accuracy measurement criteria are observed. The comparisons of the forecasting results are shown in Table 3.

TABLE 3
Electricity forecasting models performance results

	ARIMA	XGBoost
MAE	151.8	101.3
MAPE	480.7	306.9
RMSE	147.5	99.4

Acoording to results, the forecasting performance of XGBoost model is better than ARIMA also, it can be concluded that it more feasibile and practicabile than the ARIMA. However, both of them have high error rate due to the electricity price characteristic as discussed Section II.

The results of these models can be improved by taking into consideration historical load and other relevant data such as weather conditions, maintenance schedules, demand, and so on. This is the key reason that motivated researchers to study hybrid model for the electricity price forecasting.

V CONCLUSION

In this study, the strong interdependence between historical data of electricity price, coal price, natural gas and radyoactive mineral Uranium price are considered for in forecasting the day-ahead electricity price. According to the experimental results, compared to ARIMA model, the XGBoost has been proven to have the best forecasting abilities, and its average, MAE, MAPE and RMSE values are the lowest. The practicality and feasibility of the models implemented are also confirmed in this paper. In addition, to obtain better results, different volatile inputs, which can explain the reasons of the spikes, should be taken into account. The forecasting results allow decision makers to benefit quantify the uncertainty level linked with predictions, which correspond the requirements for risk management purposes and electricity market operations.

The results of these electricity price forecasting study, has high performance and explicit reasoning. Correlation of historical electricity prices and natual gas prices have high importance in this study. Although good performance of the selected models, electricity price forecasting require more information for multiple factors as whether conditions, outages of supply, contingency etc.

A future work will focus on the composit algorithms, which are highly efficient when comparing single models, it is also necessary to add some more features to obtian better accuracy. Therefore, to achieve improvements, more data should be included in the inputs and in this way, the volatility can be easily calculated in the output.

ACKNOWLEDGMENT

This study is supported in part by the Scientic Research Projects Commission of Eskişehir Technical University under the general purpose grant 19ADP005.

REFERENCES

[1] Ü. Başaran Filik, Ö. Nezih Gerek, and M. Kurban, "A novel modeling approach for hourly forecasting of long-term electric energy demand." *Energy Conversion and Management*, vol. 52, pp. 199-211, Jan 2011, doi: 10.1016/j.enconman.2010.06.059.

[2] K. Bitirgen, and Ü. Başaran Filik. "A comprehensive study on modeling of photovoltaic arrays and calculation of photovoltaic potential using digital elevation model." *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects*, vol. 42, pp. 1-25, Oct 2020, doi:10.1080/15567036.2020.1822465.

[3] M.A. Duran and Ü. Başaran Filik. "Short-term wind speed prediction using several artificial neural network approaches in Eskisehir." 2015 International Symposium on Innovations in Intelligent SysTems and Applications (INISTA). IEEE, Sept 2015.

[4] Wang, Zheng-Xin, Qin Li, and Ling-Ling Pei. "Grey forecasting method of quarterly hydropower production in China based on a data grouping approach." *Applied Mathematical Modelling*, vol. 51, pp. 302-316, 2017, doi:10.1016/j.apm.2017.07.003.

[5] J. Bastian, J. Zhu, V. Banunarayanan and R. Mukerji, "Forecasting energy prices in a competitive market", *IEEE Comput. Appl. Power*, vol. 12, no. 3, pp. 40-45, Jul 1999, doi:10.1109/67.773811.

[6] D. W. Bunn, "Forecasting loads and prices in competitive power markets", *Proc. IEEE*, vol. 88, no. 2, pp. 163-169, Feb. 2000, doi:10.1109/5.823996.

[7] L. Zhang and P. B. Luh, "Neural network-based market clearing price prediction and confidence interval estimation with an improved extended Kalman filter method", *IEEE Trans. Power Syst*., vol. 20, no. 1, pp. 59-66, Feb 2005, doi:10.1109/TPWRS.2004.840416.

[8] B.-J. Chen, M.-W. Chang and C.-J. Lin, "Load forecasting using support vector machines: a study on EUNITE competition 2001", *IEEE Trans. Power Syst*., vol. 19, no. 4, pp. 1821-1830, Nov 2004, doi: 10.1109/TPWRS.2004.835679.

[9] A. J. Conejo, M. A. Plazas, R. Espinola and A. B. Molina, "Day-ahead electricity price forecasting using the wavelet transform and ARIMA models", *IEEE Trans. Power Syst*., vol. 20, no. 2, pp. 1035-1042, May 2005, doi: 10.1109/TPWRS.2005.846054.

[10] S. Deng, "Pricing electricity derivatives under alternative stochastic spot price models", Proc. 33rd Hawaii Int. Conf. System Sciences, January 2000.

[11] X.R. Li, C.W. Yu, S.Y. Ren, C.H. Chiu, K. Meng, "Day-ahead electricity price forecasting based on panel cointegration and particle filter," *Electr. Power Syst. Res.*, vol. 95, pp. 66–76, Feb 2013, doi: 10.1016/j.epsr.2012.07.021.

[12] R. Weron, A. Misiorek, "Forecasting spot electricity prices: A comparison of parametric and semiparametric time series models," *International journal of forecasting*, vol. 24, pp. 744–763, Oct/Dec 2008, doi:10.1016/j.ijforecast.2008.08.004.

[13] J. Nowotarski, E. Raviv, S. Trück, R. Weron, "An empirical comparison of alternative schemes for combining electricity spot price forecasts," *Energy Economics*, vol. 46, pp. 395–412, Nov 2014, doi:10.1016/j.eneco.2014.07.014.

[14] T. Papadimitriou, P. Gogas, E. Stathakis, "Forecasting energy markets using support vector machines," *Energy Economics*, vol. 44, pp. 135–142, July 2014, doi: 10.1016/j.eneco.2014.03.017.

[15] A. Azadeh, M. Moghaddam, M. Mahdi, S.H. Seyedmahmoudi, "Optimum Long-Term Electricity Price Forecasting in Noisy and Complex Environments," *Energy Sources Part B Econ. Plan.* vol. 8, pp. 235–244, Feb 2013, doi:10.1080/15567249.2012.678559.

[16] H.T. Pao, "Forecasting electricity market pricing using artificial neural networks," *Energy Convers. Manag.* vol. 48, pp. 907–912, Mar 2007, doi:10.1016/j.enconman.2006.08.016.

[17] J. Contreras, R. Espinola, F.J. Nogales, A.J. Conejo, "ARIMA models to predict next-day electricity prices," *IEEE Trans. Power Syst*., vol. 18, pp. 1014–1020, Aug 2003, doi:10.1109/TPWRS.2002.804943.

[18] E. Erdogdu, "Natural gas demand in Turkey," *Applied Energy*, vol. 87, pp. 211–219, Jan 2010, doi:10.1016/j.apenergy.2009.07.006.

[19] X. Liao, N. Cao, M. Li and X. Kang, "Research on Short-Term Load Forecasting Using XGBoost Based on Similar Days," 2019 International Conference on Intelligent Transportation, Big Data & Smart City (ICITBS), Changsha, pp. 675-678, January 2019, doi: 10.1109/ICITBS.2019.00167.

[20] R.A. Abbasi, "Short Term Load Forecasting Using XGBoost." Workshops of the International Conference on Advanced Information Networking and Applications Mar 2019.

[21] C. Li et al, "Power Load Forecasting Based on the Combined Model of LSTM and XGBoost." Proceedings of the 2019 the International Conference on Pattern Recognition and Artificial Intelligence, Aug 2019.

[22] I. Oksuz, and U. Ugurlu. " Neural Network Based Model Comparison for Intraday Electricity Price Forecasting," *Energies*, vol. 4557, pp.12-23, Nov 2019, doi: 10.3390/en12234557.

[23] https://www.nyiso.com/ (Accessible on 23 Nov. 2020)

[24] https://markets.businessinsider.com/commodities/uranium-price (Accessible on 23 Nov. 2020)

[25] https://www.investing.com/commodities/energy (Accessible on 23 Nov. 2020)

[26] D. Singhal, and K. S. Swarup. "Electricity price forecasting using artificial neural networks," *International Journal of Electrical Power & Energy Systems*, vol. 33.3, pp. 550-555, Mar 2011, doi:10.1016/j.ijepes.2010.12.009.

[27] T. Chen and C. Guestrin, "XGBoost," Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '16, Aug 2016.

[28] W. Yucong, and B. Wang, "Research on EA-Xgboost Hybrid Model for Building Energy Prediction," Journal of Physics: Conference Series, Apr 2020.

[29] X. Chen, et al, "Electricity price forecasting with extreme learning machine and bootstrapping." *IEEE Transactions on Power Systems*, vol. 27, pp. 2055-2062, Apr 2012, doi:10.3390/en12234557.

K. Bitirgen graduated from Anadolu University in Turkey as an Electrical and Electronics Engineer in 2016. She got MSc from Anadolu University in 2018. She is currently a PhD student in Electrical and Electronics Engineering in Eskişehir Technical University. Her special fields of interest include power systems, electrical machines, energy systems, intelligent systems, and renewable energy.

Ü. Başaran Filik graduated from Anadolu University in Turkey as an Electrical and Electronics Engineer in 2002. She got MSc and PhD degree from Anadolu University in 2004 and 2010, respectively. She is currently an associate professor on Power System Analysis in Eskişehir Technical University. Her special fields of interest include power system analysis–optimization, renewable energy systems, and smart grids.