SAKARYA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ DERGİSİ

Sakarya University Journal of Science
SAUJS

e-ISSN 2147-835X Period Bimonthly Founded 1997 Publisher Sakarya University
http://www.saujs.sakarya.edu.tr/

Title: Electricity Load Forecasting Using Deep Learning and Novel Hybrid Models

Authors: Muhammed SÜTÇÜ, Kübra Nur ŞAHİN, Yunus KOLOĞLU, Mevlüt Emirhan ÇELİKEL, İbrahim Tümay GÜLBAHAR

Recieved: 2021-08-14 00:00:00

Accepted: 2021-12-19 00:00:00

Article Type: Research Article

Volume: 26
Issue: 1
Month: February
Year: 2022
Pages: 91-104

How to cite
Muhammed SÜTÇÜ, Kübra Nur ŞAHİN, Yunus KOLOĞLU, Mevlüt Emirhan ÇELİKEL, İbrahim Tümay GÜLBAHAR; (2022), Electricity Load Forecasting Using Deep Learning and Novel Hybrid Models. Sakarya University Journal of Science, 26(1), 91-104, DOI: 10.16984/saufenbilder.982639
Access link
http://www.saujs.sakarya.edu.tr/tr/pub/issue/67934/982639

Sakarya University Journal of Science 26(1), 91-104, 2022

Electricity Load Forecasting Using Deep Learning and Novel Hybrid Models

Muhammed SÜTÇÜ*[1], Kübra Nur ŞAHİN[1], Yunus KOLOĞLU[1],
Mevlüt Emirhan ÇELİKEL[1], İbrahim Tümay GÜLBAHAR[1]

Abstract

Load forecasting is an essential task which is executed by electricity retail companies. By predicting the demand accurately, companies can prevent waste of resources and blackouts. Load forecasting directly affect the financial of the company and the stability of the Turkish Electricity Market. This study is conducted with an electricity retail company, and main focus of the study is to build accurate models for load. Datasets with novel features are preprocessed, then deep learning models are built in order to achieve high accuracy for these problems. Furthermore, a novel method for solving regression problems with classification approach (discretization) is developed for this study. In order to obtain more robust model, an ensemble model is developed and the success of individual models are evaluated in comparison to each other.

Keywords: Load forecasting, deep learning, regression by classification.

1. INTRODUCTION

Electricity is one of the most important energy sources all around the world. It has a big impact of development of economies and nations [1]. With developing technology and population, the need for electricity has been increased. The stability of electricity has vital importance in today's society. In order to achieve sustainable economic growth, continuous electricity production and supply are important. Because of the nature of electricity, it cannot be stored. It should be generated and consumed simultaneously [2].

In Turkey, electricity market operations are done on three different markets [3]. First one is "Day Ahead Market". In Day Ahead Market, electricity producer companies and suppliers notify EPIAS (Electricity Market Operations Co.) about the consumption and production hourly quantities for the following day. Purchasing price for electricity is announced after participants notify the system, and it is named "PTF" (Market Clearing Price) [3]. Second market in the electricity operations is called "Intra-day Market". In this market, participants have opportunity to make additional purchase and sell operations in case of unexpected events, such as sudden changes in temperature and failures in the system. Participants create bids of hour, amount and price for the electricity and transactions occur if the price of the offers for given hour in the system are matched. This market

* Corresponding author: muhammed.sutcu@agu.edu.tr
[1] Abdullah Gül University, Faculty of Engineering, Department of Industrial Engineering
E-mail: kubranur.sahin@agu.edu.tr, yunus.kologlu95@gmail.com, mevlutemirhan.celikel@agu.edu.tr, ibrahim.gulbahar@agu.edu.tr
ORCID: https://orcid.org/0000-0002-8523-9103, https://orcid.org/0000-0001-9786-6270, https://orcid.org/0000-0001-6198-569X, https://orcid.org/0000-0001-9264-4345, https://orcid.org/0000-0001-9192-0782

is useful to mitigate the losses caused by the forecasts that are done in the previous day. A new type of electricity price is formed in this market, and it is named "SMF" (System Marginal Price) [3]. SMF is based on the real consumption values, and it is announced at least four hours later than the consumption. The third market is called "Balancing Market" [3]. All plus and negative deviations in the forecasts of the participants are neutralized in this market, and they pay a penalty cost for their forecast errors according to the announced PTF and SMF.

Electricity distribution companies function as a bridge between producers and consumers of electricity. One of the most important concerns of electricity distribution companies is supplying the customer demands. In order to manage electricity demand and minimize financial loss, it is necessary to forecast the usage of electricity in a distribution network. Electricity producers also need to forecast their production in order to minimize their loss from their profit and contribute the stability of Turkish Electricity Market. Because, overproduction as a result of positively deviated forecasts causes the waste of resources and financial loss and underproduction as a result of negatively deviated forecasts may cause technical issues which can lead to power outage [4]. For electricity market participants, making accurate predictions reduces their financial loss besides making their planning activities easier. With this study which will be conducted with an electricity retail company, the operations of the company in the electricity markets will be improved and its financial loss will be reduced. In the company's current system, the forecasts are done intuitional way. Their forecast based on previous year data and current weather condition and the company do not any systematical way of forecasting. The forecasting strategy of the company based on personal experience. Using the personal experience-based forecasting method is not sustainable due to the increasing number of parameters that affect consumption of electricity and complexity of the forecasting. The pattern of electricity consumption depends on many different parameters such as seasons, hours, weekdays, holidays, temperature, humidity, rainfall, cloudage and etc. These parameters make electricity demand forecasting more complex. If the electricity demand of customers is underestimated, the electricity shortage may occur locally or countrywide. Moreover, if the demand overestimated, the surplus energy should be sold in the balancing energy market with at least %3 loss. Therefore, the company faces financial losses due to inaccurate forecasting the electricity consumption. In order to decrease these losses, the company needs more accurate and systematic forecasting method for electricity consumption.

In that study, hourly electricity consumption of a city is forecasted with machine learning approaches. The aims of the study are minimizing the hourly forecasting errors.

In terms of contributions, we developed a novel ensemble method to solve a multivariable and multi-dimensional estimation problem. The proposed method involves discretizing of a large regression problem into small problems and a deep neural network model with optimized hyper parameters. The discretization and deep neural network models are combined so that we obtained a more robust forecasting model. The ensemble model outperforms the two learning models it involves

2. LITERATURE REVIEW

Solar energy systems are the most abundant and preferred renewable energy in the world [5]. There are two ways to convert solar energy into electricity. Sun Thermal Power Plants (STPP) use the heat of beams to transfer direct normal solar irradiance to electricity, whereas Photovoltaic plants turn energy directly into electricity [6]. Worldwide photovoltaic production is steadily increasing, with the International Energy Agency (IEA) estimating that solar generation will meet 2% of global electricity demand by 2030 [7]. As these technologies were more integrated into large-scale power systems, it became necessary to estimate their production [8]. Although energy production from other traditional sources may be easily computed, it is difficult to predict exactly due to the considerable unpredictability in weather conditions. To achieve a successful integration into the power systems, it is required to estimate output for the coming days and hours.

There are numerous instances of load forecasting models that scientists have applied. In load forecasting, Azedah (2008) used a fuzzy technique to determine the kind of ARMA models [9]. In an electricity forecasting problem, Wang (2008) coupled autoregressive models and moving averages with exogenous factors (such as weather conditions) [10]. To determine the load of power systems, Amjady (2007) used a hybrid model that combined the multilayer perceptron (MLP) neural network and the forecast-aided state estimator (FASE) [11]. Furthermore, several of these authors computed the load by dividing the days into 24-hour intervals or days in a week [12].

For predicting photovoltaic energy, machine learning methods are often utilized. The goal of supervised learning algorithms is to find a mapping between given inputs and outputs [13]. Unsupervised learning models, on the other hand, look for hidden structure between input values without using output variables [14]. In that way, it's equivalent to determining the input distribution. Aside from supervised and unsupervised approaches, it is also effective to combine several successful models to improve overall model performance, which is known as “ensemble learning” [15].

Vapnik developed Support Vector Machine (SVM) which is a supervised learning method for solving classification and regression applications. [16]. The goal of SVM is to determine the optimum hyperplane for accurately separating data into multiple groups. The data points with the shortest distance to the defined hyperplane are referred to as Support Vectors. Sharma, Sharma, Irwin and Shenoy (2011) used machine learning techniques and compared the forecast outcomes of National Weather Service (NWS) [17]. Weather data from January to December 2010 was utilized in their research. Temperature, dew point, wind speed, sky cover, chance of precipitation, and relative humidity are the weather metrics used in the data. Furthermore, days and hours are included in the data. Linear Regression and SVM with Radial Basis Function (RBF) kernel were employed as methods, and training data was taken between January and August. Principal Component Analysis was used to remove redundant data, and as a consequence of the research, SVM with RBF kernel was found to be more successful with a smaller RMS error. Between January 2010 and October 2010, Shi et al. (2012) conducted a study in China to forecast solar production [18]. The study's data interval is fifteen minutes, and production values are normalized to improve accuracy throughout the preprocessing step. Data was divided into four groups based on weather conditions before being used to develop the model: sunny, foggy, rainy, and cloudy days. As indicated in the study, the RBF kernel was chosen for this purpose because it is the most commonly used kernel. The study found that cloudy days had a 12.42 percent mistake rate, foggy days had an 8.16 percent error rate, rainy days had a 9.12 percent error rate, and sunny days had a 4.85 percent error rate.

Random Forest (RF) is an ensemble approach for classification and regression applications. [19]. It effectively merges K decision trees and generates a result. Huertas Tato and Centeno Brito's (2019) research focuses on estimating the photovoltaic output of six solar panels in Faro (Portugal) [20]. Temperature, meteorological factors, and radiation are all aspects in the study. Three years of data were used, with minute intervals. The most critical hyper parameter in the RF method is the number of trees, which is set to 500 via brute force. Finally, it is mentioned that the type of solar plant module influences the effectiveness of the RF approach, and that the performance can be improved by using complicated trend analysis, more relevant data, and larger prediction intervals.

Nearest Neighborhood algorithms are supervised learning algorithms that make local approximations and discover the closest data in order to find an output. The output is computed by applying weights to the determined number of closest neighbors. Voyant, Paol, Muselli and Nivet (2013), did research to evaluate the performance of forecasting systems for estimating photovoltaic production over various time frames [21]. The study employs k-NN because it is necessary to use naive models to evaluate the relevance of complicated models. As a consequence of the research, it was determined that k-NN is a suitable alternative for forecasting on a daily basis. According to the findings of the study, k-NN is not the best model for hourly forecasting.

Tibshirani (1996) proposed a new regression approach in which the sum of squares is minimized while the total of absolute values of coefficients is smaller than a constant [22]. It produces a sparse solution with a large number of zero coefficients. This regression employs the L1 norm in algebra and is also known as "LASSO" (Least Absolute Shrinkage Selector Operator). The model is applied to data on prostate cancer in three different scenarios. The first situation involves data in which a small number of attributes have a substantial impact. The second situation involves data in which a small to moderate number of attributes have a moderate influence, whereas the third scenario involves data in which a big number of attributes have a minor effect. In these circumstances, three approaches are compared: LASSO, Ridge Regression (L2 Regression), and Subset Selection. Subset Selection performs best in the first case, whereas the other two models perform badly. In the second case, LASSO is the most effective, and in the third, ridge regression and lasso are the most effective. It may be inferred that LASSO regression is a viable solution in problems similar to the second scenario.

Luo et al. (2018), provide three novel regression models to address data integrity issues that have been identified in previous models [23]. Under-forecasts are the most common cause of data integrity issues, as they might result in blackouts. The study's main goal was to address sudden anomalies in power load data rather than to solve aspects like trend, time variables, and temperature in the provided models. Two of the models were Iteratively Reweighted Least Squares (IRLS) models, in which anomalies were defined as observations with large residuals, and forecasting these anomalies contributed less to the objective function than precision on standard data points. The first IRLS technique assigns minimal weights to residuals with large values, whereas the second IRLS approach removes residuals above a threshold. The third model is an L1 regression model, which produced accurate findings even when 30% of the data contained anomalies, with an accuracy of roughly 10%, because the L1 model was less susceptible to outliers.

For a case study of short-term electrical load forecasting in Turkey, Ishik et al. (2015) constructed a feed-forward neural network [24]. The authors of this study concentrated on forecasting short-term electricity consumption. Ishik et al. used the Levenberg-Marquardt algorithm to train a feed-forward neural network to estimate the next day load in Turkey's power market. Hour, day of week, month, year, temperature of cities, and power load are all included in their data file. The input variable has six units, while the study's output variable is hourly electricity load. The data samples from 2012 weekdays were randomly divided into three groups: 70% training, 15% validation, and 15% test. The authors divided the network among seasons in Turkey and used SVM to compare it to their own network. The findings of their research demonstrate that the neural network's overall accuracy is comparable to that of SVM, and that the neural network produces better results for winter and spring data, while Support Vector Machine forecasts perform better for summer and fall. The MAPE for each season ranges from 2.0 to 3.7.

In recent studies, researchers have used ensemble models which involves combination of neural network models and conventional models such as Auto Regressive Integration Moving Average and Support Vector Regression (SVR) [25], [26], [27], [28]. When the conventional models are compared with neural network models, the neural network models outperform in load forecasting [29]. Using combined hybrid model produces better results than individual models (SARIMA and LSTM) because of linear and nonlinear data estimation [30], [31].

With this study, discretization model that is not applied in load forecasting literature is used. Ensemble of a discretization model and deep neural network model created a robust load forecasting model with high accuracy rate. In addition to that novel features are used in the study.

3. ELECTRICITY LOAD FORECASTING

3.1. Methodology

3.1.1. Data Preprocessing

Data provided by an electricity retail company contains hourly electricity usage values from November 2016 to February 2019. In literature review it has seen that weather and climate parameters are essential to forecast electricity demand.

To be able to predict distributed electricity, parameters are decided. These parameters are;

- hours (binary), electricity usage changes by hours with people's behaviors,
- months (binary), because they add seasonality to data,
- days of the week (binary), usage of electricity differs from weekdays to weekends in general, in detail people behave different in each day,
- holy and official holidays (binary), directly affects the numbers of electricity usage,
- academic calendar (school days and holidays) (binary), with the big numbers of students not going to school, electricity usage will differ,
- weather (binary), have a big impact on human behavior and heating,
- temperature as Fahrenheit (float), cold weather may lead people to use electric heaters,
- electricity usage (float).

Weekends are not counted as official holidays or academic holidays since being specified over time adds them seasonality.

One-hot encoding is applied to hour, month, day, holiday, academic calendar and weather data since they are categorical. One-hot encoding is a method that used for turning categorical data to binary vectors. A binary vector consists of all zeros and a one in the index. This way, all categorical data has the same weight on the model.

Sliding windows method is used to add recurrence effect to model which are influenced by previous values and seasonality. Sliding window method is creating a new parameter and giving it the same value as previous value of real parameter, which window is created from. For example, if the first value for a parameter is 1, second value of window parameter is 1, first value being 0. The second value of the real parameter is equal to third value of window parameter. When sliding windows are applied, rows with no value in them occur at the first rows of data since there will be no previous value for them. These rows should be deleted to prevent errors when fitting a machine learning model to data. For weather parameter, 5 windows are used for previous hourly values and 3 windows are used for previous daily values. For each temperature parameters, 5 windows are used for previous hourly values. For electricity usage parameter, 4 daily windows, 5 daily windows, and 1 weekly window are used.

3.1.2. Model Development

For such forecasting problems, understanding the data then choosing the right models is crucial. There are lots of models to fit data in, yet in this study, the data follows a nonlinear pattern. As it is seen in figure 1, data has a pattern and it has seasonality. In order to forecast hourly consumption, models that can memorize patterns and nonlinearity are needed. It is thus decided to use deep neural networks to solve this problem.

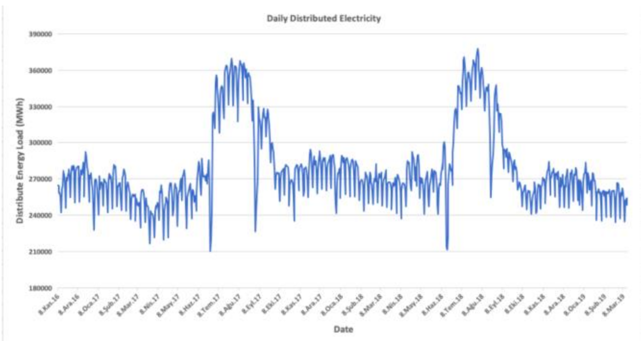

Figure 1 Daily distributed energy

3.1.3. Neural Networks

Load forecasting with traditional forecasting methods usually do not give accurate results. On account of the fact that, it is a detailed, multidimensional and multivariate forecasting problem. However, machine learning techniques especially Neural Networks are extremely good at modelling nonlinearities and the seasonal patterns in data at many fields and in various kinds of forecasting tasks. Neural networks are models that simply simulate the human brain, built to find repetitive patterns in data. Also, Neural networks have theoretically provable capability to approximate any complex functions with arbitrary precision [32]. Deep Neural networks are capable of finding correlations between current and future events. Neural networks are able to do evaluations with limitless data on timely analysis and assessment. Machine learning techniques are proactive and specifically designed for “action and reaction” industries. Essentially, systems are able to give quick response upon the outputs of the machine learning models. In that project, deep neural networks are developed to capture nonlinearities and the seasonal patterns in energy demand.

Neural Networks are composed of several layers. Layer is the ultimate key stone in deep learning. Input with weight is generally stored in a layer which is like a container. It uses a set of functions to alter the weighted inputs and then transfers the values to the next layer. The first layer of a network is known as the input layer, while the last layer is known as the output layer. All levels in between are referred to as hidden layers.

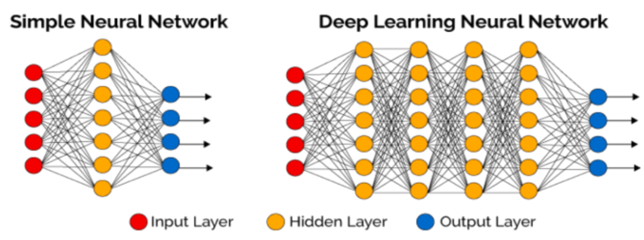

Figure 2 Neural network scheme [26]

Composition of many individual neurons form the Deep Neural Network. Inspiration comes from the architecture of a human brain for the Neural networks and as in human brain the basic building block is called a neuron. The term "artificial neuron" refers to a mathematical function. It accepts inputs, multiplies the weights, and adds them all up. The value is then given to a non-linear function known as an activation function, which transforms it into a neuron's output.

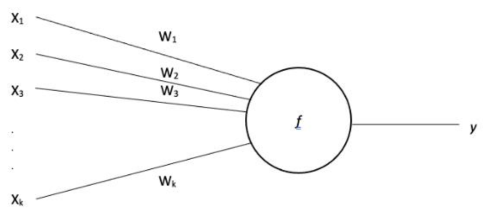

Figure 3 Neural network neuron

A signal's X value is present when it arrives at a node. This X value is passed to the activation function, which returns a value. When inputs are multiplied with weights in a neural network without activation functions, the outputs still fluctuate. However, in this case, the problem is that the model would be linear. Neural networks can handle complex models because they can implement nonlinear characteristics in the model through activation functions. Non-linearity is important because it is best suited for large and complex data sets. An exclusive activation function can be formed by anyone, yet there are some famous ones that can achieve the most out of the models. They are linear, ReLu, SeLu and ELu

The linear activation function is a simple function. It is a line function in which the activation is proportional to the input. Because it is a constant, in backpropagation, the change is constant. In this way, the changes are no longer dependent. Linear activation functions are usually not considered as preferably good.

The shortened version of rectified linear unit is known as ReLu. It captures interactions and nonlinearities very well.

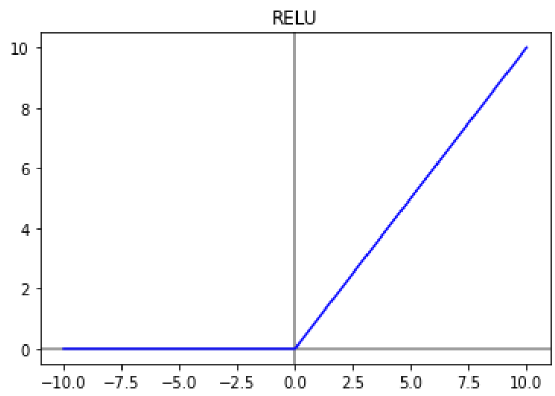

Figure 4 RELU activation function

Elu is exponential linear unit function. It is so similar to ReLu. Elu also avoids vanishing gradient. The difference of ReLu and Elu is that Elu has negative values. Negative values push mean activations closer to zero. Zero means makes learning faster and it works like a regularization method. They bring gradient closer to natural gradient and prevent overfitting.

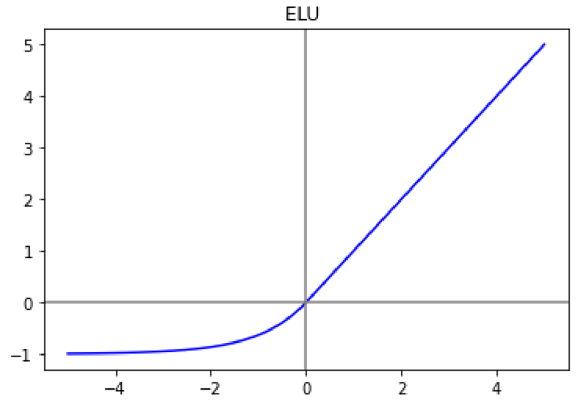

Figure 5 ELU activation function

The optimizer updates the model in response to the output of the loss function to shape and mold the neural network as accurately as possible.

Adam is an adaptive learning rate method which means, it computes individual learning rates for different parameters. Adam optimizer uses estimation of first and second moments of gradient to adapt the learning rate for each weight of the neural network. The method is computationally efficient. It has little memory requirements and is well suited for problems that are large in terms of data and parameters.

Adagrad uses frequency of features the update. It updates more frequent features. Adagrad is a stochastic gradient descent optimizer.

A loss function is an objective function that machine learning models attempt to maximize or minimize through its learning steps. The output of a loss function measures the accuracy of models.

Mean squared error (MSE) is that the sum of squared distances of data points to the regression curve. Squaring has two aims, one to extend the impact of distance and another is to get rid of negativity.

Mean absolute percent error (MAPE) is that the most used loss function in regression problems. It is used for measuring the prediction accuracy level of forecasting methods.

Epoch number implies how many times the weights will be updated. In neural networks, weights and inputs run through layers and nodes once. After the output layer, a loss function value is calculated, then weights are updated and run through again. These iterations are named epoch. Less epoch numbers may cause underfitting, more epoch numbers may cause overfitting. Thus, an optimal number is required.

3.1.4. Hyperparameter Optimization

A neural network consists of several hyperparameters. These are layer numbers, node numbers, epoch numbers, batch size numbers, activation function types and optimizer types. These parameters are highly affecting the algorithms and the results of a network. It is almost impossible to have an insight about the impact of selected parameters. In this project, models with 1 to 5 layers are applied. As number of neurons, 100, 200, 300, 500 and 1000 were applied to the models. 100, 200, 300, 400 epochs, 64, 128 and None batch sizes are applied. Relu, Elu and linear activation functions are used. As optimizer, Adam and Adagrad is used. MAPE is evaluated as loss function result. When calculated, 1800 different models will have to be evaluated. Since this process is time and energy consuming, there are hyperparameter optimization methods.

Grid search is one of the optimization methods. Grid search goes through all possible combinations with the given parameters. It builds the models and evaluates them sequentially, then returns the parameters that gives the best result. Grid search is not favorable for hyperparameter optimization because it is heavily time consuming and it requires strong devices for computation.

Random search is another method that takes given parameters randomly to build and evaluate models. Heuristically, after 30 iterations, random search is better than finding the best parameters. It consumes less time than grid search. In this project, random search method is used to achieve best hyperparameters.

Since a strong computer is needed in order not to lose time in the random search, one of the servers of AGU (Abdullah Gül University) is used. In the server, five python codes for layers from 1 to 5 is run in five different terminals.

The parameters which give best accuracy in the random search were 5 layers, 100 nodes, 400 epochs, batch size of None, “Elu” activation function and “Adagrad” optimizer. Mean validation score was 2.9 with 0.5 standard deviation. Mean test score was 2.18 when 60 days are forecasted by the model.

A deep neural network includes randomness in different stages of the algorithm. First weights are created randomly and in every run they change, thus the result changes. Reproducibility is essential for a model because it helps keeping track of the changes in hyperparameter optimization, knowing that the result only changes due to parameters. Reproducibility is achieved by using the seed function in python as well as a weight initializer for first layer. Seed function produces a sequence for a random number generation method so that python uses that same sequence every time the code runs. It is the same for weight initializer. Weight initializing has its own random number generation method. Having the same weights in first layer produces same results in every run.

3.2. Results

Best parameters of the deep neural network which give the highest accuracy rate for the hourly electricity consumption in a city are found. Model is run to forecast 60 days interval from 14 January to 14 March. Predicted values and actual values are compared and the mean absolute percentage error is calculated to be 2.18. It means that the mean of all the errors is 2.18 and the accuracy of the model is 97.82%.

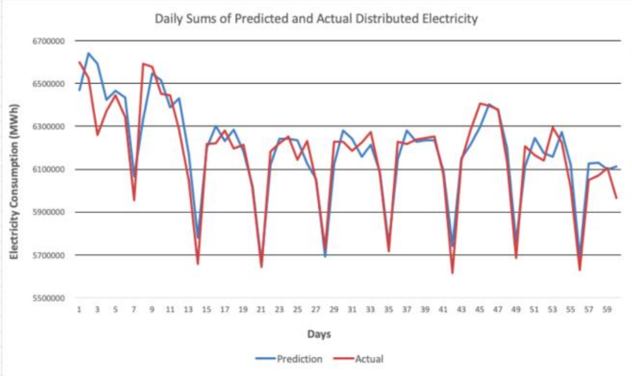

Figure 6 Predicted and Actual daily electricity distribution amounts

Figure 6 shows the daily predicted values and actual values. As it can be seen, even if at some days model fail to catch actual values, in general perspective, the model is able to detect the patterns and the trend.

If the results are inspected as hourly based, hourly errors and accuracies can be seen in the Table 1. Model is very strong on forecasting the first hours and the last hours of a day. As it goes to the mid hours, the accuracy is going down. Yet the worst accuracy is 96.12% at hour 16 which is still in acceptable range of the company so it can be still seen as a good result.

Table 1 Hourly accuracy values

Hour	Accuracy	MAPE	Hour	Accuracy	MAPE
0	98,65	1,35	**12**	97,84	2,16
1	98,48	1,52	**13**	97,74	2,26
2	98,24	1,76	**14**	97,31	2,69
3	98,43	1,57	**15**	96,49	3,51
4	98,15	1,85	**16**	96,12	3,88
5	97,94	2,06	**17**	96,43	3,57
6	97,8	2,2	**18**	96,86	3,14
7	97,87	2,13	**19**	98,03	1,97
8	97,97	2,03	**20**	98,49	1,51
9	97,94	2,06	**21**	98,2	1,8
10	97,76	2,24	**22**	98,35	1,65
11	97,39	2,61	**23**	98,71	1,29

If results are inspected as daily based, it is seen that Mondays have the least accuracy with 96.70% and the highest error rate with 3,30. In the rest of the days, the error rates are between 1.70 to 2.31 which outperforms.

Table 2 Daily accuracy values

Days	Accuracy	MAPE
Monday	96,7	3,3
Tuesday	97,82	2,18
Wednesday	97,69	2,31
Thursday	98,3	1,7
Friday	98,25	1,75
Saturday	97,98	2,02
Sunday	97,98	2,02

In order to see the improvements in the results, the accuracy and error rates of the company should be examined. In the company's experience-based forecasting method, the error rate is 2,254 in 60 days interval from 14 January to 14 March. In the developed model in that project, the error rate is 2,18 %. While the accuracy of the company's forecasting method is 97,75 %, in the model it is 97,82 %. As it can be understood, the developed model in that project is better than the company's experience-based forecasting method.

When the results compared with respect to hours, the model gives less error rate than the company's forecasting method in 15 hours during a day (See Appendix 1).

When the results compared with respect to days of a week, the model gives less error rate than the company's forecasting method in 6 days during a week (See Table 3).

Table 3 Daily MAPE comparison with the forecast of the company

Days	MAPE (Company)	MAPE (Model)	Success
Monday	2,34	3,3	0
Tuesday	2,26	2,18	1
Wednesday	2,42	2,31	1
Thursday	2,05	1,7	1
Friday	2,27	1,75	1
Saturday	2,42	2,02	1
Sunday	2,42	2,02	1

3.3. A Novel Approach

Torgo and Gama (1996) focus on applying classification methods on regression problems in their study. Main aim is to discretize a larger problem into the smaller ones, and approach the solution in that way. In the examples that are used in literature, datasets are discretized in three approaches. These approaches are splitting the values in equal probable intervals, equal width intervals, and using K-NN algorithm to create clusters. After discretizing the values, usually median values of these discretized sets are taken as the outputs.

In the methodology proposed for this study, discretizing is done iteratively on the training set. In the first step, the whole dataset is split around the value of mean. The tuples which have the output greater than the mean are labelled as “”one” and the tuples which have output lower than the mean are labelled as “zero”. This is done iteratively on the sets until a threshold point. This threshold point is determined based on the preferences of the decision maker and called “Lower Threshold”. There is also “Upper Threshold”, which is basically the two times bigger version of the lower threshold. According to the ratio proposed for this methodology, decision about splitting is determined.

After a splitting operation is done, ratio of the newly built dataset is checked. If the ratio is bigger than the upper threshold, according to the determined labels splitting is done again. If the ratio is between upper and lower threshold, a regression model is applied on the dataset. If the ratio is lower than the lower threshold, Gaussian white noise with a mean of 3 is added into the samples. Figure 7 below shows the process diagram of this iterative methodology.

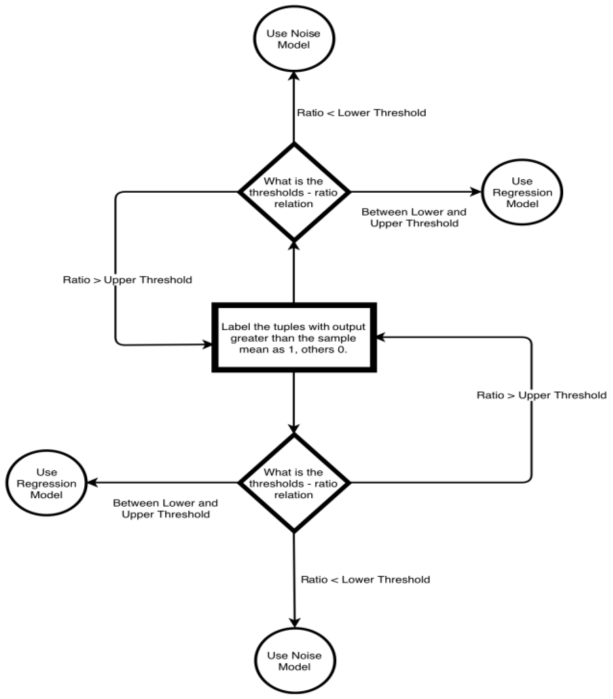

Figure 7 Subprocesses for the elimination in the novel approach

When this methodology is applied to the dataset of this study, 7 final subclasses are built. In order to learn the value of a new sample, there are 13 different ML models integrated. Classifications in this methodology are done via Random Forest, and regressions are done via MLP (Multi-Layer Perceptron). Appendix 2 shows the subclasses after the splitting operations. Figure 8 below is the process flow chart for new coming data according to dataset processed in this study. Currently, success of this methodology based on MAPE values is around 2.5 %. In the next steps, sub models will be optimized in order to obtain a better result.

3.4. Ensemble Model

In order to increase the forecasting performance deep neural network and discretization (novel) models are ensemble. The individual models are compared in terms of MAPE (See Table 5).

Table 5 Comparison of the models

Hour	Novel Method MAPE	ANN Method MAPE	Hour	Novel Method MAPE	ANN Method MAPE
0	1,47	1,35	**12**	3,22	2,16
1	1,69	1,52	**13**	3,85	2,26
2	1,35	1,76	**14**	2,2	2,69
3	1,85	1,57	**15**	2,9	3,51
4	2,25	1,85	**16**	3,2	3,88
5	1,92	2,06	**17**	3,45	3,57
6	2,7	2,2	**18**	3,07	3,14
7	2,5	2,13	**19**	2,42	1,97
8	2,3	2,03	**20**	1,49	1,51
9	2,32	2,06	**21**	2,15	1,8
10	2,45	2,24	**22**	2,76	1,65
11	3,72	2,61	**23**	1,52	1,29

With this ensemble model, we built more robust model. The accuracy results of the ensemble method are given in the Table 6.

Table 6 MAPE values of the ensemble model

Hour	Ensemble Model MAPE	Hour	Ensemble Model MAPE
0	1,35	**12**	2,16
1	1,52	**13**	2,26
2	1,35	**14**	2,2
3	1,57	**15**	2,9
4	1,85	**16**	3,2
5	1,92	**17**	3,45
6	2,2	**18**	3,07
7	2,13	**19**	1,97
8	2,03	**20**	1,49
9	2,06	**21**	1,8
10	2,24	**22**	1,65
11	2,61	**23**	1,29

The average accuracy of the ensemble model is 2,09 % so that the best model is obtained by the ensemble model.

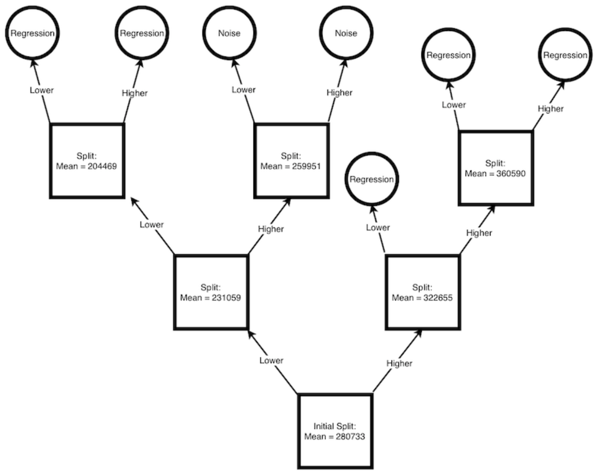

Figure 8 Prescription of the novel approach

4. CONCLUSION

Due to the competitive and dynamic nature of electricity industry, electricity load is highly critical for companies to be able to compete. Within the scope of this study, forecasting models are developed to predict hourly electricity load of a city.

In that study, Neural network are applied, trained and tested with the dataset. A novel method which include discretization of regression problem is developed. These two models are ensemble. Mean absolute percentage error is used as key performance indicators and models are assessed accordingly. The ensemble model develop for the load forecast predicts the consumption of electricity with a 97,91 % accuracy.

Funding

"The authors has no received any financial support for the research, authorship or publication of this study."

The Declaration of Conflict of Interest/ Common Interest

“No conflict of interest or common interest has been declared by the authors”.

Authors' Contribution

Under this heading, "The authors contributed equally to the study" or "The first author contributed 60%, the second author 40%." expressions such as should be included.

The Declaration of Ethics Committee Approval

"This study does not require ethics committee permission or any special permission"

The Declaration of Research and Publication Ethics

“The authors of the paper declare that they comply with the scientific, ethical and quotation rules of SAUJS in all processes of the paper and that they do not make any falsification on the data collected. In addition, they declare that Sakarya University Journal of Science and its editorial board have no responsibility for any ethical violations that may be encountered, and that this study has not been evaluated in any academic publication environment other than Sakarya University Journal of Science.”

APPENDICES

Appendix 1 Hourly MAPE Comparison with The Forecasts of the Company

Hour	MAPE (Company)	MAPE (Model)	Success	Hour	MAPE (Company)	MAPE (Model)	Success
0	1,80	1,35	1	12	2,26	2,16	1
1	1,46	1,52	0	13	2,46	2,26	1
2	1,49	1,76	0	14	2,67	2,69	0
3	1,70	1,57	1	15	2,76	3,51	0
4	1,85	1,85	1	16	3,03	3,88	0
5	1,79	2,06	0	17	3,20	3,57	0
6	1,90	2,20	0	18	3,50	3,14	1
7	2,18	2,13	1	19	2,36	1,97	1

8	2,77	2,03	1	20	1,87	1,51	1
9	2,90	2,06	1	21	2,02	1,80	1
10	2,77	2,24	1	22	2,07	1,65	1
11	2,43	2,61	0	23	1,69	1,29	1

Appendix 2 Subclasses for The Novel Method

	Min	Max	Mean	Median	Range	Data Points	Half Range Ratio	Decision
1	280741	440811	322655	312018	160070	8213	28,50848291	Classify
0	149404	280732	231059	228064	131328	6932	43,95063051	Classify
1_1	322667	440811	360590	351965	118144	2990	18,3074191	Classify
1_0	280741	322635	300950	300783	41894	5225	7,461325563	Regression
0_1	231060	280732	259951	262438	49672	3211	10,74872328	Classify
0_0	149404	231014	204469	205528	81610	3621	27,31185243	Classify
1_1_1	360682	440811	391171	387653	80129	1312	11,10798432	Regression
1_1_0	322667	360573	336698	334615	37906	1679	5,873857568	Regression
1_0_1	300963	322635	310938	310752	21672	2584	3,600442579	Noise
1_0_0	280741	300948	291161	291479	20207	2637	3,598868708	Noise
0_0_1	204471	231014	216079	215027	26543	1898	6,490651486	Regression
0_0_0	149404	204455	191688	193804	55051	1724	18,42353618	Classify
0_0_0_1	191732	204455	198526	198737	12723	984	3,3179459	Noise
0_0_0_0	149404	191678	182582	184709	42275	740	14,14774083	Regression

REFERENCES

[1] D. I. Stern, P. J. Burke & S. B. Bruns, “The Impact of Electricity on Economic Development: A Macroeconomic Perspective”, EEG State-of-Knowledge Paper Series, 1.1, 2017.

[2] About the U.S. Electricity System and its Impact on the Environment, https://www.epa.gov/energy/about-us-electricity-system-and-its-impact-environment

[3] Energy Exchange İstanbul, Elektrik Piyasaları, https://seffaflik.epias.com.tr/transparency/

[4] CRO Forum, “Power Blackout Risks, Risk Management Options”, Emerging Risk Initiative – Position Paper, 2011.

[5] P. Szuromi, B. Jasny, D. Clery, J. Austin, & B. Hanson, “Energy for the long haul,”, 2007.

[6] V. Lara-Fanego, J.A. Ruiz-Arias, D. Pozo-Vázquez, F.J. Santos-Alamillos, & J. Tovar-Pescador, “Evaluation of the WRF model solar irradiance forecasts in Andalusia (southern Spain),” Solar Energy, 86(8), 2200-2217, 2012.

[7] IEA, “Global energy technology perspectives” International Energy Agency, OECD Publication Service, OECD, Paris, 2006a.

[8] B. Espinar, J.L. Aznarte, R. Girard, A.M. Moussa, & G. Kariniotakis, “Photovoltaic Forecasting: A state of the art,” In 5th European PV-Hybrid and Mini-Grid Conference (pp. Pages-250). OTTI-Ostbayerisches Technologie-Transfer-Institut, 2010.

[9] A. Azadeh, M. Saberi, S.F. Ghaderi, A. Gitiforouz, & V. Ebrahimipour, “Improved estimation of electricity demand function by integration of fuzzy system and data mining approach,” Energy Conversion and Management, 49(8), 2165-2177, 2008.

[10] B. Wang, N.L. Tai, H.Q. Zhai, J. Ye, J.D. Zhu, & L.B. Qi, “A new ARMAX model based on evolutionary algorithm and particle

swarm optimization for short-term load forecasting," Electric Power Systems Research, 78(10), 1679-1685, 2008.

[11] N. Amjady, "Short-term bus load forecasting of power systems by a new hybrid method," IEEE Transactions on Power Systems, 22(1), 333-341, 2007.

[12] M. Khadem, "Application of kohonen neural network classifier to short term load forecasting," In Panel Session on Application of Neural Networks to Short-term Load Forecasting, 1993 IEEE Winter Meeting, 1993.

[13] R.H. Inman, H.T. Pedro, & C.F. Coimbra, "Solar forecasting methods for renewable energy integration," Progress in energy and combustion science, 39(6), 535-576, 2013.

[14] H.B. Barlow, "Unsupervised learning," Neural computation, 1(3), 295-311, 1989.

[15] Y. Gala, Á. Fernández, J. Díaz, & J.R. Dorronsoro, "Hybrid machine learning forecasting of solar radiation values," Neurocomputing, 176, 48-59, 2016.

[16] V. Vapnik, "The nature of statistical learning theory," Springer science & business media, 2013.

[17] N. Sharma, P. Sharma, D. Irwin, & P. Shenoy, "Predicting solar generation from weather forecasts using machine learning," In 2011 IEEE international conference on smart grid communications (SmartGridComm) (pp. 528-533). IEEE, 2011.

[18] J. Shi, W.J. Lee, Y. Liu, Y. Yang, & P. Wang, "Forecasting power output of photovoltaic systems based on weather classification and support vector machines." IEEE Transactions on Industry Applications, 48(3), 1064-1069, 2012.

[19] L. Breiman, "Random forests," Machine learning, 45(1), 5-32, 2001.

[20] J. Huertas Tato, & M. Centeno Brito, "Using Smart Persistence and Random Forests to Predict Photovoltaic Energy Production," Energies, 12(1), 100, 2019.

[21] C. Voyant, C. Paoli, M. Muselli, & M.L. Nivet, "Multi-horizon solar radiation forecasting for Mediterranean locations using time series models," Renewable and Sustainable Energy Reviews, 28, 44-52, 2013.

[22] R. Tibshirani, "Regression shrinkage and selection via the lasso," Journal of the Royal Statistical Society: Series B (Methodological), 58(1), 267-288, 1996.

[23] J. Luo, T. Hong, & S.C. Fang, "Robust Regression Models for Load Forecasting," IEEE Transactions on Smart Grid, 2018.

[24] M.Y. Ishik, T. Göze, İ. Özcan, V.Ç. Güngör, & Z. Aydın, "Short term electricity load forecasting: A case study of electric utility market in Turkey," In 2015 3rd International Istanbul Smart Grid Congress and Fair (ICSG) (pp. 1-5). IEEE, 2015.

[25] W. Tan and B. Khoshnevis, "Integration of process planning and scheduling— a review," Journal of Intelligent Manufacturing, vol. 11, no. 1, pp. 51–63, 2000.

[26] S. Ai, A. Chakravorty, and C. Rong. "Household Energy Consumption Prediction using Evolutionary Ensemble Neural Network." Engineering Assets and Public Infrastructures in the Age of Digitalization. Springer, Cham, 2020. 923-931.

[27] K. Bot, A. Ruano, and M.G. Ruano. "Forecasting electricity consumption in residential buildings for home energy management systems." International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems. Springer, Cham, 2020.

[28] S. Rahman, M.G. Rabiul Alam, and M. Mahbubur Rahman. "Deep Learning based

Ensemble Method for Household Energy Demand Forecasting of Smart Home." 2019 22nd International Conference on Computer and Information Technology (ICCIT). IEEE, 2019.

[29] T. Panapongpakorn, and D. Banjerdpongchai. "Short-Term Load Forecast for Energy Management System Using Neural Networks with Mutual Information Method of Input Selection." 2019 SICE International Symposium on Control Systems (SICE ISCS). IEEE, 2019.

[30] S. Chan, I. Oktavianti, and V. Puspita. "A deep learning cnn and ai-tuned svm for electricity consumption forecasting: Multivariate time series data." 2019 IEEE 10th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON). IEEE, 2019.

[31] M. Krishnan, Y.M. Jung, and S. Yun. "Prediction of Energy Demand in Smart Grid using Hybrid Approach." 2020 Fourth International Conference on Computing Methodologies and Communication (ICCMC). IEEE, 2020.

[32] H. S. Hippert, C. E. Pedreira, R. C. Souza, “Neural networks for short-term load forecasting: A review and evaluation”, IEEE Transactions on power systems, 16 (1) 44–55, 2001.