Engineering and Applied Science Research 2021;48(1):73-82 Research Article

EASR

Engineering and Applied Science Research

https://www.tci-thaijo.org/index.php/easr/index

Published by the Faculty of Engineering, Khon Kaen University, Thailand

Using a neural network model to determine electricity sales under renewable energy systems penetration consideration

Desmond Eseoghene Ighravwe* and Daniel Mashao

Department of Electrical and Electronic Engineering Science, University of Johannesburg, South Africa

Received 30 January 2020
Revised 21 May 2020
Accepted 23 June 2020

Abstract

Business systems will experience new data-driven models for their performance evaluation in the coming years, especially systems with stochastic characteristics. This development will benefit experts in energy management because more problems will be solved using machine learning algorithms - such as artificial neural networks (ANN). This research develops a machine-learning model for electricity sales using a single hidden layer ANN model. The developed model consists of six input parameters, including the number of renewable energy systems and households. This research used principal component analysis (PCA) algorithm to reduce the inputs to three parameters to improve the model performance. A TOPSIS (Technique for Order of Preference by Similarity to Ideal Solution) method was used to select the most suitable predictive models between SVR (support vector regression) and ANN. Data sets from a community in Lagos, Nigeria, were used to test the developed model performance. This research observed that a SVR model with a linear function performed better than an SVR model with a radial basis function or polynomial kernel. On the other hand, an ANN with 15 neurons outperformed ANN models with fewer nodes. The selected ANN model training and testing mean square errors are 0.00007 and 0.00028, respectively. This research recommends PCA for input parameters selection during electricity sales prediction based on the developed sales model performance.

Keywords: Electricity sales, Neural network, Support vector regression, Principal component analysis, Renewable energy system

1. Introduction

Energy diversification has encouraged several communities to explore alternative means of electricity generation for households and industrial purposes. It has not only improved electricity supply, but it has also reduced the volume of CO_2 emitted into our atmosphere. Also, energy diversification has reduced household dependence on utility firms for their daily energy needs, especially households in remote communities. This is because of innovative works in renewable and biodegradable energy systems. While researchers have reported that investment in these systems is high compared with non-renewable energy systems, their low long-term operation and maintenance cost encourage investors to acquire them. Several studies have reported that we can use these systems for small, medium, and large-scale purposes. These systems can improve small and medium scale enterprises profitability in communities where medium-class citizens live.

To reduce households' dependence on utility firms, the possibility of using renewable and biodegradable energy systems to support households' electricity needs have been reported. Some of these reports have presented information on the usefulness of solar photovoltaic (PV) systems for remote communities in developing countries. For example, there are reports on portable wind turbine systems for essential electricity needs in remote communities. Currently, researchers are examining the role of retrofitting energy systems in households' energy consumption management. Scholars believe that synergy between these systems and renewable and biodegradable energy systems will reduce a community's electricity demand from a national grid. However, this reduction does not follow a linear because renewable and biodegradable energy systems outputs are stochastic because of feedstock availability. This problem makes a utility firm's electricity sales to follows a stochastic pattern.

Despite the poor performance of several utility firms in developing countries, the stochastic nature of renewable and biodegradable energy systems outputs has made several households have these systems to maintain their connections to these firms. Utility firms' poor performance has been associated with administrative and technical problems. These firms' administrative issues are not limited to poor energy tariff system, skilled workforce shortage and unprofessional behaviours [1, 2]; studies have shown that inadequate electricity generation, supply, and faulty transformers affect these firms' performance. Beyond these issues, the living standard of a community affects these firms performance. Energy policy is also a critical parameter that affects these firms' performance [3] - especially in developing countries with energy policy inconsistency - such as a tax incentive [4]. These issues are responsible for policy inconsistency among utility firm that is operating in the same geographical space.

Despite the need to improve utility firms' sales analysis, few studies have used machine learning algorithms, as decision-making tools, for electricity sales prediction. Some studies on this subject matter have considered electricity consumption. This has made them focus more on customers' characteristics rather than

*Corresponding author.
Email address: ighravwedesmond@gmail.com
doi: 10.14456/easr.2021.9

on utility firms' chrematistics. An approach that focuses on both utility and customers' characteristics can capture energy management problems. For instance, it is possible to use administrative and technical constraints and customers' economic conditions to monitor electricity sales in a locality. The combination of electricity consumption and unit cost of electricity information allows organisations to know their expected revenue for a period. This information is useful for strategic planning, especially in this era of renewable energy penetration. As stated earlier, this penetration will cause a fluctuation in utility firms' energy revenue.

This objective of this research is to estimate energy sales using machine learning algorithms. SVR (Support vector regression) and ANN (artificial neural network), PCA (principal component analysis), and TOPSIS (Technique for Order of Preference by Similarity to Ideal Solution) are used to achieve this objective. In this research, SVR and ANN serve as regression tools, PCA serves as a data reduction tool, and TOPSIS serves as a model selection tool. This research's remaining sections are organised as follows: Section 2 contained information on energy revenue estimation, while section 3 described this research methodology. Section 4 presented a case study of the methodology application, and section 5 contained this research's concluding remarks.

2. Energy revenue

Researchers and practitioners have documented different energy revenue studies to improve utility business while ensuring that constant electricity is supplied to customers at affordable rates [1], [5, 6]. Some of these studies have used predictive models to investigate energy revenue problems, while others have used mathematical programming models to study the same issues. The remaining paragraphs in this section discussed some insights into these studies.

Ghajar and Khalife [1] analysed the impacts of deploying an automatic meter-reading system on utility firms' revenue. These authors used this system to address non-technical losses in a utility business. Using a utility firm in Lebanon as a case study, the system can make the firm break-even within 2.7 years. Gross et al. [3] evaluated the implications of energy policy on energy revenue. Their work showed that government interventions are required to manage the risk involved in the energy business. For example, Gross et al. [3] recommended that proper cost-benefit analysis is required before deploying energy technology for a community. This approach can minimise a utility firm's revenue risk. These authors also reported that simple and complex models had been used to investigate utility business risks. Consideration of a hybrid energy system often increases the complexity of utility firms' business models. Complex models can show the impact of energy mix on electricity sales. For example, Green et al. [5] reported that revenue energy systems proliferation would depress energy price.

Cai et al. [4] reported that as the investment cost for PV system price keeps dropping, it will reduce the number of households dependent on utility firms' service. These authors believe that utility firms will still break-even because fixed cost is the significant expense that utility firms incur during electricity generation. Mayr et al. [2] investigated the impact of renewable energy systems on electricity sales in South Africa. Using an optimisation model, the authors analysed the effect of a solar photovoltaic (PV) system and energy tariff on electricity sales. The model's goal is to minimise households' electricity consumption. These authors observed that PV and battery systems deployment in a community reduces electricity that the community will purchase from a utility firm. [7] reported that as households adopt PV, utility firms' return-on-equity will reduce; however, we will experience a reduction in greenhouse gas emission [8].

Since renewable energy systems penetration will cause a shortfall in electricity sales, researchers have investigated the possibility of using an optimal plant size to provide electricity for a community. These authors intend to improve utility firms' capacity to break-even on time, especially in an era where electricity price is stochastic. Lund and Anderse [6] used different electricity sales price to determine an optimal combined heat and power plant size. Technical, their work use of a switch on and off plant's concept to maximise its profit. Berry [9] used a least-squares regression to predict electricity sales. The developed model considered five input parameters, including a change in economic activity and energy efficiency programme. Their work revealed that an inverse relationship exists between energy efficiency programmes and electricity sales.

3. Methodology

This section discusses the model's parameters, machine learning algorithms for ANN and SVR, principal component analysis and TOPSIS. In this study, the parameters in Table 1 are used for electricity sales prediction. Figure 1 shows the ANN model for this sales prediction, while Equation (1) gives the expression for this sales prediction.

$$y = f(x_1, \dots, x_6) \tag{1}$$

3.1 Machine learning algorithms

Several machine learning algorithms have been used to address energy planning problems; however, ANN and SVR are the two algorithms that researchers have been used to produce practical results for energy management problems. Hence, this research used these algorithms to develop energy revenue models.

3.2 Artificial neural network

As we embrace the fourth industrial revolution, ANN applications will continue to grow because of ANN algorithms for deep learning problems – such as a convolution neural network. These algorithms have found wide application in energy management literature because of its ability to model the nonlinear relationship among system parameters. These algorithms generate the nonlinear relationship using activation functions, such as softmax (Equation 2), sigmoid (Equation 3) and tanh (Equation 4). The activation function for an ANN problem is based on the problem being solved - regression or classification. For regression problems, an output-layer activation function can be a sigmoid activation function. On the other hand, classification problems often use softmax as their activation function for an ANN output layer.

$$\sigma(z_i) = \frac{e^{z_i}}{\sum_{j=1}^{n} e^{z_j}} \tag{2}$$

$$z = \frac{1}{1 + e^{-z_j}} \tag{3}$$

$$z = 2 \times sigmoid(2x) - 1 \tag{4}$$

An ANN model's layer fires an activation function based on neurons' linear combination, connecting weights, and bias (Equation 5). For a multi-layer perceptron, the firing of activation starts from the first hidden layer and ends at the output layer. While it is rare for this model to have one input and hidden layers, the output layer can have one node - especially when dealing with regression problems. On the other hand, classification or ranking problems often have a minimum of two or more output nodes - especially when dealing with supervised learning problems.

Table 1 Description of the input and output parameters

Parameter	Description
Administrative (x_1)	It covers the non-technical factors that affect customers' decisions to use a utility firm's service as a source of electricity for their households. These factors are not limited to poor tariff plans and abnormal billing systems [1]. Also, higher reconnection and connection charges reduce households' dependence on electricity from utility firms.
Technical (x_2)	It covers the technical factors that affect customers' decision to use a utility firm's service as an electricity source for their households. These factors are not limited to faulty transfers and lack of electricity supply.
Economic (x_3)	It covers economic factors that affect customers' demand for electricity, such as a poor living standard. Also, energy tariff affects households electricity consumption [2], [9].
Energy policy (x_4)	It considers government policy implications on electricity consumption and revenue at a household level [3]. For example, the policy of renewable energy system adoption has an impact on electricity consumption.
Households (x_5)	The number of households depends on a utility firm for electricity [2], [4].
No. of RE systems (x_6)	It is the total number of functional renewable energy systems in a community [2]. Improvement in households' confidence increases a community's dependence on these systems.
Electricity sales (y)	This is the quantity of electricity sold to households in a community within a period [2]; it excludes revenue from reconnection fees.

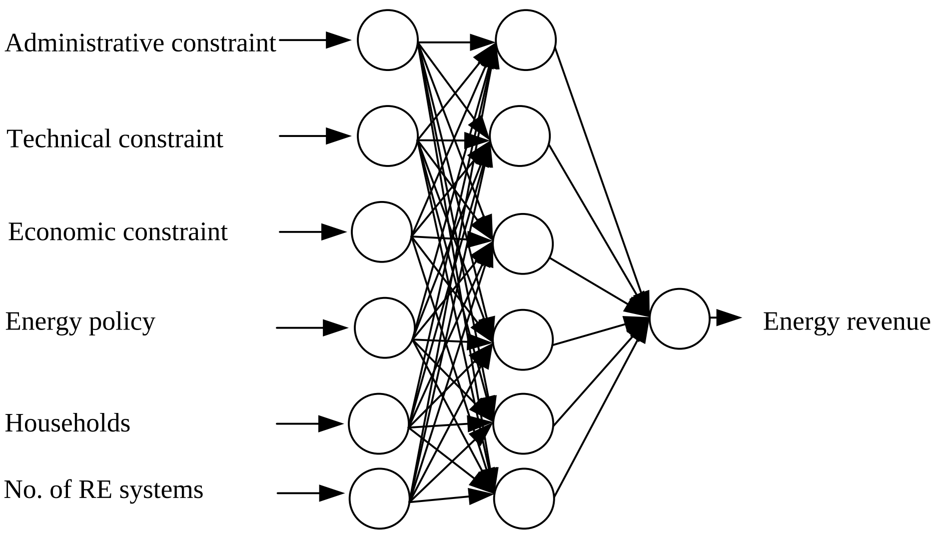

Figure 1 Proposed energy revenue model for Case I

$$z_j = \sum_{i=1}^{m} x_i w_{ij} + b_j \tag{5}$$

where, x_i and z_i denote input and output parameter i, respectively, b_j denotes bias for neuron j, w_{ij} denotes connecting weight from node i to node j.

When ANN models are trying to learn problems - which could be regression or classification problems, examples are divided into training, validation and training sets. This makes it possible for developed models to reduce the difference between targeted and predicted values, when these algorithms are used to solve real-life problems. However, most academic articles limit data splitting to training and testing sets. The former set is used to train models based on statistical measures (Equations 6 to 7), the latter set is used to evaluate developed models' performance. During data splitting into training and testing sets, the largest part represents the training sets. For example, some researchers and practitioners often use a ratio of 70:15:15 to divide data sets into training, validation and training sets [10]. On the other hand, a ratio of 80:20 is used to split data sets into training and training sets [11].

Two main approaches are used to train an ANN model. A first approach is an online approach; it involves using a data tuple's prediction error to adjust a model's connecting weights. A second approach is an off-line approach. It involves the use of a batch of data tuple to adjust the weights of the model. For both approaches, special algorithms, such as gradient descent and sequential gradient descent, are used to determine an epoch's weight [12]. Based on any of the statistical measures or performance indicators in Equations (6) to (8), a trained ANN model's performance is validated using other algorithms - such as a linear regression model, SVR and random forest.

$$MSE = \sqrt{\frac{\sum_{i=1}^{m}(A_i - P_i)^2}{m}} \tag{6}$$

$$MAE = \frac{1}{m}\sum_{i=1}^{m}|A_i - P_i| \tag{7}$$

$$r = \frac{\sum_{i=1}^{m}(A_i-\bar{A})(P_i-\bar{P})}{\sqrt{\sum_{i=1}^{m}(A_i-\bar{A})^2(P_i-\bar{P})^2}} \tag{8}$$

where, *MSE*, *MAPE* and *r* represent mean squared error, mean absolute percentage error and correlation coefficient, respectively, and A_i and P_i denote the actual and predicted values of data tuple *i*, respectively.

Before model training or testing, data sets preprocessed. In machine learning, this process is pivotal to the success of any machine learning algorithm. It covers issues that are not limited to input parameters selection, parameter reduction, missing data points analysis and parameters normalisation [13]. While the decision to select the number of potential inputs depends on domain knowledge, data reduction is based on scientific procedures. PCA and factor analysis are used to reduce the input parameters for a model. Researchers and practitioners use this process to reduce the computation cost and to improve machine learning algorithms performance. However, the cut-off point for reducing input parameters still depends on domain knowledge. Missing data analysis is a special body of knowledge in machine learning study. While some researchers and practitioners have relied on the use of mean value approach or other approaches to solve this issue, others remove data tuples with missing points from data sets [14-16].

One problem of removing data tuples is that it can make data sets to lose some of its properties, especially for a case where the data sets are small. After ensuring that data sets do not contain missing data points, data are normalised to ensure that all the inputs are within the same range. This process eliminates the possibility of one or more inputs from influencing the outputs of machine learning algorithms. Also, it helps to reduce the memory requirement to improve a machine learning algorithm and the algorithm computation time.

3.3 Support vector regression

Researchers and practitioners are attracted to a SVR algorithm because it uses a few data points to generate predictive models with practical relevance across disciplines. SVR is proposed as an extended version of a support vector machine [17]. In support vector machines, a hyperplane is used to separate classes, while SVR uses this plane to predict dependent parameters [18]. Figure 2 shows the relationship between the actual and predicted values of an SVR model [17]. Before training model can commence, users specify a model boundary lines - these are the dotted lines in Figure 2, [17]. A constant parameter "C" is used during model training to prevent SVR models from an over-fitting problem.

Based on these data point, a linear (linear kernel) or nonlinear (radial basis function- RBF) equation is used to minimises the difference between the actual and predicted values (Equation 9). This equation is subject to the distance between the bounds and predicted values outside the bounds and learned weights [17].

$$\min \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}(\xi_i^* + \xi_i) \quad (9)$$

St.

$$\begin{Bmatrix} y_i - \langle w, x_i \rangle - b \leq \in + \xi_i^* \\ \langle w, x_1 \rangle + b \leq \in + \xi_i \end{Bmatrix} \quad (10)$$

3.4 Principal component analysis

In 1933, [19] proposed PCA as an algorithm that uses the covariance among parameters to explain the importance of parameters that influence a system's output. Its application creates an opportunity to reduce the input parameters' dimension for mapping problems at minimal information loss [20, 21]. PCA is used in the machine learning domain to reduce machine learning algorithms' computational cost, especially solution quality and computational time. And it has been used to reduce data size for pattern recognition problems [22]. Because of its unique benefits, several statistical software has PCA as a package - this helps eliminate manual computation rigour during PCA implementation. Given that the parameters (n) that denote system variability can be expressed with a linear combination that is uncorrelated (Equation 11), PCA can be used to $k < n$ [23].

$$\Omega_k = \sum_{i=0}^{n}(a_{ik}x_i) \quad (11)$$

$$\sum_{i=0}^{n} a_{ik} = 1 \quad (12)$$

where Ω_k denotes the *k-th* principal component and a_{ik} denotes the coefficient of parameter i for the *k-th* principal component [23].

3.5 TOPSIS

Multi-criteria decision-making (MCDM) literature has established that TOPSIS method is among the most used MCDM tools for multi-disciplinary studies, which energy management is a sub-set. This tool uses the relationship between the distance ideal and non-ideal solutions to rank alternatives solution for MCDM problems [24]. Hence, the ranking of alternatives is based on their closeness to these solutions. Technically, the ranking process is based on four steps. The steps are data normalisation, determination distance ideal and non-ideal solutions, distance ideal and non-ideal solutions, and closeness coefficient [25].

Step 1: Data normalisation is a process used to reduce criteria numeric values within a specific range during a MCDM tool's applications. As at today, several approaches that can be used to normalised data have been proposed (Equations 13a and 13b). One unique feature of this approach is that it reduces data values between 0 and 1 [24, 25]. Another feature is the criteria orientation consideration - cost or benefit-based [26].

$$x_{ij} = \begin{cases} \dfrac{r_{ij} - r_j^{min}}{r_j^{max} - r_j^{min}} & \forall r_{ij} \in Benefit\ \{B\} \\ \dfrac{r_j^{max} - r_{ij}}{r_j^{max} - r_j^{min}} & \forall\ r_{ij} \in Cost\ \{C\} \end{cases} \quad (13a)$$

$$x_{ij} = \frac{r_{ij}}{\sum_{j=1}^{n}(r_{ij})^2} \quad (13b)$$

where, r_{ij} and x_{ij} denote the real and normalised values of performance measure j for alternative i, respectively.

Step 2: Using a criterion orientation, its ideal and non-ideal solutions are determined. The ideal solution for benefit-based criteria is taken as the higher-the-better, while cost-based criteria are taken as the lower-the-better [26, 27]. On the other hand, not-ideal solutions for benefit-based criteria are taken as lower-the-better (Equations 14), while cost-based criteria are taken as higher-the-better (Equations 15). Some studies have argued that the importance of criteria should precede their ideal and non-ideal solutions determination. Equations (14) and (15) represent the expression for criteria ideal and non-ideal solutions, respectively.

$$x_j^+ = \min_{\forall j \in C}(x_{ij}),\ \max_{\forall j \in B}(x_{ij}) \quad (14)$$

$$x_j^- = \max_{\forall j \in C}(x_{ij}),\ \min_{\forall j \in B}(x_{ij}) \quad (15)$$

Step 3: For classical TOPSIS method, the Euclidean distance between the normalised or weighted normalised values and the ideal and non-ideal solutions for a MCDM are used to determine alternatives' distance from ideal and non-ideal solutions (Equations 16 and 17).

$$d_j^+ = \sqrt{\sum_{j=1}^{K} (x_{ij} - x_j^+)^2} \tag{16}$$

$$d_j^- = \sqrt{\sum_{j=1}^{K} (x_{ij} - x_j^-)^2} \tag{17}$$

where, d_j^- and d_j^- denote the distance of alternative's i from the ideal and non-ideal solutions, respectively.

Step 4: This method uses the relationship between an alternative's ideal and non-ideal solutions to determine an alternative closeness coefficient (Equation 18). This coefficient serves as a basis for ranking alternatives. The most and least suitable alternatives are the alternatives with the highest and lowest coefficients, respectively [26, 27].

$$cc_i = \frac{d_i^-}{d_i^+ + d_i^-} \tag{18}$$

Based on the various mathematical tools described in this section, an outline for these tools application for electricity sales model selection is presented as follows:

Step 1: Consult a panel of experts to implement the proposed model in [28].
Step 2: Use the model to generate data sets for electricity sales and its corresponding inputs.
Step 3: Use PCA to determine the importance of the inputs.
Step 4: Identify the inputs that contributions at least 80 and 50% to the study of electricity sales.
Step 5: Use the identified inputs parameters to construct different electricity models.
Step 6: Select a data partitioning into training and testing sets.
Step 7: Train the electricity models using different SVR kernels.
Step 8: Train the electricity models using different ANN hidden layer nodes.
Step 9: Use TOPSIS to identify the best SVR kernels for the different electricity models.
Step 10: Use TOPSIS to determine the best ANN architectures for the different electricity models.
Step 11: Compare the best SVR and ANN models.
Step 12: Make a recommendation.

4. Case study

The implementation of the proposed model was in a local government area in Lagos, Nigeria. Using [28] simulation model, we generated 90 data points for the selected input and output parameters. This research implemented the simulated model using Vensim software package. We consulted experts to provide guidelines for the maximum and minimum values of some parameters in the selected simulation model. The experts provided information based on an urban community's attributes in Victoria Island, Lagos, Nigeria (Figure 2). We selected this community because of its capacity to acquire renewable energy system and the possibility of an independent power plant to break-even in the community. This research used a ratio of 80:20 to split simulated data sets into training and testing sets. This research uses MSE and MAE to evaluate and validate developed ANN and SVR models. Figures 3a to 3f show the input parameters profile, while Figure 4 shows the output parameter profile. This research studied different single-hidden layer ANN models for energy revenue estimation. By varying the number of nodes, it developed the different ANN models. Also, this research considered SVR three model using linear, polynomial, and radial basis function kernels. Three cases were considered during the proposed model application. The first considered six input parameters (Case I), the second case considered four input parameters (Case II) and the third case considered three input parameters (Case III).

i. PCA results

Using input data sets in Figure 3a to 3f, PCA generated the results in Table 2. This table shows that the least and most important inputs for the electricity prediction problem are administrative and technical parameters, respectively. This research used a mark-off point of 80% to reduce the input parameter size. The results obtained showed that four input parameters fail within this point. These input parameters are administrative, energy policy, the number of households, and the number of renewable energy systems. Figures 5 and 6 show the ANN model's structure for the electricity sales prediction problem for II and III cases. Also, a mark-off of 50% was considered. And it was observed that three input parameters felt within this point; the inputs are an administrative constraint, number of households, and renewable energy systems. Figure 5 shows the structure of the number ANN model for the electricity sales prediction problem.

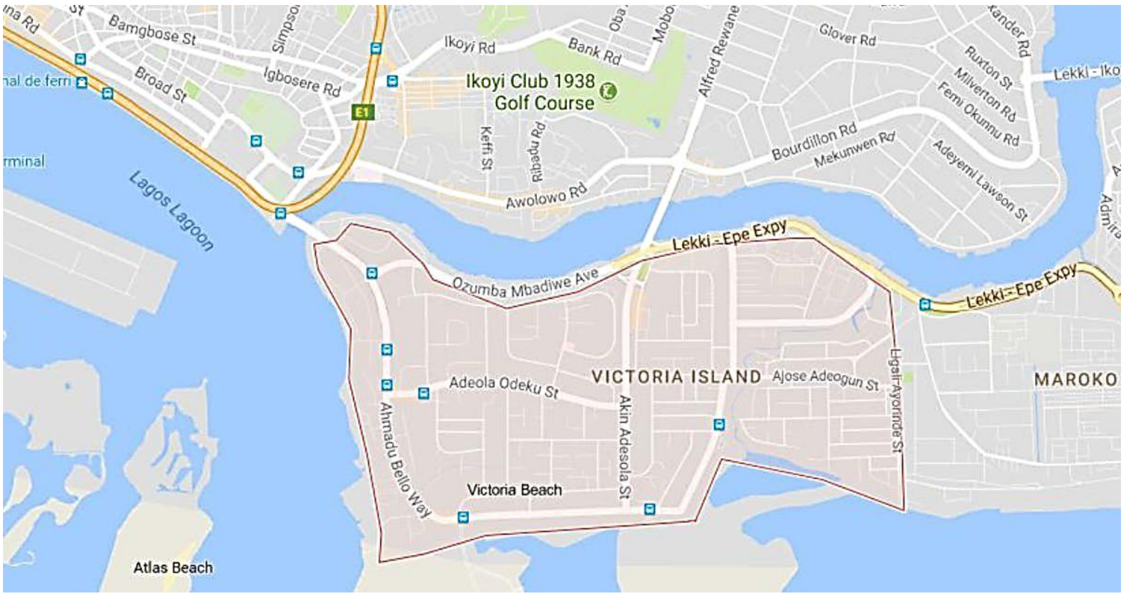

Figure 2 Map of the case study [29]

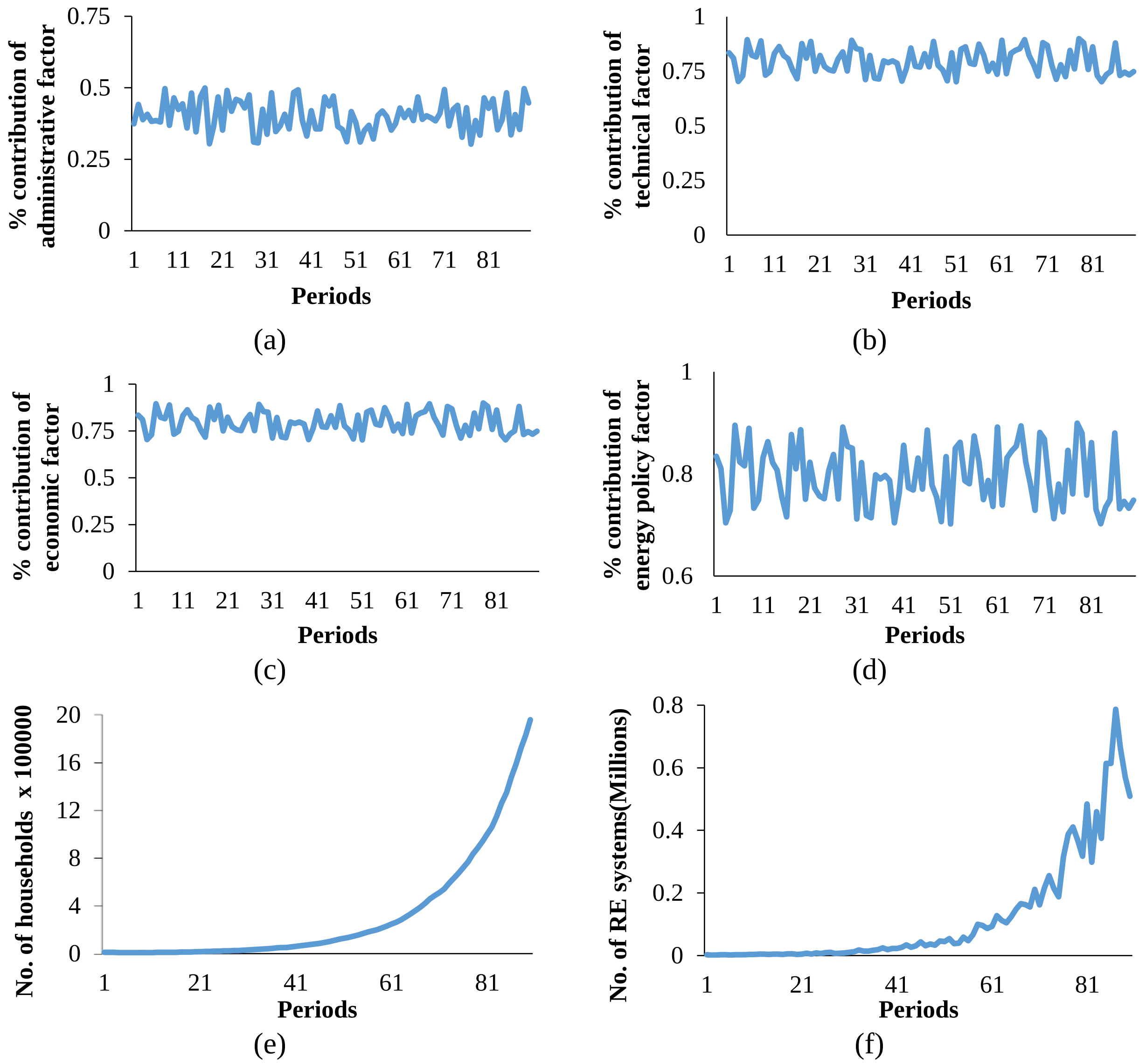

Figure 3 (a) Profile of administrative factor; (b) Profile of technical factor; (c) Profile of economic factor; (d) Profile of energy policy factor; (e) Profile of numbers of households and (f) Profile of numbers of RE systems.

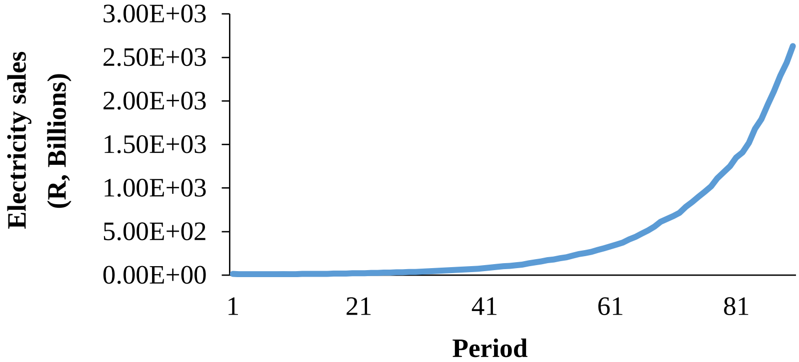

Figure 4 Profile of electricity sales

Table 2 PCA results

Parameters	Actual PCA	Normalised PCA
X1	0.7876	0.2400
X2	0.0958	0.0292
X3	0.4703	0.1433
X4	0.5419	0.1651
X5	0.6041	0.1840
X6	0.7827	0.2385

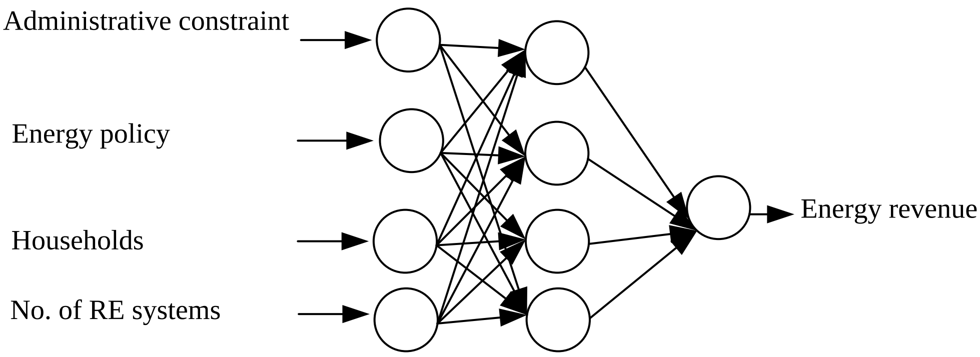

Figure 5 Proposed energy revenue model for Case II

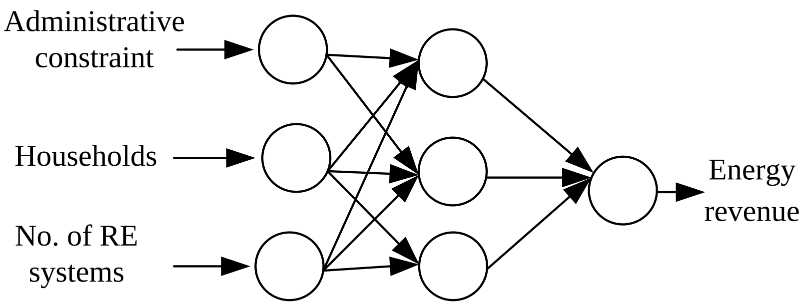

Figure 6 Proposed energy revenue model for Case III

Table 3 SVR results for electricity sales prediction

		MSE		MAE	
		Training	Testing	Training	Testing
Case I	RBF	0.0063	0.0209	0.0716	0.0981
	Linear	0.0036	0.0027	0.0516	0.0455
	Polynomial	0.2017	0.3620	0.2841	0.4072
Case II	RBF	0.0052	0.0293	0.0656	0.1113
	Linear	0.0030	0.0020	0.0465	0.0381
	Polynomial	0.3294	0.3364	0.3555	0.3890
Case III	RBF	0.0051	0.0123	0.0646	0.0839
	Linear	0.0052	0.0039	0.0694	0.0579
	Polynomial	0.3169	0.3309	0.3598	0.3778

Table 4 Closeness coefficient for the developed SVR models

	Case I	**Case II**	**Case II**
RBF	0.7272	0.8138	0.8486
Linear	1.0000	1.0000	0.9924
Polynomial	0.0000	0.0000	0.0000
Case selection	0.4880	1.0000	0.0000

ii. SVR results

This research used Scikit-learn package in Python to implement the SVR models for the three cases [30]. Table 3 shows the SVR results for the three cases. For cases I and II, the linear trained SVR model outperformed the RBF and polynomial trained SVR models for the training and testing MSE. On the other hand, the RBF trained SVR model performed better than the RBF and polynomial trained SVR model for the training and testing MSE. These observations are the same for the MAE results (Table 3). For Case III, the RBF trained SVR model training MSE and testing MAE are better than the linear and polynomial trained SVR models. On the other hand, the linear trained SVR model's training MSE and testing MAE are better than the RBF and polynomial trained SVR models.

To select the most suitable case for the current problem, we used the standard TOPSIS method described in Section 3.4 to aggregate the results in Table 3. For the current application, the importance of the statistical measure is the same (0.25).

Table 4 shows a summary of the SVR models' results. The TOPSIS method ranked the most (linear) and least (polynomial) suitable kernel for the SVR models for cases I to III as the same. Using the linear kernel performance to select the most suitable case for the current problem, Table 4 shows that Case II performed better than cases I and III.

iii. ANN results

This research used Scikit-learn, Keras and Tensor packages to implement the single layers ANN model for the three cases. This research evaluated 15 different hidden nodes structure to

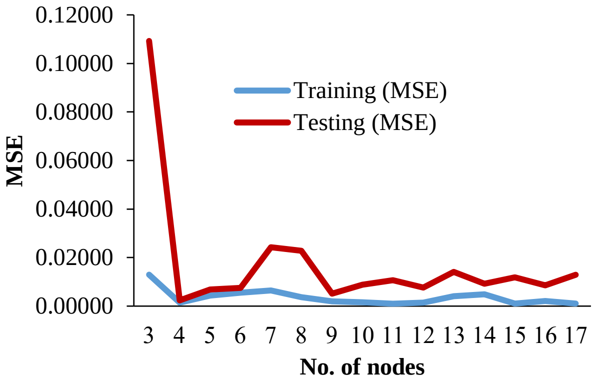

Figure 7 MSE for Case I

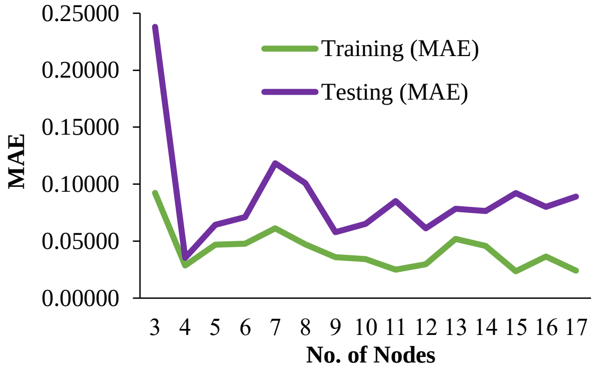

Figure 8 MAE for the Case I

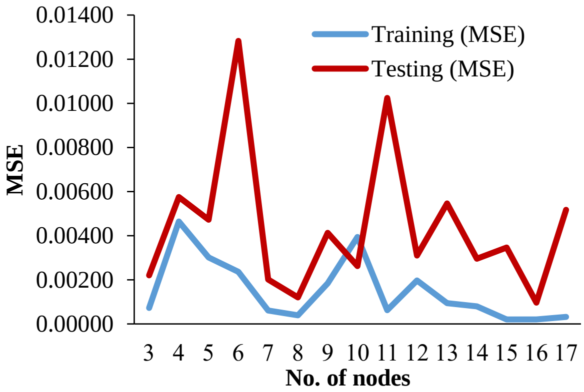

Figure 9 MSE for Case II

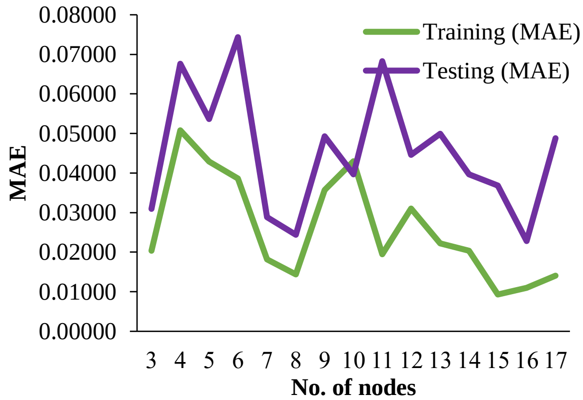

Figure 10 MAE for Case II

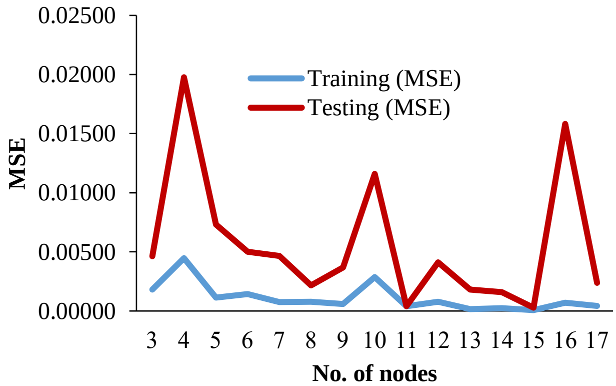

Figure 11 MSE for Case III

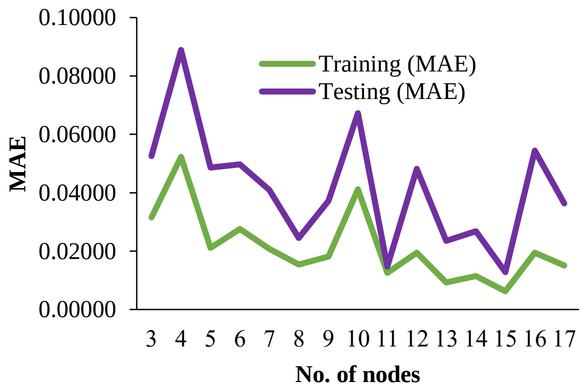

Figure 12 MAE for Case III

select the most suitable ANN model for the three cases - ANN model with three, four, and six inputs, respectively. Also, it evaluated the ANN models performance using MSE and MAE. Figures 7 to 12 show the performance of the ANN models for the three cases.

This research aggregated the results in Tables 3 to 4 using the classical TOPSIS method. Table 5 shows the TOPSIS results. From this table, it can be deduced that the most and least suitable number of nodes for Case I is nine and three nodes, respectively. For Case II, the models with 16 and 4 nodes are the most and least suitable models for electricity sales prediction. Furthermore, Table 5 shows that the most and least suitable number of nodes for Case III is 15 and 4, respectively.

This research used TOPSIS to rank the models for the electricity sale prediction (Table 6). The TOPSIS results show that the three-input model is the most suitable electricity prediction model; while the six-input model is the least suitable electricity model. The selected ANN model performed better when compared with the selected SVR model (Table 7). This research, therefore, recommends the ANN in Figure 5 for the electricity sales prediction problem.

iv. Managerial implications

This research's findings have several managerial implications. It has shown that machine learning algorithms can predict electricity sales from a multivariate perspective. It has also demonstrated that machine learning algorithms' capacity to predict electricity sales depends on the input parameters. Lastly, it has shown that three-input parameters can predict electricity sales better than a six-input parameters model.

5. Conclusions

Parameters selection is one of the building blocks of machine learning models development. Hence, this research used principal component analysis (PCA) to determine the influential parameters for utility firms' revenue prediction. The initial model selected six input parameters, but this research reduced it to four parameters using PCA. This research observed that the combined PCA and ANN improved the electricity revenue prediction. According to the developed revenue model performance, machine-learning models will improve energy management

Table 5 ANN results for the cases

S/n	Nodes	Case I	Rank	Case II	Rank	Case III	Rank
1	3	0.0000	15	0.7977	5	0.4855	14
2	4	0.8439	6	0.1048	15	0.0000	15
3	5	0.8466	5	0.3403	13	0.5711	10
4	6	0.8213	7	0.1701	14	0.5323	11
5	7	0.6150	14	0.8416	3	0.6504	8
6	8	0.6988	13	0.9216	2	0.8361	5
7	9	0.8931	1	0.4492	11	0.7006	7
8	10	0.8661	3	0.4942	10	0.2811	14
9	11	0.7939	9	0.4044	12	0.9286	2
10	12	0.8868	2	0.5457	9	0.5900	9
11	13	0.7819	10	0.5590	8	0.8774	3
12	14	0.8048	9	0.7004	6	0.8367	4
13	15	0.7658	12	0.8001	4	1.0000	1
14	16	0.8061	8	0.9757	1	0.5085	12
15	17	0.7764	11	0.6304	7	0.7256	6

Table 6 Classical TOPSIS results for the number of input selection

Case	No. of node	Training (MSE)	Testing (MSE)	Training (MAE)	Testing (MAE)	Closeness coefficient
I	9	0.00201	0.00515	0.03598	0.05782	0.48948
II	16	0.00020	0.00097	0.01096	0.02282	0.79577
III	15	0.00007	0.00028	0.00625	0.01279	1.00000

Table 7 Selected SVR and ANN model results

Inputs	Model	Training (MSE)	Testing (MSE)	Training (MAE)	Testing (MAE)
4	SVR	0.00300	0.00200	0.04650	0.03810
3	ANN	0.00007	0.00028	0.00625	0.01279

decision-making process. Also, these models will not only improve energy utility firms' performance, but it will also improve other utility firms' performance, such as water cooperation and waste management.

Some of this research's contribution to energy literature is:

- It uses principal component analysis algorithm to reduce the dimension of predictive models for electricity sales.
- It has shown that three or four input parameters can predict electricity sales under renewable energy consideration.
- It compared the capacity of SVR and ANN as predictive models for electricity sales.
- It uses TOPSIS, an MCDM tool, to select the most suitable machine learning model for electricity sales prediction.

One of the limitations of this work is that it did not consider a specific renewable energy system. Hence, our future research will use the dichotomy between renewable and biodegradable energy resources impact on electricity sales. Our future research will consider the impact of renewable energy systems' stochastic outputs on electricity sales in developing countries. We will also investigate the effect of connection and reconnection fees on electricity revenue in our future research. Having demonstrated the ANN performance for the current problem, we recommend developing machine learning models for a multi-stage energy tariff plan effects on electricity sales as a further study.

6. References

[1] Ghajar RF, Khalife J. Cost/benefit analysis of an AMR system to reduce electricity theft and maximise revenues for Electricite du Liban. Appl Energ. 2003;76(1-3):25-37.

[2] Mayr D, Schmid E, Trollip H, Zeyringer M, Schmidt J. The impact of residential photovoltaic power on electricity sales revenues in Cape Town, South Africa. Util Pol. 2015;36:10-23.

[3] Gross R, Blyth W, Heptonstall P. Risks, revenues and investment in electricity generation: why policy needs to look beyond costs. Energ Econ. 2010;32(4):796-804.

[4] Cai DWH, Adlakha S, Low SH, De Martini P, Chandy KM. Impact of residential PV adoption on retail electricity rates. Energ Pol. 2013;62:830-43.

[5] Green RJ, Léautier TO. Do costs fall faster than revenues? Dynamics of renewables entry into electricity markets. TSE Working Papers 15-591, Toulouse School of Economics (TSE). 2015:1-45.

[6] Lund H, Andersen AN. Optimal designs of small CHP plants in a market with fluctuating electricity prices. Energ Convers Manag. 2005;46(6):893-904.

[7] Satchwell A, Mills A, Barbose G, Wiser R, Cappers P, Darghouth N. Financial impacts of net- metered PV on utilities and ratepayers: a scoping study of two prototypical US utilities. USA: U.S. Department of Energy; 2014.

[8] Akinyele D, Olatomiwa L, Ighravwe DE, Babatunde OM, Monyei C, Onile, A. Evaluation of solar PV microgrid deployment sustainability in rural areas: a fuzzy STEEP approach. 2019 IEEE PES/IAS PowerAfrica; 2019 Aug 20-23; Abuja, Nigeria. USA: IEEE; 2019. p. 593-8.

[9] Berry D. The impact of energy efficiency programs on the growth of electricity sales. Energ Pol. 2008;36(9):3620-5.

[10] Theocharides S, Makrides G, Venizelou V, Kaimakis P, Georghiou GE. PV Production forecasting model based on artificial neural networks (ANN). 33rd European Photovoltaic Solar Energy Conference and Exhibition; 2017 Sep 25-29; Amsterdam, Netherlands. München: WIP; 2017. p. 1830-94.

[11] Mustafa M, Taib M, Murat Z, Sulaiman N. Comparison between KNN and ANN classification in brain balancing application via spectrogram image. J Comput Sci Comput Math. 2012;2(4):17-22.

[12] Oyebode O, Ighravwe DE. Urban water demand forecasting: a comparative evaluation of conventional and soft computing techniques. Resour. 2019;8(3):156.

[13] Anyaeche CO, Ighravwe DE. Predicting performance measures using linear regression and neural network: a comparison. Afr J Eng Res. 2013;1(3):84-9.
[14] Asparouhov T, Muthén B. Weighted least squares estimation with missing data. Mplus [Internet]. 2010;1-10. Available form: https://www.statmodel.com/download/GstrucMissingRevision.pdf.
[15] Collins LM, Schafer, JL, Kam CM. A comparison of inclusive and restrictive strategies in modern missing data procedures. Psychol Methods. 2001;6(4):330-51.
[16] Silva-Ramírez EL, Pino-Mejías R, López-Coello M, Cubiles-de-la-Vega MD. Missing value imputation on missing completely at random data using multilayer perceptron. Neural Network. 2011;24(1):121-9.
[17] Kleynhans T, Montanaro M, Gerace A, Kanan C. Predicting top-of-atmosphere thermal radiance using MERRA-2 atmospheric data with deep learning. Remote Sens. 2017;9(11):1133.
[18] Bhattacharyya I. Support vector regression or SVR [Internet]. 2018 [cited 2019 Oct 4]. Available from: https://medium.com/coinmonks/support-vector-regression-or-svr-8eb3acf6d0ff.
[19] Hotelling H. Analysis of a complex of statistical variables into principal components. J Educ Psychol. 1933;24(6): 417-41.
[20] Kramer MA. Nonlinear principal component analysis using auto-associative neural networks. AIChE J. 1991;37(2): 233-43.
[21] Ringnér M. What is principal component analysis? Nat Biotechnol. 2008;26(3):303-4.
[22] Smith LI. A tutorial on principal components analysis. Technical Report. New Zealand: University of Otago; 2002. Report No.: OUCS-2002-12.
[23] Su CT, Tong LI. Multi-response robust design by principal component analysis. Total Qual Manag. 1997;8(6);409-16.
[24] Behzadian M, Otaghsara SK, Yazdani M, Ignatius J. A state-of the-art survey of TOPSIS applications. Expert Syst Appl. 2012;39(17):13051-69.
[25] Peng Y, Wang G, Kou G, Shi Y. An empirical study of classification algorithm evaluation for financial risk prediction. Appl Soft Comput. 2011;11(2):2906-15.
[26] Ighravwe DE, Babatunde MO. Selection of a mini-grid business model for developing countries using CRITIC-TOPSIS with interval type-2 fuzzy sets. Decis Sci Lett. 2018;7(4):427-42.
[27] Lozano-Minguez E, Kolios AJ, Brennan FP. Multi-criteria assessment of offshore wind turbine support structures. Renew Energ. 2011;36(11):2831-7.
[28] Ighravwe DE, Mashao D. Development a system dynamic model for utilities revenue evaluation under distributed energy resources. The 2019 General Meeting and PIESA-IERE South Africa Forum; 2019 Oct 28-31; Sun City, South Africa.
[29] Propert P. Victoria Island Neighbourhood Guide [Internet]. 2017. [cited 2019 Oct 10]. Available from: https://www.privateproperty.com.ng/news/victoria-island-neighbourhood-guide/.
[30] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: machine learning in Python. J Mach Learn Res. 2011;12:2825-30.