Electrical Energy Demand Prediction: A Comparison Between Genetic Programming and Decision Tree

Ali DANANDEH MEHR[1,*] , Farzaneh BAGHERI[2] , Mir Jafar Sadegh SAFARI[3]

¹Department of Civil Engineering, Antalya Bilim University, Antalya, Turkey

²Department of Electrical and Electronics Engineering, Antalya Bilim University, Antalya, Turkey

³Department of Civil Engineering, Yaşar University, Izmir, Turkey

Highlights

• Design of GP model for electricity demand prediction.
• Design of DT model for electricity demand prediction.
• GP is slightly superior to DT.
• GP provides explicit model.
• GP and DT can be used for 1-day ahead forecast of electricity in Nicosia.

Abstract

Several recent studies have used various data mining techniques to obtain accurate electrical energy demand forecasts in power supply systems. This paper, for the first time, compares the efficiency of the decision tree (DT) and classic genetic programming (GP) data mining models developed for electrical energy demand forecasting in Nicosia, Northern Cyprus. The models were trained and tested using daily electricity consumptions measured during the period 2011-2016 and were compared in terms of three statistical performance indices including coefficient of determination, mean absolute percentage error and concordance coefficient. The prediction results showed that the proposed models can be effectively applied to forecasts of electrical energy demand. The results also indicated that the GP is slightly superior to DT in terms of the performance indices.

1. INTRODUCTION

The 21 century is known as the age of information means that a huge amount of data from multiple sources is produced, measured, and recorded every second. In other words, data is everywhere, available in multiple formats and expanding exponentially. To extract information from data, many learning algorithms are available that typically use computers to learn underlying knowledge in data. Data mining is a set of techniques that can be used to classify the data and distill knowledge from it. Classically, data mining is defined as the computing process of discovering patterns in large data sets [1]. It involves the use of machine learning (ML), statistics and database systems to recognize the hidden patterns among the data. Once the patterns (knowledge) are recognized, they can be used for generalization aims; however the pattern must be verified before we reach any sort of conclusions.

Using the recent advancements in computer technology, there are noticeable increase in the application of data mining techniques for descriptive and predictive modelling. Some of the tasks accomplished by mining a sort of data include anomaly detection, association rule mining (i.e., market basket analysis), clustering, classification, regression and time series forecasting. In the present study, we apply two data mining techniques, namely genetic programming (GP) and decision tree (DT) to accomplish both classification and

*Corresponding author, e-mail: ali.danandeh@antalya.edu.tr

regression tasks in order to forecast electrical energy demand (hereafter ED) based upon observed energy consumption time series.

The ED forecasting models are used in the wider domain of energy management such as power generation and distribution, energy market, power plant operation, building energy demand, etc. A number of data mining-based models have been suggested to estimate ED using either historical power demand or via regression analysis among ED and germane forcing parameters such as temperature, demography, charges, and others. For instance, artificial neural networks (ANN) were compared with regression analysis as well as DT methods and was shown that ANNs can forecast the ED more precise than the counterparts [2]. In a similar work, ANN was also used to the long-term forecast of ED in Greece [3]. The author showed that the ANN forecasts are more truthful than those of linear regression as well as support vector regression (SVR). DT method was applied for building energy use intensity modeling of Japanese residential buildings [4]. The authors elaborately demonstrated how one can generate decision trees based on a set of training data and how the results can be evaluated on test data. The authors also showed that DT is capable of classifying and predicting building energy demand levels precisely. Because of the non-stationary feature of the ED time series, the frequently used ad hoc ML methods may not be accurate enough to forecast ED. In such conditions, hybrid ML methods were implemented and promising results were reported. For instance, [5] demonstrated that a hybrid genetic algorithm (GA)-ANN model can meaningfully improve the accurateness of ad hoc ANN-based ED forecasting models. Even though these crossbreed models may create favorable forecasts, these are implicit models in which the relationship between the inputs and the output data cannot be easily expressed in a clear mathematical form. Therefore, hybrid models would be difficult to be used in practice.

The authors' review in the relevant literature revealed that the primary application of GP for ED prediction belongs to [6] where the study examined the capability of linear GP for short-term ED forecasts in the State of Victoria, Australia. The results showed that the linear GP generates direct and truthful models. Later, [7] used a more recent GP variant, namely gene expression programming (GEP) to regress ED series to the climatological features in Canada. The model was reported as efficient to be employed in practice. The advantages of GP and GEP in identifying precise and explicit models were described for electricity prediction, in which a clear mathematical form is provided or can be induced from the GP/GEP chromosome [8, 9]. Recent attempts have also discussed the usage of the smart meters (SMs) in a smart-city grid (SG) as enabling technologies for the aforementioned ML methods in order to enhance the process of the ED prediction between the supplier and the demand side considering different aspects [10-12]. Among such applications, power outage detections/allocation, and the voltage/frequency fluctuations monitoring are essential features/events to monitor not only for power utilities but also for customer benefits.

The objective of this study is to use GP and DT, as the subsets of data mining approaches, in order to develop ED forecasting models. The ultimate goal is to compare the effectiveness of the models and to show how they can be applied to evolve one-day ahead ED prediction models for Nicosia, the capital city of Northern Cyprus. To this end, the DT code, written in MATLAB, and GPdotNET, an open-source GP software, were used in this study. Both the programs provide user-friendly frameworks with the capability of exporting their tree-based results to numerically readable spreadsheet data sets. Currently, energy demand prediction for Nicosia is not efficient and therefore, our results may be used to increase the efficiency of the energy industry of Northern Cyprus.

2. METHODS AND DATA

2.1. Utilized Data

The GP and DT models developed in the present study use daily ED records from the northern part of Nicosia (Turkish Republic of Northern Cyprus) in the period 2011-2016. The city has a typical Mediterranean climate with annual average electricity consumption of about 4000 MWh on the northern part and about 6000 MWh on the southern part [13]. The local power generation and distribution authority, called KIBTEK, predicts ED occasionally based only on the weather conditions not the actual user energy

consumption. It applies the SCADA which is a sophisticated expensive solution for ED prediction with intensive administrator dependence. The electricity consumption time series is depicted in Figure 1. This is a refined data (anomaly detected and removed) indicating a positive trend in energy use, particularly since 2013. The figure also represents seasonal behavior, since the sinusoidal shape of the data can be recognized visually. The statistical properties of the electricity consumption series are shown in Table 1.
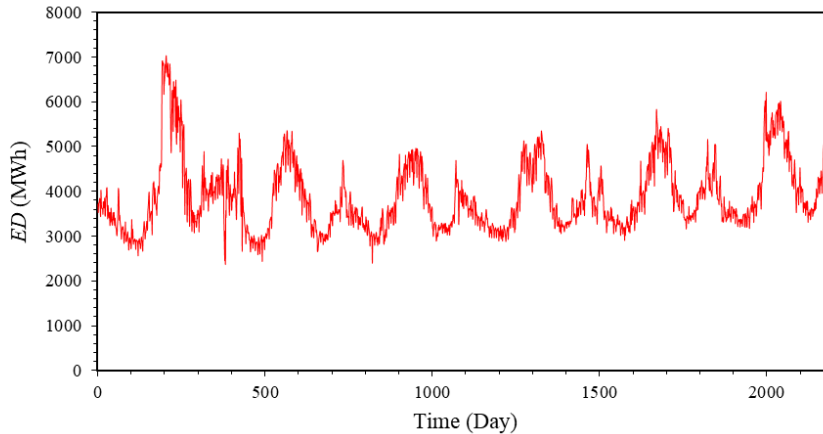

Figure 1. *Daily ED records in northern part of Nicosia*

Table 1. *Statistical properties of implemented data, using the observations 2011-01-01 to 2016-12-31*

Data set	Mean (MWh)	Minimum (MWh)	Maximum (MWh)
Entire data	3876	2130	7032
Training set	3782	2130	7032
Validation set	4097	2293	6215

2.2. Data Preparation

Developing a DT model for modeling a nonstationary time series seems natural due to the complex nature of such events. Since the ED shown in Figure 1 represents a typical nonstationary time series (obvious trend and seasonality), the DT model was developed and compared with GP that both can handle the nonstationary series. The available data (i.e., ED time series) is represented with only one variable which is the example of a univariate time series. To prepare the data for GP- and DT-based univariate modeling, it must be transformed into data frame-based sets, i.e., features (input) and label (output). The features are generated by the previously observed values, so-called time lag values while the label is the time series value at the current time step. In this way, the time series is transformed so that historical changes influence the current value. For both DT and GP model configurations, seven days of time lags were selected and normalized using mi-max normalization method so that normalized ED at time step t+1 is assumed to be the function of normalized *ED* at the current time step ED_t and previous time steps ED_{t-i} as expressed in Equation 1. Once the data frame is created, training (80%) and validation (20%) data sets were created to calibrate and verify the models

$$ED_{t+1} = f(ED_{t-i}) \ and \ (i = 0,1,...,7 \ ). \tag{1}$$

2.3. DT-based Modelling

As a data mining technique, DT is applied commonly in many engineering fields for classification and prediction problems [4, 14, 15]. DT utilizes the information available in the data set mainly for classification purposes. Respecting the initial information on each class data, a tree-based model is generated as a diagram called the decision tree. It is a nonparametric regression method in which training algorithm output is a type of discrete class. To obtain accurate results, the entropy theory is used for the characterization of impurity on a data set. DT generates a discrete response for the model through an entropy reduction at its structure. In the modeling procedure, a binary tree is used for finding the optimum tree size. The degree of importance

of variables in DT is related to their location at DT flowchart, where preceding sets have higher entropy. Lowest entropy belongs to leaves that cause a reliable decision as a decision support system [16].

DT finds a relationship between dependent and independent variables in a nonlinear manner by a series of binary splits. Assuming a sample X containing a set of independent variables (x_1, x_2, …, x_n), and the dependent variable of y (here ED_{t+1}), DT determines y via subdivision of sample X. The tree is grown through successive branches of subsets of sample X into subsequent branches. The root node is defined as X and the binary tree is finalized in the set of terminal nodes. Based on the independent variable vector value the branches are determined at each node of a binary tree. The branches selection at each parent node is based on the DT algorithm in which model variance in y has the maximum reduction. The decision at each node comes from considering a specific value of the independent variable at the data range domain. Decisions stop once the model variance reduction reaches a specific value called the critical number. Finally, in order to find the optimal tree size, the cross-validation is used [17, 18].

2.4. GP-based Modelling

Fundamentals of GP and its variants can be found in [19]. Therefore, a brief overview is presented here in order to help the reader to realize the applied methodology in this paper. GP is an evolutionary algorithm that generates and optimizes computer programs (called individuals) to solve a wide diversity of problems. For more than a couple of decades, GP has been one of the most robust ML techniques to solve symbolic regression problems [20, 21]. However, it was also used to solve binary or multiclass classification problems [22, 23]. The main reason for its great success over time is it's explicitly and ability to evolve mathematical expressions that can be easily analyzed, validated and applied in practice. Similar to most of the evolutionary algorithms, the main operators that are used to convert the initial population of individuals (i.e., random programs) to a set of preferred solutions include *crossover*, *mutation*, and *reproduction*. In contrast to the genetic algorithm, these operations act on a single or a pair of individuals having a tree shape known as the GP tree. GP trees can be commonly expressed by mathematical equations. For example, mathematical representations of a GP tree was illustrated in Figure 2.
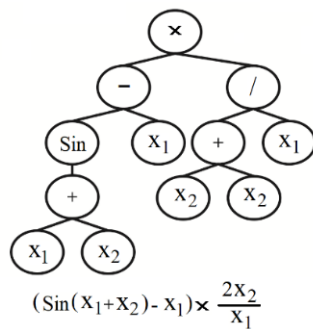

$$(Sin(x_1+x_2)-x_1)\times \frac{2x_2}{x_1}$$

Figure 2. Tree and mathematical representations of a GP program

Using crossover operation, some nodes and branches of a pair of GP trees (called parents) are exchanged. It is expected that the operation creates better programs (called offspring). Reproduction is the direct copy of the best GP tree in the initial population to a new population of individuals. Mutation is the alteration of a node or branch in a single GP tree. Each node in a tree can adopt a function or terminal variables such as x_1 and x_2 in Figure 2. One of the main issues in the GP-based modeling is the selection of a set of appropriate functions, input variables and maximum depth (also referred to as height) of GP trees. While the appropriate functions and variables can be select concerning the physics of the desired phenomenon, the tree depth is commonly constrained in order to control the complexity of the model to be developed. In an analogy with ANNs, the latter acts similar to the implementation of a restricted number of hidden neurons/layers. Complexity control in GP setup may also help the system to avoid overfitting problems where the solutions show very low errors on the training set, but high errors on the testing set.

Nowadays, a few GP tools are available to run regression or classification analysis. A list of GP software frameworks/scripts and their features were presented in [19]. To create GP solutions in the present study,

we utilized GPdotNET which uses $C^{\#}$ programming language to run evolutionary operations and to compare the solutions in terms of different goodness of fit functions.

3. RESULTS

3.1. DT Results

The DT models developed in this study is shown in Figure 3. It has to be emphasized that the priority of decision is linked to the level of decision in each branch. As it is shown in Figure 3, the DT model has five decision steps algorithm. It is seen in Figure 3 that ED_t is located at the top of the DT. It indicates that ED_t is the most important variable in the modeling procedure in the decision-making plan of this study. The aforesaid assertion is supported for levels two and three, where ED_t is located at one-three levels of DT flowchart. Looking to the forth level of DT, it can be found that behind ED_t, variables of ED_{t-5}, ED_{t-2} and ED_{t-4} have a great effect on the model output. It is also seen in Figure 3 that ED_{t-3} is located at the fifth level and variable ED_{t-1} is not listed in DT flowchart showing its lowest importance in the modeling procedure of this study.

The observed ED values are compared with corresponding values computed by the DT model in Figure 4. It can be seen that the DT-based model yield high performance, where provides results close to the measured values.
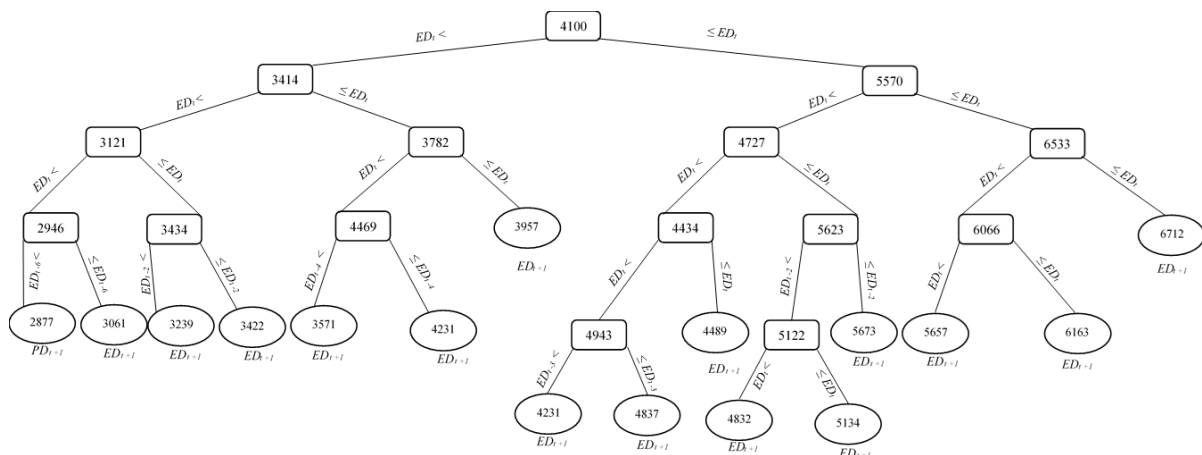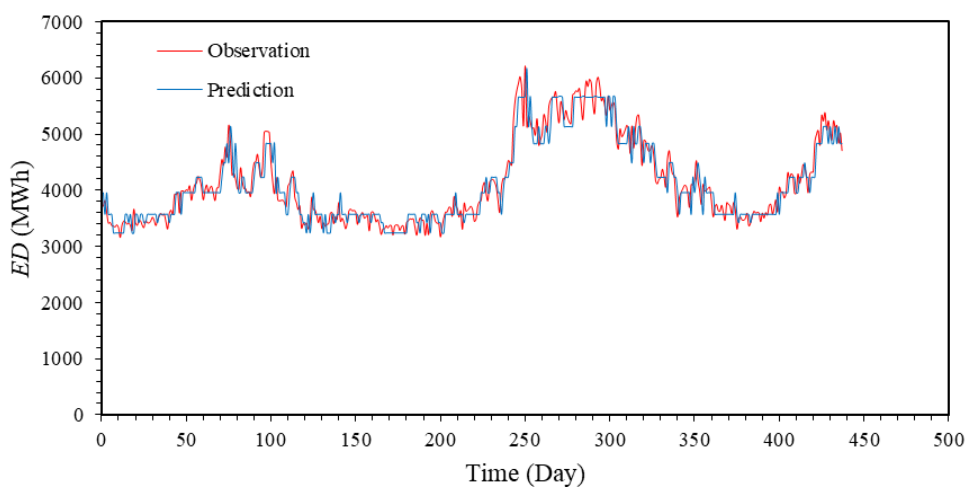

Figure 3. *DT model for energy demand prediction*

Figure 4. *Predicted demand obtained by the DT model for the validation set*

3.2. GP Results

As previously mentioned, the normalized daily ED prediction scenario was also considered to be modeled by GPdotNET. Similar to DT, the *RMSE* between the predicted and target values was selected as the objective function as suggested by Hrnjica and Danandeh Mehr [24]. This is the absolute error function often used as a fitness function in regression problems. In addition to lagged ED series (i.e., input vectors), five random constants in the range 0-1 were generated and used as members for terminal and inner nodes of GP individuals. The range is identical to the range of input vectors so the GP solutions would not be biased toward constants. To avoid more complex programs and possible overfitting problem, only four basic arithmetic operators (+,*, /, -) were selected as potential functions in the GP setup. Besides, the maximum height of the best solutions was limited to six levels. Some other important evolutionary parameters used in the setup of the GP tool are listed in Table 2.

Table 2. *Statistical properties of implemented data, using the observations 2011-01-01 to 2016-12*

Parameter	Selected values
Population Size	500
Initialization	Ramped Half and Half
Validation set	4097
Maximum Number of Generations	500
Selection Method	Tournament Selection
Probability for Crossing-Over	90%
Probability for Mutation	5%

It's worth mentioning that we monitored fitness simulation at each generation and ended the runs when the average fitness of the evolved models reaches to the fitness of the best-evolved model (see Figure 5). This procedure helps the modeler to get away from highly complex solutions.
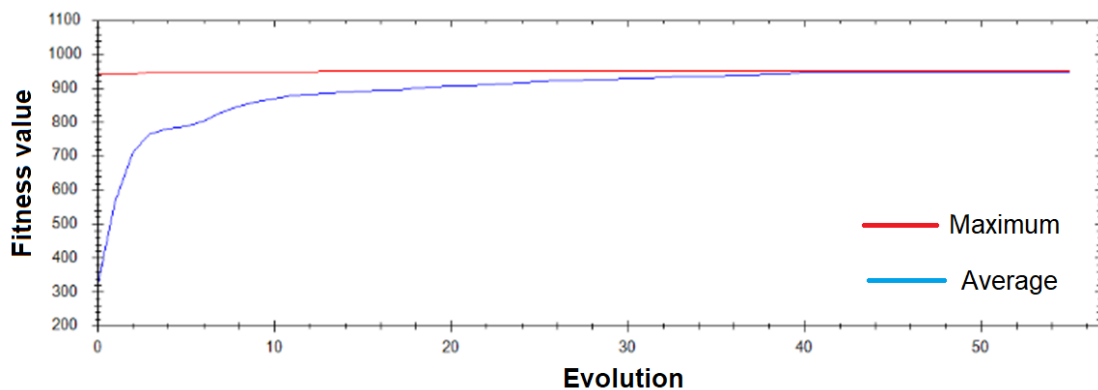

Figure 5. *Fitness values obtained by the GP model for the training set*

Figure 5 indicates that GPdotNET is able to provide the best individual approximately after 40 generations, and program growth may increase the complexity of the solutions without noteworthy improvement in prediction accuracy. Figure 6 shows the best GP model attained for 1-day ahead ED forecast at Northern Nicosia. The parameters x_1, x_2, …, and x_7 shown in the figure represent the normalized values of ED_{t-6}, ED_{t-5}, ..., and ED_t, respectively. The parameters r_i are the random constants created by GPdotNET This is the row model tree that might include some introns, i.e., a program part without any influence on the result. Thus, it requires special attention that can simplify the model.

The mathematical function of the best GP model (Figure 6) has been expressed in Equation (2). The constant 0.388 in the equation is the product of 0.542 and 0.716, which are randomly generated by GP in the initial stage of the model run.

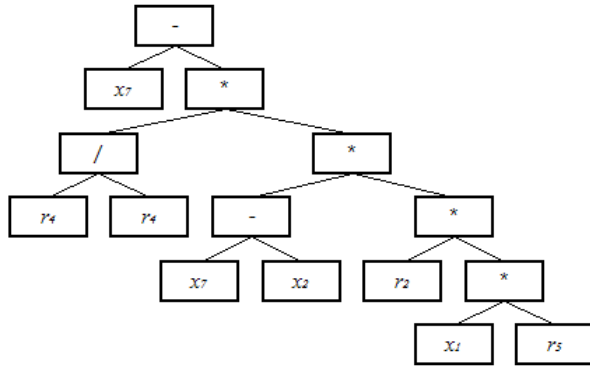$$ED_{t+1} = ED_t - (ED_t - ED_{t-5})(0.388 \times ED_{t-6}).$$ (2)

Figure 6. *The best GP model for ED forecasting*

Figure 7 plotted the best GP model forecasts and the observed ED series at the validation period. Similar to the DT results, the GP model is also able to capture ED fluctuations as well as both low and peak consumptions.
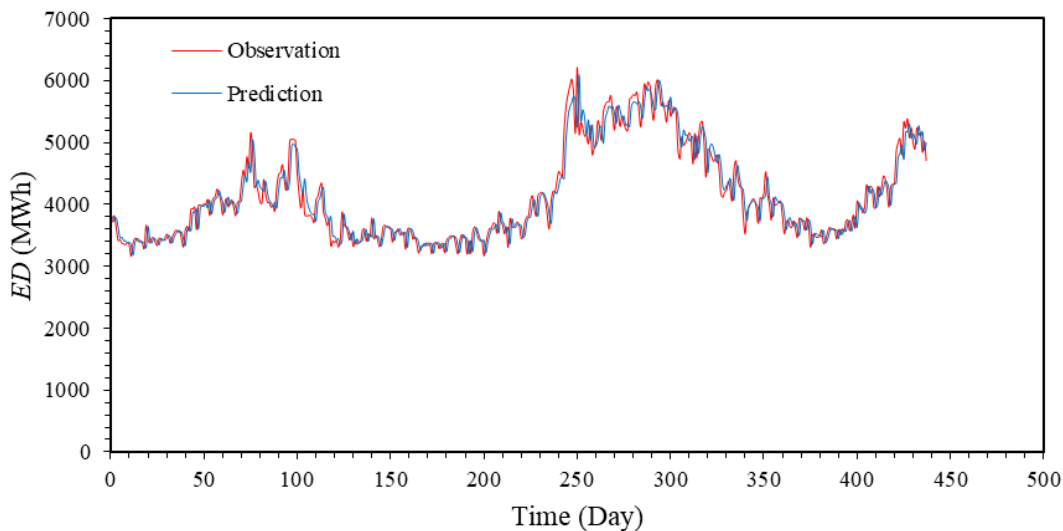

Figure 7. *Predicted demand obtained by the GP model for the validation set*

3.3. Comparison of DT and GP

To compare the accuracy of the DT and GP models developed in this study, their performances are investigated based on three statistical performance indices; coefficient of determination (R^2), mean absolute percentage error (*MAPE*) and concordance coefficient (*CC*).

The R^2 is a statistical measure of how the results are fitted to the best fit line and defined as

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(\left(ED_{t+1}^{obs}\right)_i - \left(ED_{t+1}^{pre}\right)_i\right)^2}{\sum_{i=1}^{n}\left(\left(ED_{t+1}^{obs}\right)_i - \left(ED_{t+1}^{obs}\right)_{mean}\right)^2} \tag{3}$$

where ED_{t+1}^{obs}, ED_{t+1}^{pre} and $\left(ED_{t+1}^{obs}\right)_{mean}$ respectively are, observation, prediction and mean of observation power demand, and n is number of data.

The *MAPE* calculates model errors by comparison of the observation and prediction outputs defined as

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{\left(ED_{t+1}^{pre}\right)_i - \left(ED_{t+1}^{obs}\right)_i}{\left(ED_{t+1}^{obs}\right)_i}\right| \times 100 . \tag{4}$$

The *CC* shows the concordance between observation and prediction data. It ranges from 1 to -1 and has the perfect value of 1. *CC* is calculated by

$$CC = \frac{2r\sigma_{obs}\sigma_{pre}}{\sigma_{obs}^2 + \sigma_{pre}^2 + \left(\left(ED_{t+1}^{obs}\right)_{mean} - \left(ED_{t+1}^{pre}\right)_{mean}\right)^2} \tag{5}$$

where r is correlation coefficient, σ_{obs} and σ_{pre} are standard deviation of observation and prediction power demand respectively, and $\left(ED_{t+1}^{pre}\right)_{mean}$ is the mean of prediction power demand.

Table 3 shows the performance of DT and GP models in terms of three statistical indices of R^2, *MAPE* and *CC*. The R^2 and *CC* values in both models are so close to the perfect fitness value of 1.00 with slightly higher values for the GP model. Regarding the error measure of *MAPE*, the GP model has also less error than DT. Therefore, it can be concluded that GP provides better performance than DT in the present case study. Considering the efficiency results given in Table 3, one can imply that the DT and GP models have almost similar performance with close ED forecasts. Figure 8 compares the forecasts and observed ED time series during the validation period. Although both the models seem reliable for ED prediction in northern Nicosia, the figure indicates that GP is more satisfactory than ED in capturing peak ED values. It is also worthy to mention that the GP model is simpler, explicit, and straightforward and therefore, motivating to be used in practice.

Table 3. *Model evaluation measures at validation period based on R^2, MAPE and CC*

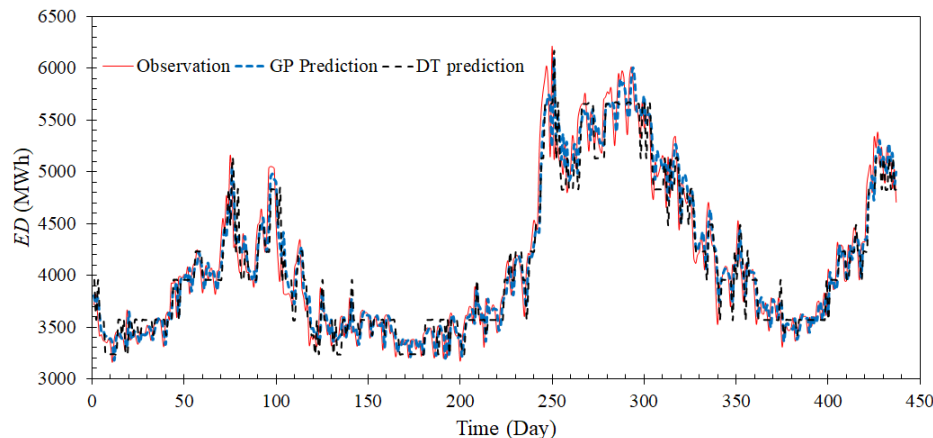| Index	DT	GP
R^2	0.901	0.918
MAPE	4.21	3.7
CC	0.950	0.958

Figure 8. *Observed and predicted electricity demand in the validation period*

Both GP and DT techniques have the ability to determine the most effective variables through the modeling procedure. For instance, looking to the ED pattern, based on DT results shown in Figure 3, the energy consumption in northern Nicosia is correlated to one-, three-, four-, five-, and seven-day ago variables, while in GP it is more correlated to the one-, six-, and seven-day antecedent values. It is understood that both GP and DT techniques can find the most effective variables within the input parameters. It makes an advantage in the application of GP and DT to consider a higher number of input parameters regardless of

their degree of impact on output. However, it should be noticed that having a higher number of input parameters may increase the time of computation. It has to be emphasized that DT can detect the importance degree of input parameters as well. As mentioned before, most important parameters are located at the top level of the tree, while less important parameters are located at the lower levels. It is the advantage of DT in comparison with GP, where the degree of effectiveness of parameters can be simply found from DT flowchart. It is seen from the results that energy consumption in northern Nicosia is most correlated to the one-day ago variable (ED_t) which have the greatest impact on the model output.

4. CONCLUSION

Using historical data, this study compared the efficiency of two data mining methods for ED prediction. To demonstrate and verify the proposed models' efficiency, a simple use case was presented using historical electricity consumption records from Nicosia and one-day ahead ED prediction models were developed. The performance of the models was evaluated in terms of coefficient of determination, mean absolute percentage error and concordance coefficient. The results showed that the proposed models can easily handle the daily ED process. However, the GP is slightly superior to DT. Both GP and DT methods are able to classify and predict ED variables using a tree shape structure that enable the modeler to quickly extract useful knowledge from the phenomenon.

The current framework was confined to historical electricity consumption records as required data for ED predictions. To decrease the uncertainty of ED predictions, particularly in long-term forecasting scenarios such as weekly, monthly, and even seasonally, it may be useful to include other effective parameters such as climate and the economic factors into the prediction process. For example, air temperature and touristic information at summer seasons may have a significant role in reliable ED prediction. Returning to the study area condition, it is also worth to mention that the almost entire ED in Northern Cyprus is produced by fuel that creates a high level of environmental pollution [24]. Undoubtedly, precise ED predictions in the country will lead to saving energy and consequently reduce in the use of fuels that is known as one of the most eco-environmental problems of the island.

CONFLICTS OF INTEREST

No conflict of interest was declared by the authors.

REFERENCES

[1] Han, J., Pei, J. and Kamber, M., "Data mining: concepts and techniques", Elsevier. (2011).

[2] Tso, G. K. and Yau, K. K., "Predicting electricity energy consumption: A comparison of regression analysis, decision tree and neural networks", Energy, 32(9): 1761-1768, (2007).

[3] Ekonomou, L., "Greek long-term energy consumption prediction using artificial neural networks", Energy, 35(2): 512-517, (2010).

[4] Yu, Z., Haghighat, F., Fung, B. C. and Yoshino, H., "A decision tree method for building energy demand modeling", Energy and Buildings, 42(10): 1637-1646, (2010).

[5] Azadeh, A., Ghaderi, S. F., and Sohrabkhani, S., "Annual electricity consumption forecasting by neural network in high energy consuming industrial sectors", Energy Conversion and management, 49(8): 2272-2278, (2008).

[6] Bhattacharya, M., Abraham, A., Nath, B., "A linear genetic programming approach for modelling electricity demand prediction in Victoria", In Hybrid Information Systems, 379-393. Physica, Heidelberg, Berlin, Germany Springer-Verlag, (2002).

[7] Bakhshaii, A. Stull, R., "Electric load forecasting for western Canada: A comparison of two non-linear methods", Atmosphere-Ocean, 50(3): 352-363, (2012).

[8] Çunkaş, M., Taşkiran, U., "Turkey's electricity consumption forecasting using genetic programming", Energy Sources, Part B: Economics, Planning, and Policy, 6(4): 406-416, (2011).

[9] Mousavi, S. M., Mostafavi, E. S. and Hosseinpour, F., "Gene expression programming as a basis for new generation of electricity demand prediction models", Computers & Industrial Engineering, 74: 120-128, (2014).

[10] Aghaei, J., and Alizadeh, M. I., "Demand response in smart electricity grids equipped with renewable energy sources: A review", Renewable and Sustainable Energy Reviews, 18: 64-72, (2013).

[11] Mwasilu, F., Justo, J. J., Kim, E. K., Do, T. D., Jung, J. W., "Electric vehicles and smart grid interaction: A review on vehicle to grid and renewable energy sources integration", Renewable and sustainable energy reviews, 34: 501-516, (2014).

[12] Mehrotra, K., Mohan, C. K., Ranka, S. Elements of artificial neural networks. 2nd ed. Massachusetts, USA, MIT press, (2000).

[13] Danandeh Mehr A., Bagheri, F. and Reşatoğlu, R., "A genetic programming approach to forecast daily electricity demand", 13th International Conference on Theory and Applications of Fuzzy Systems and Soft Computing. Warsaw, Poland, 27–28 August, (2018).

[14] Quinlan, J. R. Induction of decision trees. Machine Learning, 1: 81–106, (1986).

[15] Safari, M. J. S., "DT, GR and MARS models for sediment transport in sewer pipes", Water Science and Technology https://doi.org/10.2166/wst.2019.106, (2019).

[16] Vaheddoost, B., Aksoy, H., Abghari, H., Naghadeh, S. "Decision tree for measuring the interaction of hyper-saline Lake and coastal aquifer in Lake Urmia." In Proceeding of Environmental and Water Resource Institute (EWRI): Watershed Management Symposium, August (5-7), (2015).

[17] Breiman, L., J. Friedman, Olshen, R. C. Stone, Classification and Regression Trees, Wadsworth, Belmont, Calif., (1984).

[18] Balk, B., and Elder, K., "Combining binary decision tree and geostatistical methods to estimate snow distribution in a mountain watershed", Water Resources Research, 36(1): 13-26, (2000).

[19] Hrnjica, B., and Danandeh Mehr, A., Optimized Genetic Programming Applications: Emerging Research and Opportunities: Emerging Research and Opportunities. Hershey, PA, USA, IGI-Global, (2019).

[20] Danandeh Mehr, A., Nourani, V., Kahya, E., Hrnjica, B., Sattar, A. M., Yaseen, Z. M, "Genetic programming in water resources engineering: A state-of-the-art review", Journal of Hydrology 566: 643-667, (2018).

[21] Safari, M.J.S., Danandeh Mehr, A., "Multigene genetic programming for sediment transport modeling in sewers for conditions of non-deposition with a bed deposit." International Journal of Sediment Research 33(3): 262-270, (2018).

[22] Danandeh Mehr, A., Nourani, V., Hrnjica, B., Molajou, A., "A binary genetic programing model for teleconnection identification between global sea surface temperature and local maximum monthly rainfall events", Journal of Hydrology, 555: 397-406, (2017).

[23] Danandeh Mehr, A., Jabarnejad, M., Nourani V., "Pareto-optimal MPSA-MGGP: A new gene-annealing model for monthly rainfall forecasting", Journal of Hydrology, 571: 406-415, (2019).

[24] Hrnjica, B., Danandeh Mehr. A., "Energy Demand Forecasting Using Deep Learning." Smart Cities Performability, Cognition, & Security. Springer, Cham, 71-104, (2020).