ISSN 1330-3651 (Print), ISSN 1848-6339 (Online)
https://doi.org/10.17559/TV-20220430111309
Subject review

# An Overview of Forecasting Methods for Monthly Electricity Consumption

Sofija KRSTEV, Jovana FORCAN, Dragoljub KRNETA*

**Abstract:** Mid-term electricity consumption forecasting is analysed in this paper. Forecasting of electricity consumption is regression problem that can be defined as using previous consumption of an individual or a group with the goal of calculation of future consumption using some mathematical or statistical approach. The purpose of this prediction is multi beneficial to the stakeholders in the energy community, since this information can affect production, sales and supply. The Different methods are considered with the main goal to determine the best forecasting model. Considered methods include Box-Jenkins autoregressive integrated moving average models, state-space models and exponential smoothing, and machine learning methods including neural networks. An additional objective of the conducted research was to determine if modern methods like machine learning are equally precise in forecasting mid-term electricity consumption when compared to traditional time series methods. The performances of forecasting models are evaluated on the monthly electricity consumption data obtained using real billing software owned by the Distribution System Operator in Bosnia and Herzegovina. Mean absolute percentage error is selected as a measure of prediction accuracy of forecasting methods. Every forecasting method is implemented and tested using the R language, while data is collected from Data Warehouse in the form of total monthly consumption. The efficiency of presented solution will also be discussed after presentation of the results.

**Keywords:** electricity consumption; machine learning; mid-term load forecast; state-space models; time series data

## 1 INTRODUCTION

The economic development of a country depends on the availability of cheap and sustainable energy, especially since most of the industry depends entirely on its use. Electricity consumption forecasting plays a key role in ensuring a reliable supply of electricity to the system and reducing operational costs. In energy company operations, decisions are also made in line with electric load forecasting. It can be very helpful and economic to predict as accurately as possible in order to better plan power production in power plants, to plan energy market transactions to negotiate better energy prices, and to plan fuel and raw materials procurement. Energy consumption is also used in maintenance planning [1]. The accuracy of load forecasting is of even greater importance in deregulated electricity markets due to the financial risks of market participants. The fluctuating electricity and fuel prices, combined with the uncertainties of customer demand and solar, wind, or hydro-power generation impose significant risks for utility companies [2].

According to various publications, load forecasting can be broadly divided into three categories relative to their time horizon: short-term forecast (STLF) which covers forecasting in time horizon from one hour up to two weeks, mid-term forecast (MTLF) which covers forecasting from time horizon from two weeks to two years, and long-term forecasts (LTLF) which cover forecasting for a period longer than two years [2, 3].

All three types of forecasting are important for electric utilities because of the different usages of all three forecasting categories in servitude for different company goals and divisions [1]. Energy companies use STLF to support decision-making with data in order to better plan, produce, keep reserves, and operate but also to bring higher security and financial stability to daily operations [4]. STLF has become increasingly important in daily market operations [5] mostly because of the increase and strengthening of the competition in the energy markets [6]. MTLF is often used in power plant production planning, and it is of great importance for grid upkeep planning. MTLF is very important for countries whose power systems operate in a deregulated environment since it gives the decision makers information about market consumption behaviour which would lead to better procurement planning and negotiating precedence. Additionally, MTLF is essential for individual customers who operate structured procurement or portfolio management in deregulated markets [2]. LTLF is the basis for energy investment planning and plays a vital role in developing countries [7].

During the last 50 years, a great number of load forecasting methods have been developed. Forecasting methods are often divided into two groups: statistical approaches and artificial intelligence-based (AI-based) techniques [8]. The first group consists of naïve approaches, exponential smoothing - ETS, linear or logistic regression, methods based on time series, etc. The AI-based group includes artificial neural network (ANN) models, support vector machines - Linear and Radial, Elastic nets, tree-based methods, similarity methods, boosting, bagging, etc. In some cases, hybrid methods are developed by combining several methods [9,10]. These hybrid approaches often include domain logic in modelling. Many of above mentioned are mostly applied to STLF problems [9-16]. For example, in [9] a hybrid approach based on the wavelet transform and support vector machine is used to solve the STLF problem. In [10] the fuzzy linear regression method is used for load forecasting for non-working days and Monday consumption, while the general exponential smoothing method was used for weekday consumption forecasting. AI-based forecasting methods, Neural networks [11-13], fuzzy logic [14, 15], and genetic algorithm [16], show promising results for STLF.

Often considered to be a special category, the Kalman filtering technologies have thus far proved to be the most complex, yet reliable methods in time series forecasting [17]. Several applications can be found in [18, 19].

Compared to STLF, mid-term and long-term forecasting are rarely seen in research publications [20]. According to [21], accurate mid-term forecasting is the bigger challenge, because consumption habits are not the only factor affecting the pattern, but also random factors such as the country's political or economic decision-making and governance. It is stated in [20] that MTLF and LTLF are highly complex challenges that require not only fitting multiple models and tweaking the parameters of it but more about learning about data and focusing on analysis and data-centric thinking. It depends highly on domain experience in the utility sector, primarily energy companies, i.e., how electricity companies work, how they depend on technology breakthroughs, or even different economic or political factors or global events [20].

Also, MTLF data contains a lot of noise since those data contain information that is better seen in STLF data. For example, in the area of energy consumption, move-in, move out and temporary abandonment cannot be recognized as events, which can be clearly recognized in STLF data. On the other hand, LTLF data contains less information and can be generalized or can ignore those kinds of events in forecasting.

As seen by the authors, research gap is seen through a smaller number of papers exploring possibilities in application of different models and methods in MTLF prediction. Moreover, dataset used in training the model is the first of this kind in Bosnia and Hercegovina and can be insightful for the energy community and regional academia.

### 1.1 MTLF Overview

MTLF plays a great role in decision-making for countries whose power systems operate in a deregulated environment. The special significance of MTLF is reflected in its contribution to system reliability, as it can be used to optimize maintenance schedules. Also, an accurate MTLF is necessary for an economically viable system. Techniques developed for mid-term load forecasting that use statistical methods are considered in [22-25]. In [22], authors fit a SARIMA model to predict electricity consumption in Saudi Arabia. The SARIMA model is shown to outperform the regression and abductive network machine-learning models developed earlier on the same data. In [23], a parametric regression method based on STLF correlation is proposed for MTLF/LTLF. In [24], an autoregressive model utilizing meteorological parameters is proposed to estimate monthly demand predictions for a period of 1-year and applied as a pilot in Greek power utility. The model provides high accuracy forecasts. A semi-parametric additive model is proposed in [25] to forecast a huge number of electricity consumption series on the distribution grid in France at both the daily horizon and yearly horizon. The yearly prediction was based on monthly peaks calculated from daily peaks in network load. The model showed good and sufficient performances for the industrial perspectives. In [26], the authors suggested a combination of Bootstrap aggregating (Bagging) and time series methods for monthly consumption prediction in different parts of the world. The proposed methodology showed great performance in different parts of the world.

AI-based approaches are often used in mid-term forecasting. In [27] authors used neural network models to predict monthly load using historical temperature data for the Israel power grid. Their proposed ANN model gave better predictions compared with statistical approach methods. In [28], the authors implemented a statistical approach (ARIMA) and an AI-based (a simple neural network and fuzzy neural network) and compared the performance. The conclusion was in favour of the AI-based approach. In [29] authors proposed dynamic ANN, called DAN2, which is based upon the principle of adaptive learning for each network layer, propagating knowledge in lower layers, resulting in targeted network performance in the last layer. Forecasted results in such dynamic neural networks showed good performance which, when measured by MAPE values to be comparable, showed to be more accurate compared to traditional statistical, time series, or standard neural network models.

An overview of SVM in terms of estimations of function curve was described in [30] while load prediction using the same method was discussed in [31-35]. Various mid-term forecasting methods are developed and tested on different data sets. Some research recommends using proposed statistical methods while some show that AI-based methods are superior to statistical techniques.

According to [36], AI-based approaches are still not the first choice when predicting univariate time series. There is a lower number of publications using machine learning (ML) approaches than classical statistical ones. In [37] authors showed that classical statistical methods give better performance in terms of accuracy than machine learning methods for univariate time series and they pointed out that the reasons are still unknown and need to be defined to improve the performance of already proven powerful AI approaches [37]. In the analysis, the authors used 1045 monthly time series data from the M3 Competition. Motivated by [37], in [36], the authors did a data-centric analysis of the sample size influence on the final performance of a method. They showed that a small sample size suits better statistical approaches, while the big data analysis is stronger in AI-based approaches [36].

In the last three years (2018-2022), focus of the academia stayed the same, Neural networks and fuzzy approach dominate published papers in the AI-based methods, while exponential smoothing dominates statistical methods. Moreover, published works now tend to combine several methods or approaches to get better performance or accuracy. In [38], long-term prediction is presented with fuzzy Bayesian method combined with expert prediction. Holt-Winters exponential smoothing was enhanced by fruit-fly optimization algorithm in [39] and achieved better performance in terms of training and execution time, while in [40], same method was applied with extreme learning machine with same benefits in short term prediction. Extreme learning machine has been applied also in [41] with online search data which lead to improving accuracy. In [42], three ensemble learning models are developed and the respective results compared: gradient boosted regression trees, random forests and an adaptation of Adaboost, which showed best accuracy in short term prediction. In [43] ARIMA model with SVM was combined to achieve higher accuracy on targeted model comparing to single model application. As for the Neural network implementation, LSTM dominate most cited papers in the last three years, such as [44-47] and [48].

A concise overview of different forecasting methods is given in Tab. 1.

**Table 1** Different forecasting methods for MTLF

| Statistical methods | AI-based methods | Hybrid methods |
|---|---|---|
| 22, 23, 25, 28, 36, 37, | ANN: 24, 27-29, 44-48<br>ML: 36, 37<br>Fuzzy: 28,38<br>SVM: 31-35, 43 | 26, 39, 40, 41, 42 |

In this article, the standard time series methods with the modern machine learning methods are compared for MTLF. This paper includes an analysis of twelve methods. Used classical time series models include season naïve, exponential smoothing models, ARIMA models, and structural models. Regarding machine learning methods, this article considered the following: Linear regression, Elastic net, K-Nearest neighbours, Random forest, Extreme gradient boosting machine, Support vector machines (linear and radial), and finally the neural network (neural network autoregression). All of the previously mentioned models are used for regression problems, and wide number of approaches is used intentionally to explore possible good results of high accuracy with, until now, not published approach in this domain.

The performances of forecasting models are evaluated on the monthly electricity consumption data for the period from January 2000 to March 2020. The first 228 months are used for training, while the last 15 were used for testing and validation. The considered models calculate the forecast for the whole test period from January 2019 to March 2020.

A comparative analysis of such a large number of methods has not been done so far for MTLF. As emphasized earlier a limited number of research papers deal with mid-term forecasts of electricity consumption, so this analysis could be very useful, especially for Distribution System Operators. The electricity demand data considered in this research are obtained from the Distribution System Operator (DSO) "Elektrodistribucija" Pale, Bosnia and Herzegovina which is in the final stage of the process of division distribution and supply. So, the results of this paper can give significant support to the electric companies that start to operate in deregulated electricity markets.

## 2 MATERIALS AND METHODS

Bosnia and Herzegovina has been an open electricity market since January 1st, 2015.

Since then, all buyers have a free choice of choosing an electricity supplier. Supply of the market with electricity is done by energy stakeholders who are registered for trade and supply of the Bosnian market with electricity. Supplier is a name for a utility provider with the permit of electricity supplying activity to tariff buyers and an entity with the permit of trading and supplying electricity issued by the regulatory body.

Distribution System Operator (DSO) is a utility company focused on the activities related to the distribution network. All countries in the region have opened the market of the electricity for legal trade and have divided distribution and supply. In Bosnia and Herzegovina, the process of division distribution and supply is in its final stage.

### 2.1 Data Set Used

Data used for training, testing, and validating models are obtained from billing software used in the Distribution System Operator "Elektrodistribucija" Pale, Bosnia and Herzegovina. DSO Pale currently has more than 60 000 measuring points. Data consists of monthly electricity consumption of active energy (in kWh) of high tariff for metering points in a time frame from January 2000 to March 2020. Data is collected from their Data Warehouse, in the form of total monthly consumption. First, data is explored to find outliers. For outliers' detection, Loess' method for decomposition of seasonality and trend was used [49], and several points were detected and eliminated. After the elimination of the outliers, the missing values are imputed with linear interpolation. Fig. 1 shows the monthly electricity consumption data for the period from 2000 to 2018. Detected outliers are presented with dashed lines.

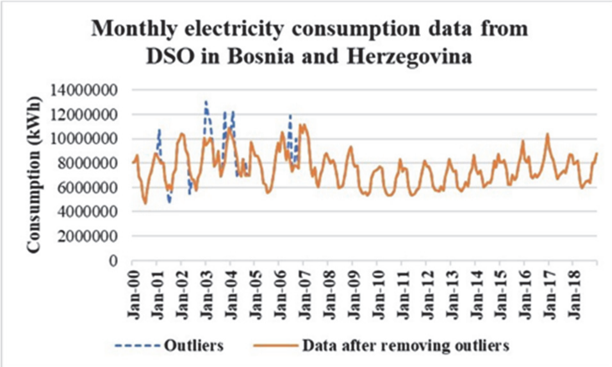


**Figure 1** Outliers detection on a monthly basis from 2000 to 2018

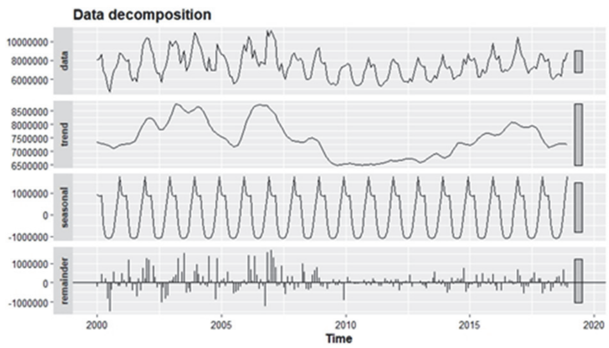


**Figure 2** Decomposition of data into a trend, seasonal, and remainder component

Data decomposition shows the existence of trend and seasonality components in time series, as shown in Fig. 2. The positive trend slope is visible from 2010 to 2017. There is a sudden drop in electricity consumption in 2017, which makes precise forecasting of future electricity consumption very challenging. The changes in seasonality over time can be seen clearly in Fig. 3. From January consumption is decreasing, and from September consumption is increasing. The minimum electricity consumption is reached in June, July, and August.

To select the optimal forecasting model, we divide the data set into two parts: train data is data from 2000 to the end of 2018. Test data is data of energy consumption for 2019 and the first quartile of 2020.

In the following, the methods and their implementation used for mid-term forecasting on a given data set are described.

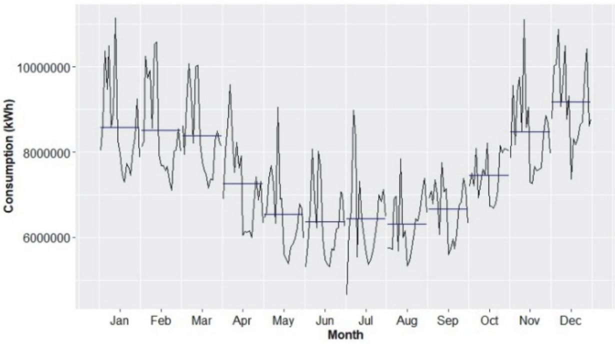


**Figure 3** The seasonal plot of monthly electricity consumption data obtained from DSO in Bosnia and Herzegovina

## 3 CLASSICAL TIME SERIES FORECASTING MODELS

### 3.1 Seasonality-Based Naive Method

This approach is often used for data that is highly seasonal [49]. The method is based on the value of the same season part in the previous cycle. By $y_1$, $y_2$, ..., $y_T$ denote historical data and by $y'$ forecasted data. The prediction for the time $T + h$ can be written as $y'_{T+h|T} = y_{T+h-m(k+1)}$ where $m$ is the seasonal part, $h$ is horizon and $k$ is the fraction of $(h - 1)/m$, meaning the number of whole time fractions on which method is based before $T + h$. A seasonal naive forecasting method is implemented in this way in using the function snaive() from the R package called "forecast". A seasonal naive forecasting method is often used as starting point, i.e. the basis, to evaluating any other algorithm. To produce forecast point, the function snaive() would repeat the value from the same period a season ago, for example, to produce forecast for January 2021, the method would take the value January 2020.

### 3.2 Box-Jenkins ARIMA Models

Univariate time series data is most often predicted with ARIMA models. It is based on the description of historical data of a single variable and describing the autocorrelation in the data. The non-seasonal model ARIMA($p$, $d$, $q$) is described with the following three parameters $p$ - the number of autoregressive terms, $d$ - the degree of nonseasonal differences needed, and $q$ - the number of lagged forecast errors. Since ARIMA models ($p$, $q$, $d$) cannot be used for data without a seasonal component, a seasonal ARIMA model (SARIMA) ARIMA($p$, $d$, $q$)($P$, $D$, $Q$)$_m$ is formed to support time-series data with a seasonal component. It contains three new parameters ($P$, $D$, $Q$) to specify the part of autoregression, differencing, and moving average, and it also adds a parameter m which represents the seasonal period [49]. In this paper, the Box-Jenkins approach is used for model definition and training. The procedure consists of data transformation and an iterative process of model identification, model estimation, model checking, and forecasting.

**ARIMA model selection**

To obtain the optimal ARIMA model the auto.arima function from the earlier mentioned "forecast" package with Box-Cox transformation in R is used. Auto.arima is based on the Hyndman-Khandakar algorithm [50], which returns a model with the smallest AIC. The model obtained by the function is ARIMA(1, 0, 1)(1, 1, 1).

### 3.3 State-Space Models and ETS

ETS (Error, Trend, Seasonal) method is another approach for forecasting univariate time series data. It is an innovation state-space model described in [51] for every exponential smoothing method. Predictions based on the exponential smoothing methods are formed on averages of previous records but weighted based on their "age" - older records have lower significance to the prediction than the newer ones [49]. The exponential smoothing method can be represented as a pair of trend and seasonality descriptors, which can be non-existing ($N$), additive ($A$), multiplicative ($M$), or damped additive/multiplicative ($Ad$/$Md$), giving 15 different methods [50].

**ETS model selection**

The function ets() from the forecast package is used, with Box-Cox transformation, which implements a state-space modelling framework. The model chosen via algorithm for our data sets is ETS($A$, $N$, $A$).

### 3.4 Structural Time Series Models and the Kalman Filter

The Kalman filter refers to an algorithm for time series data with state space model descriptors, which recursively predicts and calculates variance with the strength to predict in real-time, i.e., to predict a value for the next time point at any part of the model based on the state-space descriptors [52].

In this paper, the Kalman filter is applied to the BSM model, which is a basic structural time series model, which consists of mutually independent trend, seasonal, and noise components. Implementation is performed using the R language.

The structural time series model is a state-space model by which time series can be decomposed into components: trend, cycle, seasonal, and noise. Each of these components is considered a different space and can be analysed separately.

Compared to the ARIMA model, where trend and seasonal components are removed by applying differencing on data before the analysis, the structural time series model allows getting the specific information and identifying any characteristics in each explicitly formulated component. This is the main advantage of the structural model.

## 4 MACHINE LEARNING METHODS

Predictions on time-series-based data are a supervised learning problem. The process of observing different time frames and applying different models opens the possibility of describing parts of data. The process of training uses predictions on frames and creates updates based on the overall or partial result when compared to real data, which is standard for supervised learning problems. Training of the models is stopped when targeted performance is achieved [53]. Supervised learning problems are divided into regression problems, where an expected prediction is an integer-based number and a model is a function with one numerical output, and classification problems, where expected prediction is a true-false/category-based output. In order to create a supervised learning problem in time-series data, a previous value can be used as output for future time points. After choosing the variable for prediction, the feature engineering can define that is informative enough to make a good description of future time steps [53].

In this paper the class of lag features is used, that is, the values at prior time steps. The purpose of lag features is re-framing of time series data in order to observe previous timeframes as features with the same length and, possibly, information. The width of the time-frame window depends on the number of created lagged features [53]. For our data set 24 lagged variables are created. Denote them by: $t - 1$, $t - 2$, …, $t - 24$. To estimate the relative usefulness of input features two methods are used, the linear method (lm) and Principal Component Analysis (PCA). After feature selection, the following machine learning methods are applied: Linear regression, Elastic net, K-Nearest neighbours, Random Forest, Extreme gradient boosting machine, and Support vector machine. The description of these methods can be found in [54].

When using linear method, student t-test shows significance of $t - 1$, $t - 3$, $t - 12$, $t - 22$. Thus, those values will be preselected as additional attributes. To select relevant features by PCA method the preProcess function from the caret package in R is used. The package caret (classification and regression training) is used as a tool for developing initial forecasting models, with the possibility of tweaking the models based on performance and available model parameters. The goal of the caret package is fast model setup, training, and modification, with the purpose of exploring different options regarding modelling. Caret also covers pre-processing, component analysis, feature selection, and model visualisation.

The *train* function from the caret package is used to evaluate the effect of model tuning parameters on performance, choose the optimal model across these parameters, and estimate model performance from a training set using resampling. For the train function from the caret package, the possible resampling methods are bootstrapping, k-fold cross-validation, leave-one-out cross-validation, and leave-group-out cross-validation [55]. To modify the resampling method, a trainControl function is used. The option method controls the type of resampling. For building some models timeslice method is used, that is time series cross-validation which is also known as "evaluation on a rolling forecasting origin" [56]. Some of the methods, such as Random Forest or Linear regression diverged when using cross-validation, so the resampling needed to be done manually. When using the timeslice method, parameters initialWindow and horizon were set in a way that has given the best results on the training set. The starting value of repeating values in both training and test set are set in favour of the training set results. With a window of 5 years of historical data (60 lags) and yearly prediction, training of the models gave the best results.

**Neural networks**

Neural networks are currently the most popular methods for short-term consumption prediction. It has proven to be highly accurate when dealing with a high number of records. Neural networks are an artificial concept based on the human brain. It can be observed as a though interconnection of "neurons", brain cells in brain layers that propagate information between themselves and further to the surface. The input neurons form the bottom (first) layer, and the prediction forms the top layer consisting of multiple neurons in a classification problem, or a single neuron in regression problem modelling. Often, there are multiple hidden layers, consisting of a network of neurons used for propagating the knowledge accumulated in the previous layers of the network [49].

The basic neural network model does not have hidden layers and behaves as a linear regression function with weights on each input neuron. The weights (the coefficients attached to predictors) are calculated in the training process and are based on the learning algorithm and the cost function, which ought to be minimized in the training process [49]. A multilayer perceptron, or MLP, is a neural network with at least one hidden layer, where multiple layers of neurons are interconnected in a way that each layer of nodes receives inputs from the previous layers. That approach to information sharing is called a feed-forward network.

The neural network model where along with the time series data the lagged values of the data are added as input is called a neural network autoregression or NNAR model. In this paper, a three-layer NNAR model for seasonal time series data prediction is considered. A notation for the model is NNAR $(p, P, k)_m$, where $p$ and $P$ are the numbers of non-seasonal and seasonal lagged inputs, respectively, and k is the number of neurons in the hidden layer. [38]. In this research to fit NNAR($p$, $P$, $k$) $m$ model the function nnetar() function from package forecast in R is used. The values of parameters $p$ and $P$ are selected automatically. The parameter $P$ is set to 1 by default, $p = 3$ is chosen from the optimal linear model fitted to the seasonally adjusted data and the number of neurons in the hidden layer is set $k = 6$.

## 5 RESULTS

This section presents the forecasting results obtained by using different models on time series data obtained from DSO in Bosnia and Herzegovina. Models are applied and tested to find the optimal forecasting model for this type of data. The electricity consumption for the period from January 2019 to March 2020 is predicted on the monthly resolution with historical data from 2000 to 2018. In the analysis of the forecasting results, the Mean absolute percentage error (MAPE) is measured.

### 5.1 Forecasts from the Classical Time Series Methods

The forecasting results obtained by the seasonal naïve model, ARIMA(1, 0, 1)(1, 1, 1) model, ETS($A$, $N$, $A$) model, and BSM are shown in Fig. 4. It can be noticed that classical time series methods ARIMA, ETS, and BSM perform slightly better than SNAIVE.

The best accuracy according to MAPE is achieved with the ETS with MAPE 3.28 %. MAPEs for ARIMA(1, 0, 1)(1, 1, 1) and BSM are 3.36% and 3.87%, respectively. The seasonal naive model gave MAPE 4.16%.

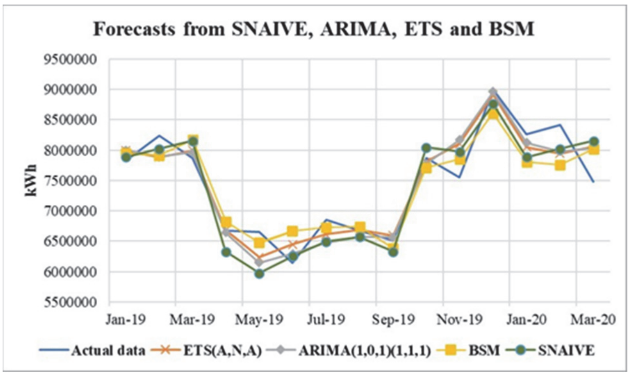


**Figure 4** Forecast from the SNAIVE, ARIMA model, ETS model, and BSM

## 5.2 Forecasts from the Machine Learning Methods

Fig. 5 and Fig. 6 show the forecasting results obtained by Machine learning methods (Linear regression, Elasticnet, KNN, Random forest, XGBM, Linear SVM, and Radial SVM) using the lm and PCA feature selection method, respectively.

From these figures, machine learning methods are competitive to classical methods. The results obtained using PCA feature selection are slightly better than the corresponding results obtained using lm feature selection. Tab. 2 presents MAPE comparison for machine learning methods which are used for prediction after lag feature selection.

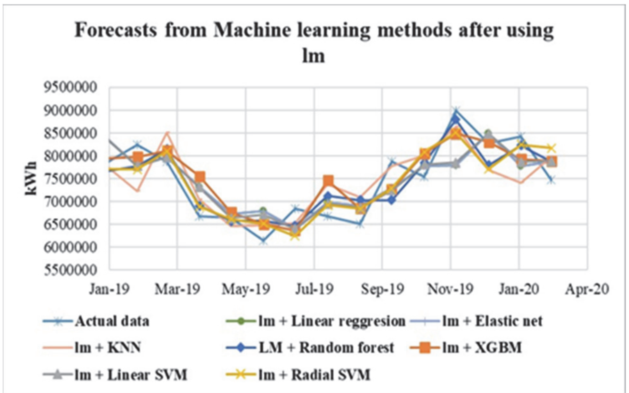


**Figure 5** Forecast from Machine learning methods after using lm method for feature selection

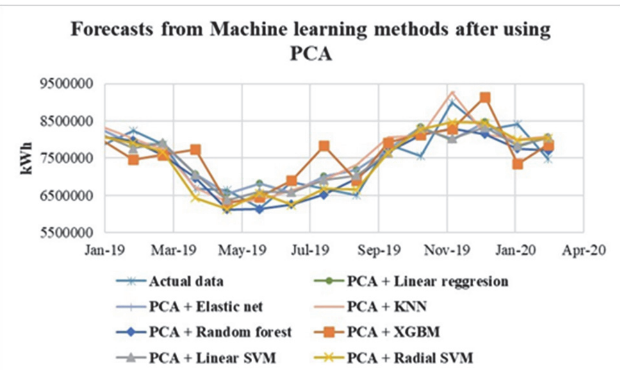


**Figure 6** Forecast from Machine learning methods after using PCA method for feature selection

**Table 2** Accuracy comparison of the combination algorithms

| FSM | Classifier | MAPE / % |
|---|---|---|
| lm<br>PCA | Linear regression | 6.14<br>5.95 |
| lm<br>PCA | Elasticnet | 6.14<br>5.81 |
| lm<br>PCA | KNN | 6.34<br>4.38 |
| lm<br>PCA | Random forest | 4.88<br>4.72 |
| lm<br>PCA | XGBM | 5.63<br>7.35 |
| lm<br>PCA | Linear SVM | 5.86<br>5.46 |
| lm<br>PCA | Radial SVM | 5.06<br>4.81 |

From Tab. 2 XGBM is the only model that shows smaller MAPE in combination with lm feature selection. All other models achieve better accuracy according to MAPE in the combination with PCA feature selection. The smallest MAPE of 4.38% is achieved with PCA+KNN, while the largest MAPE of 7.35% is obtained with PCA+XGBM.

## 5.3 Forecasts from Neural network

The forecasting results from the neural network model also show quite a good accuracy which can be seen in Fig. 7. The obtained MAPE of model NNAR(3, 1, 6) is 2.67%.

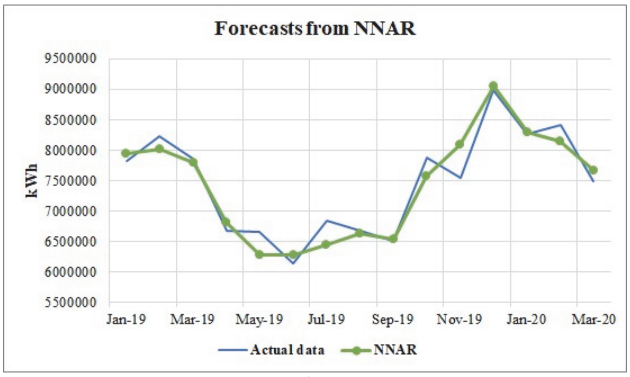


**Figure 7** Forecast from Neural networks

# 6 DISCUSION

A graphical comparison of MAPEs for different methods is shown in Fig.8. For all models, error ranges between 3% and 8% which is acceptable for application in practice for MTLF. The NNAR model performed the best accuracy, with MAPE of 2.67 %. The classical time series methods showed better accuracy than other machine learning methods. Having in mind works [36, 37], this was expected as the sample size is quite small.

To compare absolute relative errors per month, three methods are chosen (one from each category): ETS($A$, $N$, $A$) which showed better accuracy than other classical time series methods, PCA+KNN which performed better than other machine learning methods used with PCA or lm, and the neural network model NNAR(3, 1, 6). The comparison of absolute relative errors per month for these methods is given in Tab. 3. When considering absolute relative errors per months then the models' results can be very different from each other. Some models give a better prediction for some months than others. That depends on the models and on the behavior of the data for the observed month in the given years.

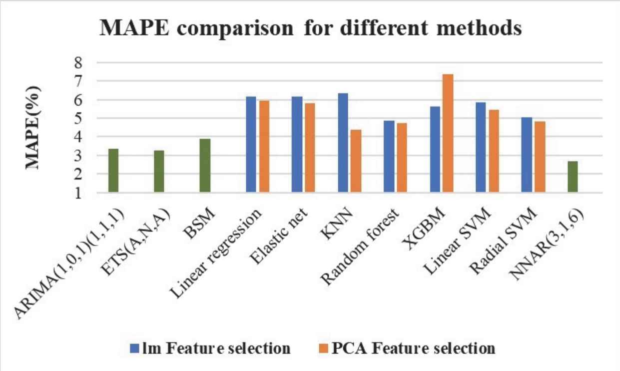


**Figure 8** MAPE comparison for different methods

**Table 3** Absolute relative error per month for ETS(*A*, *N*, *A*), PCA+KNN, and NNAR(3, 1, 6) model

| Month | Absolute relative error / % | | |
|---|---|---|---|
| | ETS(*A*, *N*, *A*) | PCA+KNN | NNAR(3, 1, 6) |
| January 2019 | 2.20 | 6.71 | 1.55 |
| February 2019 | 4.24 | 2.67 | 2.74 |
| March 2019 | 1.63 | 1.61 | 0.86 |
| April 2019 | 0.07 | 0.19 | 2.02 |
| May 2019 | 6.15 | 4.72 | 5.47 |
| June 2019 | 4.91 | 3.76 | 2.27 |
| July 2019 | 3.46 | 2.98 | 6.01 |
| August 2019 | 0.15 | 3.44 | 0.83 |
| September 2019 | 1.42 | 12.57 | 0.57 |
| October 2019 | 0.79 | 2.44 | 3.83 |
| November 2019 | 7.40 | 7.32 | 7.19 |
| December 2019 | 0.91 | 3.11 | 0.76 |
| January2020 | 2.55 | 1.46 | 0.28 |
| February2020 | 5.59 | 4.82 | 3.14 |
| March2020 | 7.68 | 7.95 | 2.48 |

Tab. 3 shows that for the months April 2019, August 2019, and October 2019 ETS(*A*, *N*, *A*) model performs more accurately in forecasting electricity consumption than the other models. For February 2019, May 2019, and July 2019 the smallest error is achieved by forecasting with the PCA+KNN, while for January 2019, March 2019, June 2019, September 2019, November 2019, December 2019, January 2020, February 2020, and March 2020 the best results are obtained by implementing NNAR(3, 1, 6) model. For July 2019 the forecasting from PCA+KNN gives a much smaller error than the other two models, while for September it gives a much larger error than ETS and NNAR(3, 1, 6). Forecasting from ETS(*A*, *N*, *A*) gives a very small absolute relative error, less than 1%, for April 2019, August 2019, October 2019, and December 2019. The smallest absolute relative error with PCA+KNN is achieved for April 2019, while with NNAR(3, 1, 6) the error of less than 1% is achieved for March 2019, August 2019, September 2019, December 2019, and January 2020. The largest deviations of the NNAR(3, 1, 6) model, above 5%, are obtained for the months of May, July, and November 2019. PCA+KNN is at least accurate for January, September, November 2019, and March 2020. For May 2019, November 2019, February 2020 and March 2020 ETS(*A*, *N* ,*A*) showed the smallest accuracy. For all three models, the obtained MAPE for the month of November is larger than 7%, and the MAPEs of these methods are the most comparable for this month.

Although, the MAPE is the main chosen indicator for selecting the best forecasting model in this paper, absolute relative errors per month can provide an additional insight that can be useful for DSO special applications, such as comparison of errors per month or seasons.

## 7 CONCLUSION

In this paper, twelve different methods for MTLF are compared to see which method gives the best results and whether the errors obtained by using these methods are comparable. An exhaustive comparative analysis of such a large number of methods has not been done so far for MTLF. The analysis given in this paper could be very useful for further analysis and improvement of methods for MTLF and gives significant support to the electric utilities that plan to operate in deregulated electricity markets.

The efficiency of provided models is competitive with the current publicly available models in the terms of accuracy and their application in the real market, thus it represents contribution to the academia and energy community.

It can be concluded that for this type of data, time series models perform better than classical machine learning methods. Even though the best results were expected from neural networks, it can be concluded that MTLF does not have enough data to create the model with significant accuracy.

Comparing to overviewed references, the future progress could be made with combining several methods into one model, since newest publications show biggest improvements in performance and accuracy in that way. Comparing to above mentioned references, MAPE of 3.28% with ETS model is competitive, but authors cannot make final conclusions and comparisons with other publications, since the dataset needs to be the same in order to do so.

Future work should focus on creating a general algorithm for choosing the right model for MTLF on any kind of consumption data and improving data quality with pre-processing and analysis.

## 8 REFERENCES


[1] Kyriakides, E. & Polycarpou, M. (2007). Short Term Electric Load Forecasting: A Tutorial. In: Chen, K., Wang, L. (eds) Trends in Neural Computation. *Studies in Computational Intelligence*, *35*. https://doi.org/10.1007/978-3-540-36122-0_16

[2] Berk, K. (2015). *Modeling and forecasting electricity demand: A risk management perspective*. Springer Spektrum.

[3] Singh, A., Ibraheem, K., & Muazzam M. (2013). An overview of electricity demand forecasting techniques. *Proceedings of- National Conference on Emerging Trends in Electrical, Instrumentation & Communication Engineering*, *3*, 38-48.

[4] Fallah, S., Ganjkhani, M., Shamshirband, S., & Chau, K.-W. (2019). Computational intelligence on short-term load

forecasting: A methodological overview. *Energies*, *12*(3), 1-21. https://doi.org/10.3390/en12030393

[5] Cerjan, M., Petričić A., & Delimar, M. (2019). Hira model for short-term electricity price forecasting. *Energies*, 12, 1-32. https://doi.org/10.3390/en12030568

[6] Hippert, H. S., Pedreira C. E., & Souza, R. C. (2001). Neural networks for short-term load forecasting: a review and evaluation. *IEEE Transactions on Power Systems*, *16*(1), 44-55. https://doi.org/10.1109/59.910780

[7] Kaytez, F., Taplamaciog˘lu, M. C., Çam, E., & Hardalaç, F. (2015). Forecasting electricity consumption: A comparison of regression analysis, neural networks and least squares support vector machines. *International Journal of Electrical Power & Energy Systems*, *67*, 431-438. https://doi.org/10.1016/j.ijepes.2014.12.036

[8] Weron, R. (2007). *Modeling and Forecasting Electricity Loads and Prices: A Statistical Approach*. The Wiley Finance Series. Wiley.

[9] Abdoos, A., Hemmati, M., & Abdoos A. A. (2015). Short term load forecasting using a hybrid intelligent method. *Knowledge-Based Systems*, *76*, 139-147. https://doi.org/10.1016/j.knosys.2014.12.008

[10] Song, K.-B., Ha, S.-K., Park, J.-W., Kweon, D.-J., & Kim, K.-H. (2006). Hybrid load forecasting method with analysis of temperature sensitivities. *IEEE Transactions on Power Systems*, *21*, 869-876. https://doi.org/10.1109/TPWRS.2006.873099

[11] Khwaja, A., Naeem, M., Anpalagan, A., Venetsanopoulos, A. & Venkatesh, B. (2015). Improved short-term load forecasting using bagged neural networks. *Electric Power Systems Research*, *125*, 109-115. https://doi.org/10.1016/j.epsr.2015.03.027

[12] Bhattacharyya S. C. & Thanh, L. T., (2004). Short-term electric load forecasting using an artificial neural network: Case of Northern Vietnam. *International Journal of Energy Research*, *28*(59, 463-472. https://doi.org/10.1002/er.980

[13] Kandil, N., Wamkeue, R., Saad, M., & Georges, S. (2006). An efficient approach for short term load forecasting using artificial neural networks. *International Journal of Electrical Power & Energy Systems*, *28*(8), 525- 530. https://doi.org/10.1016/j.ijepes.2006.02.014

[14] Mamlook, R., Badran, O., & Abdulhadi, E. (2009). A fuzzy inference model for short-term load forecasting. *Energy Policy*, 37(4), 1239-1248. https://doi.org/10.1016/j.enpol.2008.10.051

[15] Aggarwal, S., Kumar, M., Saini, L., & Kumar, A. (2011). Short-term load forecasting in deregulated electricity markets using fuzzy approach. *Journal of Engineering & Technology*, *1*, 24.

[16] Ling, S. H., Leung, F., Lam, H.-K., & Tam, P. (2004). Short- term electric load forecasting based on a neural fuzzy network. *Industrial Electronics, IEEE Transactions*, *50*, 1305-1316. https://doi.org/10.1109/TIE.2003.819572

[17] Papaioannou, G. P., Dikaiakos, C., Dramountanis, A., & Papaioannou, P. G., (2016). Analysis and modeling for short- to medium-term load forecasting using a hybrid manifold learning principal component model and comparison with classical statistical models (SARIMAX, Exponential Smoothing) and Artificial Intelligence Models (ANN, SVM): The Case of Greek Electricity Market. *Energies*, *9*, 1-40. https://doi.org/10.3390/en9080635

[18] Shankar, R., Chatterjee, K., & Chatterjee, T. K., (2012). A very short-term load forecasting using Kalman filter for load frequency control with economic load dispatch. *Journal of Engineering Science and Technology Review*, *5*, 97-103. https://doi.org/10.25103/jestr.051.17

[19] Guan, C., Luh, P. B., Michel, L. D., & Chi, Z. (2013). Hybrid Kalman Filters for very short-term load forecasting and prediction interval estimation. *IEEE Transactions on Power Systems*, *28*, 3806-3817. https://doi.org/10.1109/TPWRS.2013.2264488

[20] Khuntia, S. R., Rueda, J. L., & van der Meijden, M. A. (2016). Forecasting the load of electrical power systems in mid- and long-term horizons: a review. *IET Generation, Transmission Distribution*, 10(16), 3971-3977. https://doi.org/10.1049/iet-gtd.2016.0340

[21] Xia, C., Wang, J., & McMenemy, K. (2010). Short,medium and long term load forecasting model and virtual load fore- caster based on radial basis function neural networks. *International Journal of Electrical Power & Energy Systems*, *32*(7), 743-750. https://doi.org/10.1016/j.ijepes.2010.01.009

[22] Abdel-Aal, R. & Al-Garni, A. (1997). Forecasting monthly electric energy consumption in eastern Saudi Arabia using univariate time-series analysis. *Energy*, *22*(11), 1059-1069. https://doi.org/10.1016/S0360-5442(97)00032-7

[23] Al-Hamadi, H. & Soliman, S. (2005). Long-term/mid-term electric load forecasting based on short-term correlation and annual growth, *Electric Power Systems Research*, *74*, 353-361. https://doi.org/10.1016/j.epsr.2004.10.015

[24] Mirasgedis, S., Sarafidis, Y., Georgopoulou, E., Lalas, D., Moschovits, M., Karagiannis, F., & Papakonstantinou, D. (2006). Models for mid-term electricity demand forecasting incorporating weather influences. *Energy*, *31*, 208-227. https://doi.org/10.1016/j.energy.2005.02.016

[25] Goude, Y., Nedellec, R., & Kong, N. (2013). Local short and middle term electricity load forecasting with semi-parametric additive models. *IEEE transactions on smart grid*, *5*, 440-446. https://doi.org/10.1109/TSG.2013.2278425

[26] de Oliveira E. M. & Oliveira, F. L. C. (2018). Forecasting mid-long term electric energy consumption through bagging arima and exponential smoothing methods. *Energy*, *144*, 776-788. https://doi.org/10.1016/j.energy.2017.12.049

[27] Doveh, E., Feigin, P., Greig, D., & Hyams, L. (1999). Experience with FNN models for medium term power demand predictions. *IEEE Transactions on Power Systems*, *14*(2), 538-546. https://doi.org/10.1109/59.761878

[28] Elkateb, M., Solaiman, K., & Al-Turki, Y. (1998). A comparative study of medium-weather-dependent load forecasting using enhanced artificial/fuzzy neural network and statistical techniques. *Neurocomputing*, *23*(1), 3-13. https://doi.org/10.1016/S0925-2312(98)00076-9

[29] Ghiassi, M., Zimbra, D. K., & Saidane, H. (2004). Medium term system load forecasting with a dynamic artificial neural network model. *Electric Power Systems Re- search*, *76*, 302-316. https://doi.org/10.1016/j.epsr.2005.06.010

[30] Smola A. & Schölkopf, B. (2004). A tutorial on support vector regression. *Statistics and Computing*, *14*, 199-222. https://doi.org/10.1023/B:STCO.0000035301.49549.88

[31] Hong, W.-C., Dong, Y., Lai. C.-Y., Chen, L.-Y., & Wei, S.-Y. (2011). SVR with hybrid chaotic immune algorithm for seasonal load demand forecasting. *Energies*, *4*, 960-973. https://doi.org/10.3390/en4060960

[32] Ye, S., Zhu, G., & Xiao, Z. (2012). Long term load forecasting and recommendations for China based on support vector regression. *Energy and Power Engineering*, *04*, 380-385. https://doi.org/10.1109/ICIII.2011.418

[33] Wang, J., Niu, L. Li., D. & Tan, Z. (2012). An annual load forecasting model based on support vector regression with differential evolution algorithm. *Applied Energy*, *94*, 65-70. https://doi.org/10.1016/j.apenergy.2012.01.010

[34] Hong, W.-C., Dong, Y., Zhang, W. Y., Chen, L.-Y., & Panigrahi, B. K. (2013). Cyclic electric load forecasting by seasonal SVR with chaotic genetic algorithm. *International Journal of Electrical Power & Energy Systems*, *44*(1), 604-614. https://doi.org/10.1016/j.ijepes.2012.08.010

[35] Hu, Z., Bao, Y., Chiong, R., & Xiong, T. (2015). Mid-term interval load forecasting using multi-output support vector regression with a memetic algorithm for feature selection. *Energy*, *84*, 419-431. https://doi.org/10.1016/j.energy.2015.03.054

[36] Cerqueira, V., Torgo, L., & Soares, C. (2019). Machine learning vs statistical methods for time series forecasting: Size matters. https://doi.org/10.48550/arXiv.1909.13316

[37] Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2018). Statistical and machine learning forecasting methods: Concerns and ways forward. *PLoS ONE*, *13*. https://doi.org/10.1371/journal.pone.0194889
[38] Tang, L., Wang, Z., Wang, X., Shao, C., Liu, S., &Tian, S. (2019). Long-term electricity consumption forecasting based on expert prediction and fuzzy Bayesian theory. *Energy, 167*, 1144-1154. https://doi.org/10.1016/j.energy.2018.10.073
[39] Jiang, W., Wu, X., Gong, Y., Yu, W., & Zhong, X. (2020). Holt-Winters smoothing enhanced by fruit fly optimization algorithm to forecast monthly electricity consumption. *Energy*, *193*. https://doi.org/10.1016/j.energy.2019.116779
[40] Liu, C., Sun, B., Zhang, C., & Li, F. (2020). A hybrid prediction model for residential electricity consumption using holt-winters and extreme learning machine. *Applied Energy*, *275*. https://doi.org/10.1016/j.apenergy.2020.115383
[41] Gao, F., Chi, H., & Shao, X. (2021). Forecasting residential electricity consumption using a hybrid machine learning model with online search data. *Energy Procedia*, *158*, 2713-2718. https://doi.org/10.1016/j.apenergy.2021.117393
[42] Pinto, T., Praça, I., Vale, Z., & Silva, J. (2021). Ensemble learning for electricity consumption forecasting in office buildings. *Neurocomputing*, *423*, 747-755. https://doi.org/10.1016/j.neucom.2020.02.124
[43] Kaytez, F. (2020). A hybrid approach based on autoregressive integrated moving average and least-square support vector machine for long-term forecasting of net electricity consumption. *Energy*, *19*. https://doi.org/10.1016/j.energy.2020.117200
[44] Zheng, Z., Chen., H., & Luo, X., (2019) Spatial granularity analysis on electricity consumption prediction using LSTM recurrent neural network. *Energy Procedia*, *158*, 2713-2718. https://doi.org/10.1016/j.egypro.2019.02.027
[45] Shao, X., Kim, C. S., & Sontakke, P. (2020). Accurate Deep Model for Electricity Consumption Forecasting Using Multi-Channel and Multi-Scale Feature Fusion CNN-LSTM. *Energies*, *13*(8), 1881. https://doi.org/10.3390/en13081881
[46] X. Wang, F. Fang, X. Zhang, Y. Liu, L. Wei, Y. Shi (2019). LSTM-based Short-term Load Forecasting for Building Electricity Consumption. *IEEE 28th International Symposium on Industrial Electronics (ISIE)*, 1418-1423. https://doi.org/10.1109/ISIE.2019.8781349
[47] Lin, Z., Cheng, L., & Huang, G. (2020). Electricity consumption prediction based on LSTM with attention mechanism. *IEEJ Transactions on Electrical and Electronic Engineering*, *15*(4). https://doi.org/10.1002/tee.23088
[48] Almalaq, A. & Zhang J. J. (2019). Evolutionary Deep Learning-Based Energy Consumption Prediction for Buildings. *IEEE Access*, *7*, 1520-1531. https://doi.org/10.1109/ACCESS.2018.2887023
[49] Hyndman R. & Athanasopoulos, G. (2014). Forecasting: principles and practice. *OTexts*.
[50] Hyndman R. J. & Khandakar, Y. (2008). Automatic time series forecasting: the forecast package for R. *Journal of Statistical Software*, *27*, 1-22. https://doi.org/10.18637/jss.v027.i03
[51] Hyndman, R., Koehler, A., Ord, J., & Snyder, R. (2008). *Forecasting with Exponential Smoothing: The State Space Approach*. Springer.
[52] Harvey, A. C. (1990). *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge University Press. https://doi.org/10.1017/CBO9781107049994
[53] Brownlee, J. (2017). Introduction to Time Series Forecasting with Python: How to Prepare Data and Develop Models to Predict the Future. *Machine Learning Mastery*.
[54] Brownlee, J. (2016). Master Machine Learning Algorithms: discover how they work and implement them from scratch. *Machine Learning Mastery*.
[55] Kuhn, M. (2008). Building Predictive Models in R Using the caret Package. *Journal of Statistical Software*, *28(*5), 1-26. https://doi.org/10.18637/jss.v028.i05
[56] Tashman, L. J. (2000). Out-of-sample tests of forecasting accuracy: an analysis and review, *International Journal of Forecasting*, *16*(4), 437-450. https://doi.org/10.1016/S0169-2070(00)00065-0

**Contact information:**

**Sofija KRSTEV**, MSc
Dwelt Ltd.,
Bulevar srpske vojske 17, 78000 Banja Luka
E-mail: sofija@dwelt.net

**Jovana FORCAN**, PhD
Dwelt Ltd.,
Bulevar srpske vojske 17, 78000 Banja Luka
E-mail: jovana@dwelt.net

**Dragoljub KRNETA**, PhD
(Corresponding author)
Dwelt Ltd.,
Bulevar srpske vojske 17, 78000 Banja Luka
E-mail: dragoljub@dwelt.net