Predicting Monthly Electricity Demand Using Soft-Computing Technique

Isaac Kofi Nti[1], Asafo-Adjei Samuel[2] and Agyemang Michael[2]

[12]*Department of Electrical & Electronic Engineering, Sunyani Technical University, Sunyani Ghana*
[1]*Department of Computer Science, Sunyani Technical University, Sunyani Ghana*

Abstract –Electricity is an essential commodity for all. The generation, transmission and distribution (GTD) of electrical energy needs much planning since every megawatt transmitted for distribution depends on consumers demand. An accurate estimation (prediction) of future demand helps prevent a power shortage and load scheduling ("Dumsor"). This study proposed a soft-computing technique based on multi-layer perceptron (MLP), support vector machine (SVM) and decision tree (DT) algorithms to predict a 30-day head electrical energy demand. Using three years of real-world historical weather and electrical energy demand data from Bono region of Ghana, we experiment with the proposed model. The obtained accuracy of 80.57% for DT, 95% for MLP and 67.2 for SVR and RMSE values of 0.064221, 0.021184 and 0.100776 for DT, MLP and SVR respectively revealed the efficiency of the proposed model in predicting future electrical load.

Key Words: **Multilayer Perceptron, Short –Term Electrical Load Prediction (STLP), Support Vector Regressor, Decision Tree, Bono-Region**

1. INTRODUCTION

In this 21st century, electricity plays an essential part in social and economic development; almost everything depends on electricity; the absence of electricity would have resulted in a stuck life [1]–[3]. Every electricity generating firm aim at providing excellent and efficient service with minimal loss [4]. For optimal planning and operation of the electric power system, the authorities need a proper evaluation of present-day and future electric power demand [5]. To achieve this electric power industry depends significantly on load-demand prediction. Predicting electrical Load demand is a central and integral process in the planning and operation of electric utility companies [6].

Electrical load prediction (ELP) is a critical issue in the electrical industry for effective management and efficient planning. Several prediction techniques exist, however selecting the right technique for establishing future load requirements is a non-trivial in itself [5]. Therefore, depending on the nature of load variations, one particular method may be superior to another. Overestimate or underestimate of electrical load prediction will significantly affect the revenue of the electric utility company.

The Overestimation of the long-term load could result in spending extra money in building new power stations to manage the extra load. Moreover, an underestimate of the load could cause troubles on the market for electrical providers and manufacturing shortage within the spinning reserve of the system, which will result in an insecure and unreliable system. Therefore, an efficient technique is required to predict electrical demand. Thus an accurate predictive model takes into consideration the factors that affect the expansion of the load over several years [7].

Despite the numerous works of literature on short term load prediction (STLP) published since the 1960s, studies in this area are still a dare to the electrical engineering researchers and professionals due to its difficulty. Thus, accurate estimation of future load with the past data remains a challenging task to date, notably, for load prediction of days with risky weather, holidays and other abnormal days. Employing the recent advancement in statistical models, and artificial intelligence algorithms and data mining tools make it potentially possible to improve the prediction result [8], [9].

Despite this, Ghana's power sector has suffered numerous challenges over the past decays resulting in considerable waves on the economic status quo of the country which brought about a familiar word 'Dumsor' among the Ghanaian populace [3], [10]. However, according to [11], the lack of adequate planning by electrical energy managers and generator lead to electrical energy shortage (Stavast, 2014).

Hence this paper seeks to propose a predictive model for identifying the factors that influence the demand of electrical energy, and predict a 30-day head electricity demand based on past weather parameters and electric load demands, using decision tree (DT), multilayer perceptron (MLP) and support vector regression (SVR). Thereby helping electrical energy generators and distributors to make and inform decisions ahead of time.

The remaining section of this paper is categorised as follows; Section 2 discusses the overview of machine learning, the requirement of electrical load prediction and review of some related works. Section 3 presents the study methodology. Section 4 presents the results and discussion, while section 5 concludes the study and direction for future studies.

2. LITERATURE REVIEW

This section discusses the general overview of machine learning, the basic concept and various approaches to electrical load (EL) prediction, factors which affect EL prediction; requirement of EL prediction and other related works on the same field of study.

2.1 Machine Learning (ML)

"Machine Learning is the process where, a computer program is said to learn from experience *E* concerning some class of tasks *T* and performance measure *P*, if its performance of tasks in *T*, as measured by P, improves with experience *E*", thus making machines behave like humans [12]. Machine learning is closely associated with computational statistics, which also focuses on prediction-making through the use of computers. ML can be grouped into three broad sections, namely supervised learning, unsupervised learning, learning and reinforcement learning.

2.2 Basic Concept of Electrical Load Prediction

The electrical load demand is the sum of all the consumers' demand at any given time, and the goal of electrical load prediction is to predict the future system load demand. Various factors influence the load demand behaviour, which includes weather, economy, random disturbance, customer category factors and time [9], [13], [14].

2.3 Requirements of Electrical Load Prediction

In most of the energy management systems (EMS) and load dispatch centres there is an STLP module, and a sound STLP system should fulfil the following requirements; accuracy, detection of bad data, user-friendly, speed and automatic forecasting.

In general, electrical load prediction can be clustered as short-term (a few hours), medium-term (a few weeks up to a year) or a long term (over a year). Table 1 shows the summary of types of load forecasting and their applications.

Table 1: Summary of Types of Load Prediction

Nature of Forecast	Lead Time	Application
Very short term	A few seconds to several minutes	Generation, distribution schedules, contingency analysis for system security.
Short term	Half an hour to a few hours	Allocation of spinning reserve; operational planning and unit commitment; maintenance schedule.
Medium term	A few days to a few weeks	Planning for seasonal peak-winter, summer
Long term	A few months to a few years	Planning generation growth

2.4 Prediction Techniques

Prediction techniques may be grouped into three (3) broad classes, namely extrapolation, correlation and a combination of both. Extrapolation is further divided into time series or traditional data-based method. Extrapolation techniques involve fitting trend curves to primary historical data adjusted to reflect the growth trend itself. With the curve, the prediction is obtained by evaluating the trends curve function at the desired future point. Correlation techniques for prediction, relate system load to various demographic and economic factors. This methodology is advantageous in compelling the analyst to understand clearly the interrelationship between the load growth patterns and other measurable factors. No one prediction method is useful in all situations.

2.5 Related Studies

This section presents a brief review of existing electrical load prediction studies.

Long short-term memory (LSTM)-based neural network (NN) enhanced with genetic algorithm (GA) for feature selection with a different configuration for predicting short-medium term load demand was proposed by [15]. An artificial neural network (ANN) based predictive model was proposed in [9], [16] for predicting the short-term load demand.

In another study, a curve fitting algorithm was proposed for the prediction of electrical load [17]. A DT based predictive model for short-term electrical load prediction was presented in [5]. In [1], the authors presented a convolutional neural network (CNN) and enhanced (SVR) for prediction of the Short-Term electricity demand [1]. A random forest (RF) and SVR based predictive model was proposed for predicting the hourly electrical load by [18].

A deep learning predictive model for predicting the long-term electrical load based on economic and social factors in the Kuwait Region was proposed by [6]. A day head electricity demand Fuzzy Logic based predictive model was proposed by for predicting Albania power demand [13].

3. METHODOLOGY

This section presents the procedures and steps adopted for the current study.

3.1 Study Design

Figure 1 shows the dataflow diagram of the current study. Four main phases are considered, phase 1 covers the data collection phase, data integration and data preprocessing phase; phase 2 deals with data partitioning, phase 2 deals with the building, training and testing of models and phase 4 covers model performance measure.

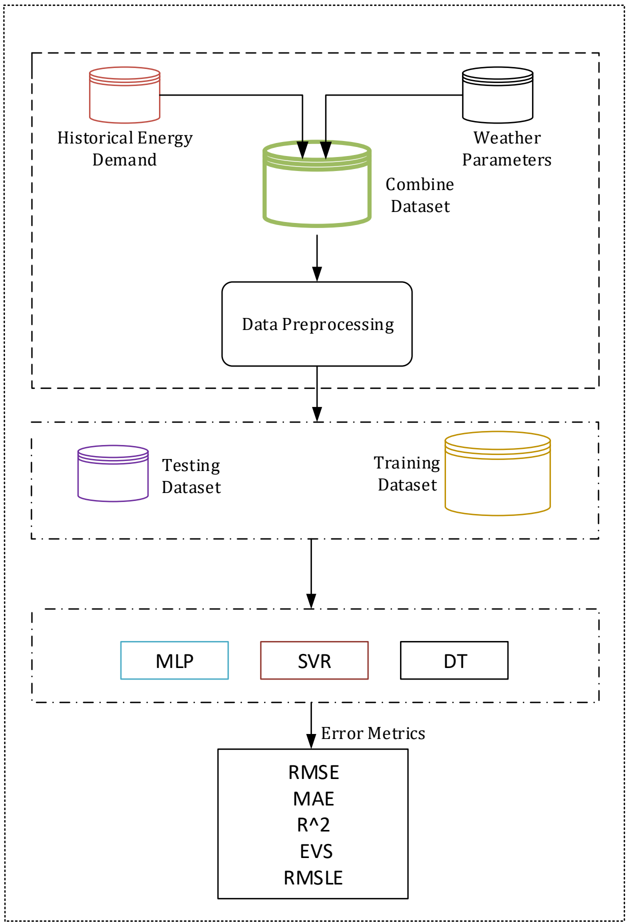

Figure 1: Dataflow Diagram of Study

Data collection: The dataset for the current study was collected from Ghana Grid Company (GRIDCo) (historical energy demand) and the University of Energy and Natural Resource (UENR) weather station and Ghana meteorological agency Sunyani (the weather parameters); from January 2015 to December 2017. The dataset includes the following features: air temperature and heat index (TEMP), barometric pressure (PB), humidity (H), precipitation (P), solar radiation (RD) and wind speed (WS) and direction and historical energy demand (LOAD).

Two variables were considered, weather parameters as the independent variables while energy demand as the dependent variable.

Date Pre-Processing: The performance of a predictive model dramatically depends on the cleanest of the data used. The combined dataset (DS) was treated in two phases, (i) data cleaning, which includes filling in missing values, smoothing noise, identification and removing of outliers where necessary and resolving data inconsistency. The average technique was used to compute a value for any missing value in DS. (ii) Data normalisation and aggregation, this was performed to scale all data in the same range of [0,1].

Training Techniques: A supervised machine learning process was used in this study, where the intended input variables (weather parameters) and output target (load demand) were entered into the network.

3.2 Predictive Models

Three machine learning algorithms, namely decision trees (DTs), neural network (NN) and support vector machine (SVM), were used in implementing the proposed predictive model. These algorithms were chosen for the current study based on their supremacy in the machine learning task similar to the current study, reported in [1], [5], [8], [16] and researcher familiarity with their concept and application. All adopted machine learning algorithms in the current study were implemented with Python on the Anaconda framework.

Decision Trees (DTs): DTs are non-parametric supervised machine learning algorithm cable for performing both classifications, and regression machine learning tasks. DTs are simple tree-like structure, with each leaf represents an outcome label. The ordinary CART algorithm was adopted in implementing the DT design in this study. The dataset was recursively divided throughout the training process in agreement to the split benchmark [19] until the optimum DT hierarchy of nodes was obtained. The Gini Index (GI) optimisation measure given in equation (3.1) was adopted in this study. Apure node is a node with only one observational class; a pure node is assumed to 0 GI. Other additional parameters set for the DT in the current study were a minimum of 4 observations for the node to develop into a full branch with at least one observation per leaf node.

$$GI\ (DS) = 1 - \sum_{n=1}^{m} Pn^2 - - - (3.1)$$

Where pn is the probability that a tuple in DS belongs to class C_n.

Artificial Neural Network (ANN): The multilayer perceptron (MLP) was adopted for the current study. The MLP is believed to be one of the most commonly used artificial neural networks. MLPs have at least three layers of artificial neurons (ANs). The input layer (IL) is made up of simple input neurons [20]. The input neurons are connected to at least one hidden layer of ANs. The hidden layer (HL) signifies latent variables (100 was used in the current study); the input and output of this layer cannot be perceived in the training data. As a final point, the last HL is connected to an output layer (OL). The logistic function is given in equation (3.2) was adopted as an activation function for the HL, while the Backpropagation training algorithm was used in training the MLP of this study. The maximum number of iterations was set to 5500 for this study. Figure 2 shows a simplified architecture of the MLP with two hidden layer and three input neurons.

$$f(x) = \frac{1}{1+e^{-x}} ------(3.2)$$

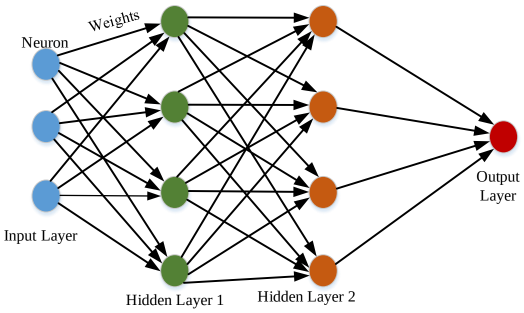

Figure 2: A simplified Architecture of MLP

Support Vector Machine (SVM): The SVM is a supervised machine learning algorithm for both classification and regression machine learning tasks. The SVM serves as the linear separator inserted between two data nodes to distinguish two unrelated classes in the multidimensional environs [20]. In this study, the Support vector regression (SVR) was adopted. The radial basis function (RBF) given in equation (3.3) was used in this study.

$$\text{RBF: } K(x_i,x_j) = exp\left(-y\|x_i - x_j\|^2\right), y > 0 ----(3.3)$$

$$\text{Where } K(x_i,x_j) \equiv \emptyset(x_i)^T \emptyset(x_j).$$

3.3 Evaluation Metrics

To measure the error rate of our proposed model, the following error metrics which include root mean squared error (RMSE), mean absolute error (MAE) and correlation coefficient (R) as defined in [12] were used as performance metrics measure.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}}(y_i - \hat{y}_i) ------(3.4)$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i) ------(3.5)$$

$$R = \frac{\sum_{i=1}^{n}(t_i - \bar{t})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(t_i - \bar{t})^2 . \sum_{i=1}^{n}(y_i - \bar{y})^2}} ----(3.6)$$

$$where, \bar{t} = \frac{1}{n}\sum_{i=1}^{n} t_i \text{ and } \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

Where (y_i) is the actual value and ($\hat{y}_i$) is the predicted value.

4. RESULTS AND DISCUSSION

This section presents the outcome of the current study and its implications.

4.1 Visualization of Processed Data

Figure 3 shows the weather parameters and consumption pattern for the year 2015. It was observed that the load demand in May 2015 was less as compared with that of January, February, March and April. However, a rise in load demand was seen in June but not as high as January to April. The reason for the low demand may be attributed to the load schedule activities that occurred in 2015, which became crushing in the month of May 2015.

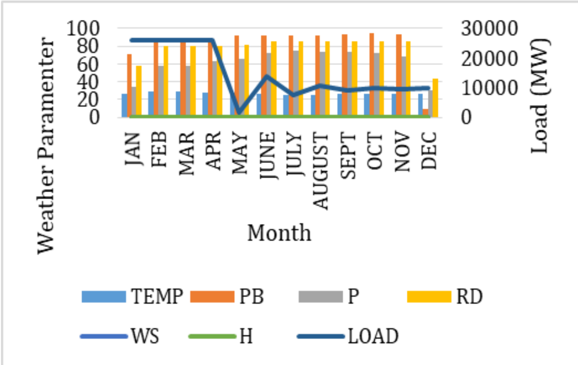

Figure 3: Load Demand in 2015

Figure 4 shows the weather parameters and consumption pattern for the year 2016. It was observed that the demand in July and December 2016 is higher as compared with the other months.

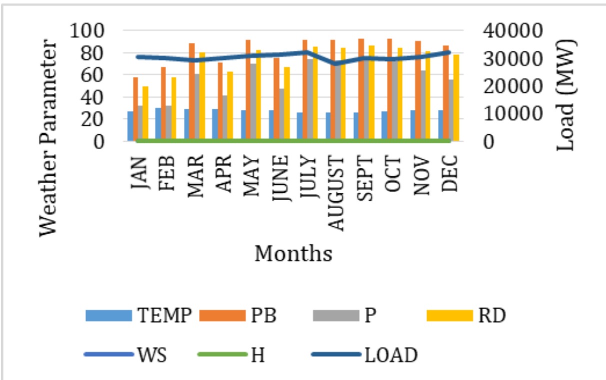

Figure 4: Load Demand in 2016

Figure 5 shows the weather parameters and consumption pattern for the year 2017. It was observed that the demand in 2017 was relatively stable from January to October, however a little drop in load demand in November, and a rise in load demand in December 2017.

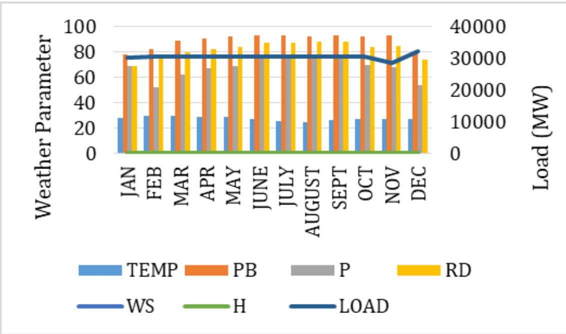

Figure 5: Load Demand in 2016

4.2 Descriptive Statistics

Table 2 shows the statistical analysis of the dataset. The value of the standard deviation in the monthly demand reveals a high rate of variation in demand from January 2015 to December 2017, confirming the difference seen in figures 3 - 5.

Table 2 Standard Deviation and Mean of Load Demand from 2015-2017

	Mean	Std. Deviation
LOAD	25186.660	9025.9490
MONTH	6.50	3.501
TEMP	27.350	1.3386
PB	85.050	15.7031
P	62.778	13.8107
RD	78.428	11.2395
WS	54.889	14.9031
H	81.506	69.6620

4.3 Experimental Setup

The preprocessed data was partition into two training dataset (January 2015 – June 2017) and testing dataset (July 2017 – December 2017).

A careful examination of the dataset reveals that the electricity demand is highly related to a specific month. Hence a 30 days look-back gives a better prediction of next month's load demand. The outcome of the experiment is discussed in the following sections.

4.3.1 Features Importance Ranking

Figure 6 shows the feature importance ranking of the independent variables. The results reveal that the month is of high importance (0.178505) as compared to the other six (6) features. Thus, the outcome reveals that the number of holidays in a particular month has a partial effect on the amount of electrical energy demand by consumers. The results confirm the increase in energy demand in the month of December 2016 and 2017, as observed in section 4.1. The temperature is next in the importance ranking with a value of (0.156331).

The results revealed that rainfall has a little effect on the amount of electrical energy demand by consumers.

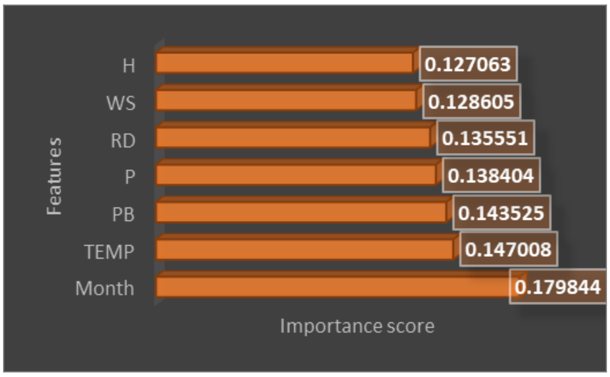

Figure 6: Features Importance Ranking

4.3.2 Model Prediction

Figure 7 shows the comparison between the actual values (y) the predicted values ($\hat{y}$) of MLP, DT and SVR models. The results reveal that the proposed predictive models obtained an average accuracy of 80.57% for DT, 95% for MLP and 67.2 for SVR. This implies that the proposed predictive model can provide sufficient prediction of future energy demand in the Bono Region with a high accuracy level, which can help the management of electricity supply in the Bono region and beyond to plan and make an informed decision.

Furthermore, the accuracy measure of MLP compared with the DT and SVR confirms the high rate of ANN in prediction studies in the literature.

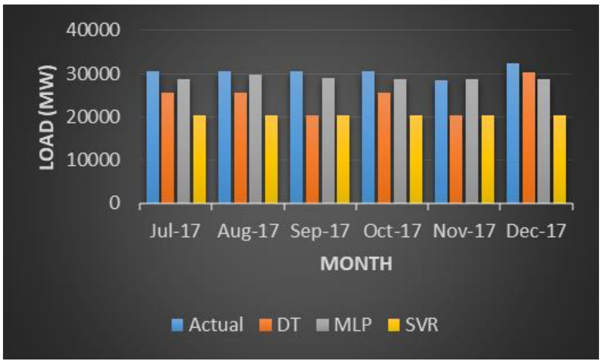

Figure 7: A comparison of Actual and Predicted Value Among Algorithms

4.3.3 Error Metrics

Table 3 shows the error metrics measure of the proposed predictive model. The table shows how close the predicted values are to the actual values; confirming the accuracy measure of our model.

Table 3: Error Measure of Models

Error Metrics	DT	MLP	SVR
RMSE	0.064221	0.021184	0.100776
MAE	0.058463	0.018565	0.100191
R^2	-34.0959	-2.81865	-85.422
EVS	-5.01108	-0.00318	0.000029
EedAE	0.049314	0.018031	0.10003
RMSLE	0.051009	0.016295	0.080301

5. CONCLUSIONS

Prediction of future power consumption plays an essential role in power conservation and efficient power use. In this study, a predictive model established on DT, MLP and SVR algorithms was proposed for predicting the 30-day head electrical energy demand in the Bono Region of Ghana. The model efficient integrated the weather parameters in discovering the most influential factors affecting electrical load demand and prediction. To demonstrate the trustworthiness of the proposed system, we have experimented the system with real-world data sets from January 2015 to December 2018.

The study contributes to the body of knowledge via an amalgamation of efficient three machine learning algorithm for predicting the future electrical load demand of the Bono region of Ghana and the critical weather parameter impacting demand in the Region. The accuracy and error metrics of the MLP was found to be better as compared with DT and SVR. Hence the MLP was used for the final model.

It was observed that the data size for the study was small, which we believed affect the performance of models in a way. Future, model's performance can be improved if the data size is increased and categorised. Again, additional features such as holidays, user lifestyle and customer categorisation can also be added to give a better prediction. Since the amount of energy one consumed is believed to depend on one's lifestyle, such as house size, the number of electrical equipment used and the kind of services received from the supply authority.

Acknowledgement

The authors of the current paper wish to express our appreciation to all staff of electrical and electronic engineering department of the Sunyani technical university. We again appreciate the effort made by the Ghana grid company (GRIDCo) and the University of energy and natural resource (UENR) weather station and Ghana meteorological agency Sunyani.

Conflicting interests

The authors declare that there are no impending conflicts of interest concerning the research, authorship, and publication of this article.

REFERENCES

[1] M. Zahid *et al.*, "Electricity Price and Load Forecasting using Enhanced Convolutional Neural Network and Enhanced Support Vector Regression in Smart Grids," *Electronics*, vol. 8, no. 2, p. 122, 2019.

[2] H. T. Akuribire, C. Puri, A. G. Abalena, and K. I. Nti, "Assessment of Natural and Artifical Lighting Levels in Lecture Rooms (A Case of Sunyani Technical University)," *Glob. Sci. Journals*, vol. 7, no. 6, pp. 149–153, 2019.

[3] E. N. Kumi, "The Electricity Situation in Ghana: Challenges and Opportunities," Washington, DC, 2017.

[4] K. R. Selvaraj, S. Sundararaj, and T. Ravi, "Artificial Neural Network Based Load Forecasting and Economic Dispatch with Particle Swarm Optimization.," vol. 4, no. 5, pp. 139–145, 2013.

[5] M. Hambali, A. Akinyemi, J. Oladunjoye, and N. Yusuf, "Electric Power Load Forecast Using Decision Tree Algorithms," *Comput. Inf. Syst. Dev. Informatics Allied Res. J.* vol. 7, no. 4, pp. 29–42, 2016.

[6] S. Zakarya, H. Abbas, and M. Belal, "Long-Term Deep Learning Load Forecasting Based on Social And Economic Factors in the Kuwait Region," *J. Theor. Appl. Inf. Technol.*, vol. 95, no. 7, pp. 1524–1535, 2017.

[7] A. Bogomolov, B. Lepri, R. Larcher, F. Antonelli, F. Pianesi, and A. Pentland, "Energy consumption prediction using people dynamics derived from cellular network data," *EPJ Data Sci.*, pp. 2–15, 2016.

[8] N. I. Nwulu and O. P. Agboola, "Modelling and Predicting Electricity Consumption Using Artificial Neural Networks," *IEEE*, pp. 1–5, 2012.

[9] H. Kuhba and H. A. H. Al-tamemi, "Power System Short-Term Load Forecasting Using Artificial Neural Networks," vol. 4, no. 2, pp. 78–87, 2016.

[10] E. Owusu-Adjapong, "Dumsor: Energy Crisis In Ghana," *Submitted as coursework for PH241, Stanford University, Winter 2018*, 2018. [Online]. Available: http://large.stanford.edu/courses/2018/ph241/owusu-adjapong1/.

[11] P. Stavast, "Prediction of Energy Consumption Using Historical Data and Twitter Master thesis," University of grningen, 2014.

[12] I. K. Nti and J. A. Quarcoo, "Self-motivation and Academic Performance In Computer Programming Language Using a Hybridised Machine Learning

Technique," *Int. J. Artif. Intell. Expert Syst.*, vol. 8, no. 2, pp. 12–30, 2019.

[13] J. A. Konica, E. Scheduling, and L. Hanelli, "Forecasting Next-Day the Electricity Demand Based On Fuzzy Logic Method Case for," vol. 3, no. 12, pp. 6172–6180, 2016.

[14] C. H. J. Kumar and M. Veerakumari, "Load Forecasting of Andhra Pradesh Grid using PSO, DE Algorithms," vol. 1, no. 9, 2012.

[15] L. Approaches, "Optimal Deep Learning LSTM Model for Electric Load Forecasting using Feature Selection and Genetic Algorithm : Comparison with Machine Learning Approaches †."

[16] S. Ghore and A. Goswami, "Short Term Load Forecasting Of Chhattisgarh Grid Using Artificial Neural Network," *Int. J. Eng. Dev. Res.*, vol. 3, no. 4, pp. 391–397, 2015.

[17] A. Soni and A. K. Sharma, "Electricity Load Forecast for Power System Planning," vol. 2, no. 9, pp. 52–57, 2013.

[18] D. Project, I. N. Mathematics, and S. Cycle, "Predicting Hourly Residential Energy Consumption using Random Forest and Support Vector Regression An Analysis of the Impact of Household Clustering on the Performance Accuracy," 2016.

[19] L. Breiman, J. H. Friedman, R. Olshen, and C. J. Stone, *Classification and Regression Tree Wadsworth & Brooks/Cole Advanced Books & Software*. Pacific California, 1984.

[20] G. Hackeling, *Mastering Machine Learning with scikit-learn*. 2014.

AUTHORS BIOGRAPHIES

Isaac Kofi Nti holds HND in Electrical & Electronic Engineering from Sunyani Technical University, Sunyani, Ghana BSc Computer Science from Catholic University College Sunyni-Fiapre, Ghana MSc Information Technology from Kwame Nkrumah University of Science and Technology Kumasi, Ghana. Mr Nti is a Lecturer at the Department of Computer Science, Electrical & Electronic Eng. Sunyani Technical University, Sunyani, Ghana and currently a PhD candidate in the Department of Computer Science and informatics, University of Energy and Natural Resources Sunyani, Ghana. His research interests include machine learning, artificial intelligence and data security.

Asafo-Adjei Samuel holds HND in Electrical & Electronic Engineering from Sunyani Technical University (Sunyani Polytechnic), BSc. In Computer Science from the Catholic University of Ghana. Mr **Asafo-Adjei** is a final year student at Sunyani Technical University pursuing a Bachelor of Technology in Electrical and electronic engineering

Agyemang Michael holds HND in Electrical & Electronic Engineering from Sunyani Technical University (Sunyani Polytechnic), BSc. In Computer Science from the Catholic University of Ghana. Mr **Agyemang** is a final year student at Sunyani Technical University pursuing a Bachelor of Technology in Electrical and Electronic Engineering.