ENHANCING ENERGY EFFICIENCY IN A SMART HOME THROUGH WINDOW-BASED SUPPORT VECTOR REGRESSION FOR ENERGY CONSUMPTION PREDICTION

*B. S. E. Zoraida (1), J. Jasmine Christina Magdalene (2)**

(1) Computer Science and Engineering, Bharathidasan University, Tiruchirappalli;
(2) Computer Applications, Bishop Heber College, (Affiliated to Bharathidasan University), Tiruchirappalli
India

* Corresponding Author, e-mail: jjasminebhc@gmail.com

Abstract: Efficient energy management is greatly facilitated by accurately predicting energy consumption in a smart home, benefiting both consumers and utilities alike. The conventional forecasting techniques rely on pre-trained statistical models built upon extensive historical data, which may experience performance degradation due to the dynamic nature of power load demands. To address this limitation, this study proposes a novel approach employing Window-based Support Vector Regression (WSVR) to accurately estimate energy requirements from a smart grid within a smart home. The dataset utilized for this research is sourced from Pecan Street in Texas, USA. To assess the efficacy of the proposed model, it is compared to several other time series data prediction models, including ARIMA, Holt Winter's, Linear Regression, Support Vector Machine, and Support Vector Regression. The performance of each model is evaluated, and the results are thoroughly examined and discussed.

Key words: smart grid, energy management, arima, sliding window.

1. INTRODUCTION

The efficient and dependable performance of electricity systems now heavily relies on energy forecasting. The operation and production costs of the electrical utility depend heavily on the accuracy of prediction. Enhancement in accuracy will not only help in revenue management of utilities but it will also help in meeting the energy needs of the consumers. The efficiency in energy management will also enhance the security of the power system [1]. Maintenance is planned by scheduling the smart appliances in a smart home to manage energy during the most needed time thereby reducing the cost for energy consumption [2].

Three categories of load forecasting are generally used: short-term load forecasting, mid-term load forecasting and long-term load forecasting. The number of units consumed, and the economical dispatch are two processes that depend on short-term load forecasting (STLF), which can last from an hour to a week [3], planning of energy transfer as well as immediate operation and management using the past data [4, 5]. This has led to the development of several forecasting models, the bulk of which may be roughly brought under the categories of machine learning and statistical methods. Most of the statistical techniques are based on analysis of linear series of data have trouble tackling the prediction of load problem since load series are often nonlinear. Neural networks and other AI-based methods have become highly popular in recent years for producing promising outcomes. In this work, an adaptive model for time series forecasting is proposed wherein the concept of sliding window is modified and combined with the Support Vector Regression technique. Comprehensive dataset from a real-world smart home in Texas is used for this work. New insights and findings in energy consumption forecasting are discussed in the result section. The contribution of this work includes a novel methodology conducted through data analysis and pre-processing and providing evidence of the effectiveness of the model through comprehensive evaluation and validation.

2. SURVEY OF RESEARCH IN THE FIELD

Different models are commonly used to model a regression problem. The traditional approach for addressing regression issues uses statistical techniques like the ARIMA model, in which the time series dataset is broken into numerous elements, such as trend and seasonality, and develops models for each element. To ensure that the auto correlation information is preserved, creating a list of temporal attributes is another method for modelling the historical data. The duration of an event, the interval between two occurrences, entropy measurements and other temporal aspects are some of the most often utilised temporal features. To calibrate its model parameters, a person must be an expert in statistics [6]. Optimization of the devices that use electricity is presented in [7] to enhance the efficiency of the electrical energy. Article [8] have done research in identifying the problems in energy management when a home undergoes automation. Support Vector Machine algorithm is not efficient, and the output is affected by noisy data [9]. Prediction of electrical load using the classification techniques and feature extraction techniques along with SVM is done in [10]. Support Vector Regression is proved to perform well for the dataset which shows trend and seasonality.

Regression analysis is carried out using SVR, an extension of SVM. Typically, the demand of load data in the past, changes in the climatic conditions etc. are used to create models for energy requirement forecasting. Sliding Windows Regression, an intriguing forecasting method that is capable of producing excellent results with little to no prior knowledge, is also used as a strategy for prediction. Support Vector Regression along with Fog and Cloud Computing for the forecasting of energy consumption is used in [11]. One model for predicting loads across the entire power plant's area system serving a huge geographic area may not always ensure sufficient forecasting accuracy. One of the main causes is the area's widespread load diversity, which is typically brought on by localised weather variations. Forecasting the load for multiple regions will be a useful way to give the utilities with regional predictions as well as more precise forecasting findings. A discussion for evaluation of power dissipation and energy in a robotic system is made in [12]. Article [13] proposes a model for the prediction of load by using SVR and LSTM algorithm for a micro-grid in a rural household of Africa. Nevertheless, most of these techniques rely on every application and don't have a universal fix that can be used across other domains. Forecasting in different domains have gained lots of attention in research where energy management is also an important area. For energy management different techniques and algorithms are used by different researchers. Support Vector Regression when combined with sliding window concept can be considered for forecasting by taking the seasonality component of time series dataset. The Pecan Street, Texas dataset is taken for investigation. Forecasting of energy consumption is needed as it helps the utilities as well as the consumers in managing the energy efficiently and to plan for continuous energy supply even at the time of natural calamities. The average energy consumption from the dataset and the decomposition of data for this work is the same as in [14].

3. DESCRIPTION OF THE PROBLEM AND PROPOSED SOLUTION:

3.1. Preliminaries

3. 1. 1. Linear Regression

Linear regression is widely recognized as the most straightforward regression technique rooted in statistical analysis for predictive analysis in machine learning. The concept of linear regression revolves around establishing a relationship that follows a straight line between the independent variable (represented by the X-axis) and the dependent variable (represented by the Y-axis). This technique aims to capture and quantify the linear association between the variables, enabling predictions to be made based on this linear relationship. If the only input variable, G (the independent variable), is present, this sort of linear regression is referred to as simple linear regression. When using the linear regression approach to find the line that best fits the data, B0 and B1 must be set to their ideal values. With the following equation, one can calculate linear regression:

$$G_i = f(D_i + \beta_0) + E_i$$

where G_i is the independent variable, f is the function, D_i is the dependent variable, β_0 is the unknown parameter and E_i is the error factor.

3.1.2. ARIMA

ARIMA is a forecasting model which is one of the types of regression technique that assesses the supremacy of a single dependent variable in respect to a number of variables which keeps changing. The ARIMA model forecasts future movements by examining the differences between the values in the series rather than actual values.

Auto Regression (AR): Depicts the variable that changes and regresses by itself based on previous values.

Integrated (I): To enable the stationary component that is present in the time series dataset, (I) indicates the differencing of the unprocessed information.

Moving Average (MA): This tells the relationship between the value that is observed and the residual error. When this model is applied to lagged values that are observed, the moving average (MA) integrates the relationship between an observation and a residual error. The representation of the ARIMA model is as follows:

$$b'_t = c + \phi_1 b'_{t-1} + \cdots + \phi_p b'_{t-p} + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q} + \varepsilon_t$$

where b't indicates the data which is differenced, and the other predictors indicate the lagged values and errors.

3.1.3. Support Vector Regression

To predict discrete values, support vector regression, a method for supervised learning, is utilised. The same theory underlies the operation of both SVMs and Support Vector Regression. Cornerstone of SVR is locating the best fit line. In SVR, the line that fits the hyperplane the best is the one with the greatest points. In contrast to other Regression models, which seek to reduce the difference that is present between the original value and the projected value, the SVR aims to identify the best line which lies within the appropriate value. The threshold value is the area that is within the boundary line and the hyper plane. It is challenging to scale SVR to datasets with more than a few ten thousand samples. The Support Vector Regression in the feature space can be calculated using the formula given below:

$$y(x,w)=(w.f(x) + e)$$

where "w" is the factor related to weight, f(x) is the function of the feature and e is a constant.

The preliminaries that are given in this section are the existing mathematical equations related to the various algorithms that are used for forecasting. These equations are used to compare the efficiency of this regression model which is given by modifying the concept of sliding window and combining it with the Support Vector Regression model to give better accuracy in forecasting.

3.2. Proposed Solution

For the prediction of dynamic load data, the Window-based Support Vector Regression (WSVR), model is proposed which is adaptive in nature. Large volumes of historical data are generally used to construct prediction models, and once the dataset is trained; it will not get upgraded because of the limitations imposed by a protracted period of training. Performance may suffer as a result of a context change for the application or the data that is real. The way the data is applied in a particular context, or the actual data may undergo changes, which result in the prediction model's performance to get degraded. A prediction model is developed that trains using the data within the sliding window for scenarios like the forecasting of real-time load data, once the model is complete. As data is received, an error is calculated and appropriately incorporated into the model. This work is implemented by using R tool.

The energy consumption in a smart home from the smart grid is predicted using SVR and the model is trained using sliding windows as the data on smart grid is dynamic. The size of the window is optimized by taking the seasonality based on the Pecan Street dataset from Texas which shows four different seasons consisting of three months each. The dataset is taken for a single smart home in Texas and the average amount of energy is calculated by summing up the energy that is consumed from the grid by various appliances. The suggested model is adaptive since it lowers forecasting errors by retraining the model on an on-going basis based on seasons. The efficiency of the prediction models is measured calculating the various errors like Mean Absolute Error (MAE) Root Mean Square Error (RMSE), the Mean Absolute Percentage Error (MAPE), and other metrics. The forecasting of energy consumption can be of great help when it comes to energy management. The consumers as well as the utilities can know the average amount of energy that is required for a smart home during various seasons and based on this other energy resources can be installed and utilized so as to avoid energy shortage. If energy consumption is less, then it can be given back to the grid. This forecasting helps the consumers and all the needed utilities to manage the distribution of energy. The utilities can also ensure uninterrupted power supply during the time of outages.

3.3. Methodology of the Proposed Work

The dataset taken for this work is from Pecan Street Texas, USA. The average amount of energy consumption in a smart home from the smart grid is taken for developing this forecasting model. The energy consumed by various smart appliances in a smart home is taken for every 15mins. Then the average amount of energy consumption in KW is calculated for every month. Texas has four seasons namely summer, winter, autumn, and spring. Based on the months in the respective seasons the average amount of energy consumption is taken, and the window size is kept as 3. This is because three months constitute a season. The average amount of energy consumption from the smart grid in the smart home from the Pecan Street dataset for the years 2013 to 2019 is given in Table 1. This dataset when decomposed shows that it is a time series data as it has randomness, seasonality, cyclicality, and trend. This is depicted in Fig1. As the dataset is a time series dataset the suggested prediction model is developed and contrasted with alternative time series forecasting methods. Various techniques that can be used for prediction of time series data ranging from statistical to Machine Learning algorithms are discussed in the above section. In this work Support Vector Regression method is used as the dataset shows non-linearity and can be modelled using the kernel functions.

The following are the steps that are taken for the proposed model. Initially the dataset is split into test data and training data. The average amount of energy consumption in a smart home for five years i.e., 2013 to 2017 is taken as the training data. The other two years data is considered as the test data. After the dataset is split into training and test data, the window size is fixed based on the seasons. As Texas has four different seasons with three months each, the value of the window is taken to be as 3 which indicates that g(t), g(t-1) and g(t-2) is considered for forecasting. Then the appropriate regression technique is chosen. SVR algorithms are more sophisticated than their competitors which produce models that are more accurate. Hence SVR forecasting model is taken for this proposed work. However, as a narrow training window is taken for this work, the increased complexity becomes insignificant for this dataset. After fixing the window, SVR is used to forecast the average amount of energy consumption in the smart home for the same season in the coming year. Fig 2 shows the sequence of steps involved in this work.

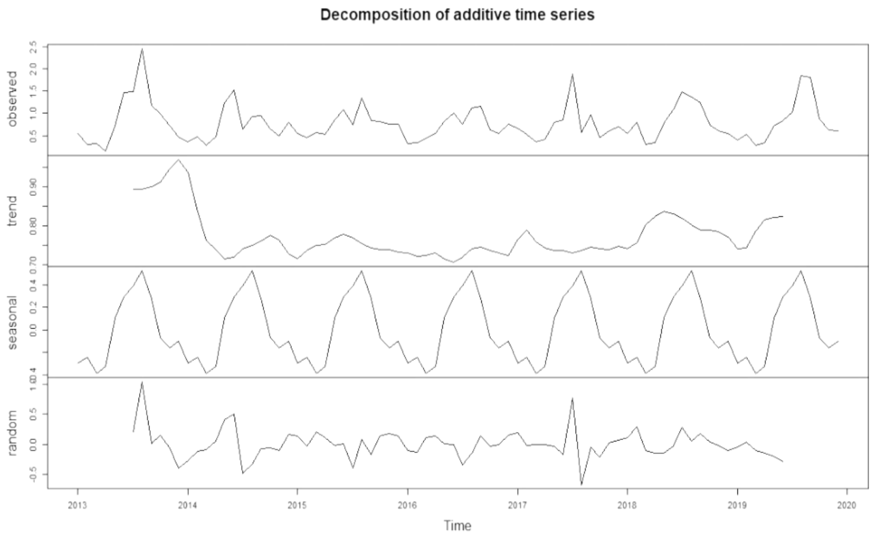
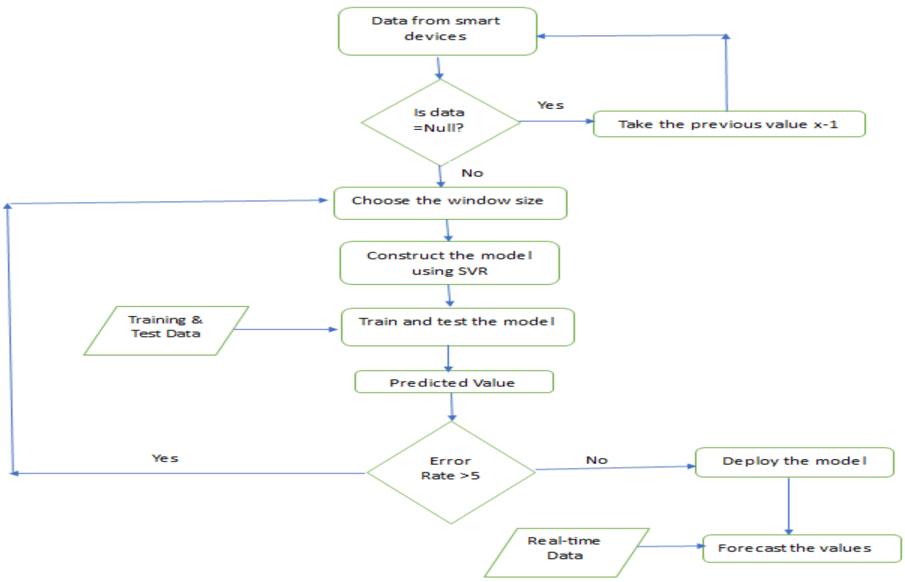


Table 1. Average Amount of Energy consumption from the smart grid in KW

Year	***Jan***	***Feb***	***Mar***	***Apr***	***May***	***Jun***	***Jul***	***Aug***	***Sep***	***Oct***	***Nov***	***Dec***
2013	*0.55*	*0.3*	*0.32*	*0.15*	*0.72*	*1.47*	*1.49*	*2.46*	*1.81*	*0.99*	*0.72*	*0.47*
2014	*0.36*	*0.47*	*0.29*	*0.47*	*1.23*	*1.52*	*0.64*	*0.94*	*0.96*	*0.64*	*0.5*	*0.8*
2015	*0.55*	*0.46*	*0.57*	*0.54*	*0.85*	*1.08*	*1.38*	*1.36*	*0.84*	*0.81*	*0.76*	*0.77*
2016	*0.33*	*0.34*	*0.44*	*0.55*	*0.82*	*1*	*0.76*	*1.12*	*1.16*	*0.63*	*0.56*	*0.77*
2017	*0.66*	*0.53*	*0.36*	*0.41*	*0.8*	*0.85*	*1.89*	*0.58*	*0.97*	*0.46*	*0.6*	*0.71*
2018	*0.55*	*0.8*	*0.31*	*0.35*	*0.79*	*1.09*	*1.48*	*1.37*	*1.24*	*0.75*	*0.61*	*0.56*
2019	*0.4*	*0.53*	*0.29*	*0.35*	*0.72*	*0.83*	*1.03*	*1.85*	*1.8*	*0.88*	*0.63*	*0.61*

For machine learning models, choosing the best window size for training the data is still an optional area for research. It is recommended to use past data with huge velocity is needed for prediction model to be trained which will help in uncovering all the possible patterns within the given time-period. Typically, the accuracy of the prediction model increases if the size of the training window is increased. There is also a drawback in training a huge dataset as the prediction model may produce false readings if the model shows some behavioural changes. In contrast, in this proposed work, a window is used for training the model which takes the amount of energy consumption of a month which falls under a particular season and predicts the energy consumption for the same season for the forthcoming year.



3.4. Algorithm of the proposed model

Step 1: Get the data from smart appliances in a smart home
Step 2: Check the data. If it is null, goto step 1 and take the previous value
Step 3: Choose the size of the window
Step 4: Build the model using SVR
Step 4: Train and Test the model
Step 5: If error rate is greater than 5 goto step 3 else
Step 6: Deploy the model
Step 7: Read the real-time data and forecast
Step 8: Stop

4. RESULTS

In this work the sliding window technique is incorporated with the Support Vector Regression method to predict the amount of energy consumption on an average during a particular from the smart grid by a smart home. The energy consumption for a smart home from the Pecan Street dataset is taken for forecasting. The data that is collected and divided into training and test data and the methodology given in the section given above is followed. The accuracy of this model is assessed by calculating the RMSE, MAE, MSE.

RMSE is calculated using the formula, $\sqrt{[\ \Sigma(F_k - O_k)^2 / \ n\ ]}$ where F_k is the forecasted value of the k^{th} term in the dataset and O_k is the observed value of the k^{th} term and n denotes the sample size that is taken for forecasting.

MAE is calculated using $(1/c)*\sum|x_k-y_k|$ where c is the number of observations, x_k is the value that is obtained for k^{th} data value and y_k is the forecasted value for k^{th} data value.

MSE is calculated by using the formula $(1/n)* \sum(\text{actual value} - \text{predicted value})^2$

The performance is compared with the performance of the models like ARIMA, Holt Winter's, linear regression, Support Vector Machine and Support Vector Regression. The results that are obtained is shown in Table 2. The result that is obtained is shows that the RMSE is high fir ARIMA when compared to the other models and the proposed technique shows that it is better when compared to other time series forecasting techniques. The graphical representation of the obtained results is shown in Fig 3.

Table 2. Comparison of the performance of time series forecasting techniques

	ARIMA	***Linear Regression***	***Holt-Winter's***	***SVM***	***SVR***	***WSVR***
RMSE	*0.5575*	*0.228*	*0.75*	*0.29*	*0.16*	*0.09*
MSE	*0.31*	*0.1672*	*0.531*	*0.25*	*0.235*	*0.13*
MAE	*0.37*	*0.33*	*0.65*	*0.35*	*0.31*	*0.29*

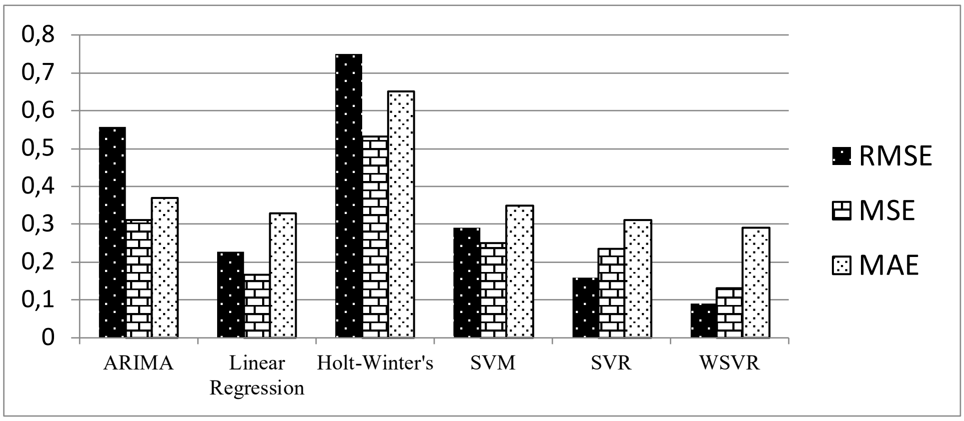


Figure 3: Graphical Representation - Comparison of various forecasting models

5. DISCUSSION

According to the results presented in the previous section, the Window-based Support Vector Regression model, proposed in this study, demonstrates lower Root Mean Square Error (RMSE) values compared to the traditional ARIMA model commonly employed for time-series forecasting. This indicates that the proposed model outperforms ARIMA in terms of accuracy and predictive capability for the given dataset. Holt Winter's forecasting shows the highest error rate and hence less efficient for time series forecasting. Linear Regression technique is better than ARIMA as only one dependent variable is taken. Support Vector Machine and Support Vector Regression also proves to be better than ARIMA which indicates that regression techniques are better when only one dependent variable is taken for consideration. When the sliding window is taken, the seasonality factor will not be considered as the window size keeps moving. So, this proposed work WSVR proves much better with 0.9, 0.13, 0.29 errors for RMSE, MSE and MAE respectively as the window size is kept as a constant.

6. CONCLUSION

This proposed work uses support vector regression with constant window size based on seasonality for the forecast of energy requirement in a smart home. Based on the output that is obtained which is deliberated in the previous sections, it is concluded that the suggested model performs better when compared with the ARIMA, Linear Regression and Support Vector Regression based on seasonality. This model forecasts the amount of energy that will be consumed in the forthcoming year for the same season. One limitation of the proposed model is that it relies solely on the average energy consumption from a single energy source, namely the smart grid, without considering other energy resources present in a smart home. This oversight hampers effective energy management, as it fails to account for the contribution of other energy sources within the system. Additionally, the model does not incorporate weather data as an independent variable, which is a drawback considering the significant impact of weather on household energy consumption. Integrating weather data into the model could provide more accurate and comprehensive energy consumption forecasts. In the future, prediction of energy consumption for a day in a month can be done which can incorporate other energy resources in a smart home. This prediction will not only be helpful to the consumers but also to the utilities so that they can plan ahead on the amount of energy that will be required so as to avoid uninterrupted power supply even at the time of outages.

REFERENCES

[1] Medhat Rostum, Hassan Moustafa, Amr Zamel, Ibrahim Zieden. Electrical load forecasting: A methodological overview. *International Journal of Engineering & Techology,* vol. 9, no 3, art. 842, Oct. 2020. http://dx.doi.org/10.14419/ijet.v9i3.30706

[2] A.Ranjani, B.S.E.Zoraida. Intelligent residential energy management in smart Grid. *Indian Journal of Science and Technology*, vol. 9, no. 45, 2016, pp. 1-8.

[3] I.Khan, Y.Xu, H. Sun, V. Bhattacharjee. Distributed optimal reactive power control of power systems. *IEEE Access*, vol. 6, 2018, pp.7100-7111.

[4] Kei Hirose. Interpretable modelling for short-and medium-term electricity demand forecasting. *Frontiers in Energy Research*, vol. 9, December 2021, https://doi.org/10.3389/fenrg.2021.724780.

[5] Khan, Irfan, Yinliang Xu, Soummya Kar, Mo-Yuen Chow, Vikram Bhattacharjee. Compressive sensing and morphology singular entropy-based real-time secondary voltage control of multi-area power systems. *IEEE Transactions on Industrial Informatics*, 2018.

[6] R.N. Calheiros, E.Masoumi, R.Ranjan, R.Buyya. Workload prediction using ARIMA model and its impact on cloud applications' QoS. *IEEE Transactions on Cloud Computing*, vol. 3, no. 4, 2015, pp.449-458.

[7] Bogdan Gilev, Nikolay Hinov, Stanislav Stefanov. Software implementation of methods for optimal design of power electronic devices, *International Journal on Information Technologies and Security*, vol. 15, no. 2, 2023, pp. 35-46. https://doi.org/10.59035/SZAK4809

[8] Saiful Islam, Lukasz Rojek, Michael Hartmann, Goran Rafajovski. Artificial intelligence in renewable energy systems based on smart energy house. *International Journal on Information Technologies and Security*, vol. 12, no. 4, 2020, pp. 3- 12.

[9] Yusuf Essam, Yuk Feng Huang, Jin Lin Ng, Ahmed H., Ali Najah Ahamed, Ahamed El-Shafie. Predicting streamflow in Peninsular Malaysia using Support Vector Machine and deep learning algorithms. *Scientific Reports*, vol. 12, 3883, 2022. https://doi.org/10.1038/s41598-022-07693-4

[10] Usman Ali, Abdul Rauf, Umer Iqbal, Ijaz Ali Shoukat, Abu ul Hassan. Big Data Analytics for a novel electrical load forecasting technique, *International Journal on Information Technologies and Security*, vol. 11, no. 3, 2019, pp. 33-40.

[11] Sofiene Haboubi Salem. Predictive analysis for energy consumption in smart homes with fog and cloud computing using Support Vector Regression. *International Journal on Information Technologies and Security*, vol. 14, no. 1, 2022, pp.49-60.

[12] Ahmed A. Alsheikhy. Evaluation of power dissipation and energy in a robotic system, *International Journal on Information Technologies and Security*, vol. 14, no. 2, 2022, pp.51-62.

[13] Arash Moradzadeh, Sahar Zakeri, Maryam Shoaran, Behnam Mohammadi-Ivatloo, Fazel Mohammadi. Short-term load forecasting of microgrid via Hybrid Support Vector Regression and long short-term memory algorithms. *Sustainability-Smart Grid Microgrid Systems*, vol. 12, no. 17, 7076, Aug 2020. https://doi.org/10.3390/su12177076

[14] J, Jasmine Christina Magdalene and B.S.E.Zoraida. Prediction of energy consumption from renewable solar resources in a smart home using Markov chain. *International Journal of Computer Science and Network Security*, vol. 22, Sep 2022, pp. 433-438.

Information about the authors:

Dr. B. S. E. Zoraida is an Assistant Professor in the Department of Computer Science, Bharathidasan University, Tiruchirappalli, India. She did her Ph. D at National Institute of Technology, Trichy. She has 18 years of teaching experience and 14 years of profound experience in Research. She has more than 40 research publications to her credit. Her areas of interests include DNA Computing, Bio-inspired Algorithms in Energy Management and Semantic Web.

J. Jasmine Christina Magdalene is an aspirant Research Scholar who works as an Assistant Professor in the Department of Computer Applications, Bishop Heber College, and Affiliated to Bharathidasan University, Tiruchirappalli, India. She has 14 years of teaching experience, and her areas of interests include Data Analytics, Internet of Things, Cloud Computing and Machine Learning.

Manuscript received on 11 June 2023