RESEARCH ARTICLE OPEN ACCESS

Prediction of Predicting the Monthly Electricity Charges of Households

Adil Ahmad Gada[1], Er. Baljit Kaur[2]

[1]M. Tech Scholar, Department of Electrical Engineering, IETBhaddal,Punjab, India
[2]Assistant Professor,Department of Electrical Engineering, IETBhaddal, Punjab, India

***Abstract*:** Electricity is the lifeline of almost everything in this 21st century. Residential electricity consumption has seen an increase both locally and globally. Therefore, it has become a global concern of significant importance to promote electrical energy consumption reduction (energy conservation) within the household for a viable development of a nation in the case of resource limitations. The current study seeks to identify the social psychology (lifestyle) factors that significantly influence the residential electricity consumption, and predict future electricity consumption using an artificial neural network (ANN) based on lifestyle data collected from three hundred and fifty (350) households in the Sunyani Municipality. The performance metrics RMSE, MSE, MAPE, and MAE, were used to estimate the performance of the proposed model. The RMSE (0.000726) and MAE (0.000976) of the proposed model compared to (RMSE = 0.0657 and MAE = 0.05714) for Decision Trees (DT) and (RMSE = 0.08816 and MAE = 0.06911) for Support Vector Regression (SVR) shows a better fit of the proposed model. Furthermore, it was observed that the type of vehicle (saloon or sport utility vehicle) used by the head of a household was the most significant lifestyle feature in forecasting residential electricity consumption. Future studies would focus on developing a vigorous model using a combination of weather parameters and several socio-economic factors based on hybrid machine-learning algorithms to increase forecasting accuracy..

Keywords: Forecast, Bills, Consumption, Lasso, Monthly .

I. INTRODUCTION

Electricity, with the passage of time, has become one of the essential forms of energy to humankind in light of the explicit fact that almost everything depends on electricity [1] [2]. The quantity of electrical energy consumed has become one of the critical concerns of the electricity industry for strategic planning and expansion [3]. Hence, the local and global energy producers consider that the efficient use of electrical energy is a key aspect to be addressed as a priority. Consequently, energy efficiency is a crucial challenge for building sustainable societies. However, the world's primary energy consumption is estimated to increase by 1.6% yearly as a result of increasing incomes, growing populations and the industrialisation of developing countries [4]. This scenario raises issues associated with the increasing paucity of natural resources, the increase in environmental pollution, and the looming menace of global climate change. This warrants a call for the efficient management of electrical energy by both domestic and industrial consumers.

Studies show that 30% of the electricity generated globally is consumed by residential users [5]. In 2015, the global average yearly electricity consumption by residential users reached 10,812kWh. Again, a report in 2017 revealed that 25,019kWh/year of electricity was used in Norway, 11,974kWh/year in the US, 6,400kWh/year in France, 4,656.52kWh/year in the UK, 26kWh/year in Sierra Leone, and an average of 21 372TWh globally, which is 2.6% higher than in 2016 [6] [7]. Similarly, residential electricity consumption in Ghana for the past nine years kept increasing in value, thus 1,996GWh in 2007, 2,168GWh in 2008, 2,275GWh in 2009, 3,060GWh in 2013 to 3,932GWh in 2016 [8].

Notwithstanding, the increase in residential electricity consumption, Nishida et al. [9] suggest that residential (domestic) energy consumption differs depending on the lifestyle of the family. A family lifestyle, according to [10], is a set of factors such as family composition, house type, age, home appliances possessed and their usage, family income, cultural background, social life and lifestyle habits which include how long to stay at home and how to spend holidays. For this reason, it is difficult to grasp the factors that have significant influences on electricity consumption by a residential consumer.

Despite the difficulties associated with electrical load forecasting, knowing the expected residential load demand is desirable for electrical energy generators and distributors to make the right decision ahead and for consumers to know how future energy consumption (demand) will change in line with their future lifestyle [1] [11]. Further, it helps to accurately determine the right time for buying and selling electrical energy, which contributes to costs savings or even earning an income. Moreover, Kong et al. [12] argue that as the power system is facing an evolution toward a new flexible, intelligent, and collaborating system with sophisticated infiltration of renewable energy generation, a short-term electric load forecasting for residential (individual) electricity customers, plays a progressively more essential role in the future grid planning and operation [12].

In [2], the authors argue that the basic unit of electrical energy consumption is the home. Hence, the reduction in residential electricity consumption will reduce the consumption of electrical energy in society. This raises the need to find the features/factors that have significant effects on household electricity consumption, and to address such features/factors to minimize electricity consumption. Simultaneously, the consumers could also be informed of their energy consumption pattern in the future, so that they could plan in time and make an appropriate decision. Therefore, in affirmation to [13] opinion, energy management systems (EMS) are required to monitor the generation, distribution, consumption, and storage and also to make the paramount decisions according to input signals and the user's requirements and preferences.

Based on the issues detailed above, the current study seeks to propose a predictive model for useful and accurate prediction of Now a days, consumption of electricity is increasing along with the financial growth and it is also essential in our day to day life [1], [3], [4]. In recent years, the participants confronted many challenges in the electricity market with the concept of deregulation in the power industry.

The majority of the methodologies for prediction of the consumption of the electrical energy are categorized into two, i.e conventional statistical methods and ML methods [5].

In this Paper, basically centers around on Neural Networks (NN), Linear Regression (LR), Support Vector Regression (SVR), K-Nearest Neighbors (KNN), Random Forests (RF), Gaussian Process Regression (GPR) to predict the electricity consumption [2]. In addition to these a recent model such as an adaptive network-based fuzzy inference system (ANFIS) and Gray relational analysis (GRA) utilized to get a calculation of the electricity consumption of a building based on human exercises and climate conditions [1].

II. LITERATURE REVIEW

Hung Nguyen et. al. (2017) [3] introduced ARIMA and SARIMA models for short term load forecasting using historical data of electricity load for past 14 years. The objective of this analysis was to separate the time series random components from deterministic components like trend, seasonality and cyclicality of the dataset and apply the ARMA processes to construct the models. In the study, SARIMA model provide a better fit to time series data as it allows randomness in the seasonal pattern. For accuracy measurement Mean Absolute Percentage Error (MAPE), Mean Absolute Deviation (MAD) and Mean Squared Deviation (MSD) statistics were calculated for different length of input data and recorded as 9.13% of MAPE for ARIMA and 4.36% MAPE for SARIMA model.

A comparative study of load forecasting approaches using three different methods for occupancy prediction and context-driven control of a smart building's appliances was reported by **S. Hadri et.al. (2019) [4]**. The proposed methods were ARIMA and SARIMA, which were based on statistical method, while XGBoost and Random Forest (RF) are based on machine learning method and LSTM is based on deep learning method. This study also investigated the performance of the prediction results of the five methods by evaluating the error metrics. It was reported that the XGBoost model gives better performance in terms of execution time and also better accuracy for short term load forecasting and occupancy prediction

III. OBJECTIVES

1.Calculate precision of four ML models for dairy ranch E&W forecast from a unique dataset of indicator factors using a scope of data mining techniques. These data mining techniques included: variable choice techniques to remove high prescient yield factors, grid search hyper-parameter tuning to improve the forecast exhibition of each ML model, and defined settled cross-approval compute the expectation execution properly using information not utilized for model preparing or approval.

2. Dissect the month to month forecast inclination of every ML algorithm to decide factors, which may impact model execution. Like to [12] the presentation of every ML algorithm was benchmarked against that of the MLR model. Forecast precision and inclination of the ML models were benchmarked against results from recently created MLR models utilizing basic model approval rules utilized in horticultural research [9], [13]. These recently created MLR models for dairy ranch E&W utilization were created utilizing similar information utilized for this investigation and determined model correctness's utilizing concealed information.

3. Analyse the absolute prediction accuracy for the most accurate ML model for both E&W (selected from section one) according to the number of dairy cows.

IV. METHODOLOGY

.This section discusses the pattern of electrical energy consumption globally and locally, the social-psychological factors the affect residential electrical energy consumption, methods of modelling energy consumption, machine learning, and its application in energy modelling and finally related studies on electricity prediction.

A. ELECTRICITY CONSUMPTION LOCALLY AND GLOBALLY

This figure has risen over the years that in 2016 it became 87% and 88.87% in 2017. The primary reason for this drastic increase is the constantly growing urbanization [14] - [16].

Electrical energy consumption may be classified as residential (domestic), commercial (non-residential) or industrial. Residential or domestic refers to the home or a dwelling where people globally live from day-to-day. Also, domestic electricity consumption happens on a considerably small scale compared to commercial users (businesses) who deal with heavy-duty machinery and lighting or appliances [17]. Fig.1 shows the electricity consumption per capita (kWh per person) in 2017. The individual consumption (IC) of electricity in every country using Eq.(1) shows that there is a higher disparity in electricity consumption among developed countries.

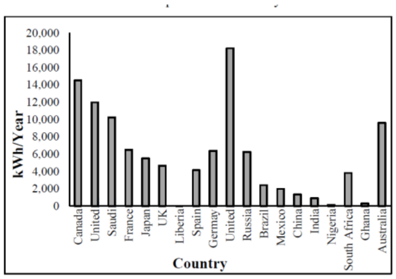

Figure 1Electricity consumption per capita [6]

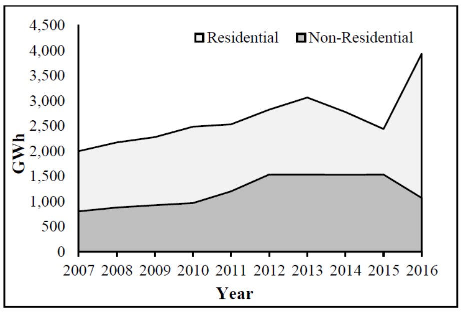

Figure 2 Residential and commercial electricity demand in Ghana

The situation in Ghana is of no difference; a report by the Energy Commission of Ghana shows that residential electricity demand has increased over the past nine-year as compared with

commercial/non-residential demand for electricity (Fig.2). The average electricity consumption in 2016 in Ghana was 2,647GWh for residential and 1,196GWh for commercial [8].

B. SOCIAL PSYCHOLOGICAL FACTORS INFLUENCING RESIDENTIAL ELECTRICITY CONSUMPTION BEHAVIOUR

As per [2], the home is the basic unit of electrical energy consumption, and hence, there is a need to study and identify the factors that contribute to energy consumption in the home. The literature argues that diverse families have diverse structures, ideological concepts, and cultural backgrounds. Every family is anticipated to have different electricity load profiles based on the influence of categories of factors and their interactions. Subsequently, a different load profile also mirrors diverse family types and consumption patterns [2] [5] [9] [13] [18]. The section briefs some of the household characteristics that are reported to influence domestic electricity consumption.

The Family Size (FS), study reveals that there is a direct positive relationship between the FS and electricity consumption (EC), thus FS □ EC. The family size includes the nuclear family (children and parents), extended family, and domestic staff in the house [2] [9].

Size of the House (SH): the SH (the type of house) and the number of rooms in a house is also believed to partially affect the electricity consumption pattern in a home [2] [9].

The Age Composition of Family Members (ACFM), the age of the family member significantly influences the electricity consumption in a home [9] in their studies. Reports showed that electricity consumption in a home is minimal (lesser) when the age of the family members is below fifty (50) years and above sixty-five (65) and higher when the age of the family members is within fifty to sixty-five (50-65) years [2].

Family Economic Situation (FES), the family FES thus the disposable family income, and the family income is believed to influence household electricity demand [2] [9].

Time Staying in the House (TSH), the daily average time spent in the house, is reported to influence the electricity consumption in the home [9].

Educational Level (EL): the EL of the family member is believed to influence the consumption of electricity home Gram-Hanssen et al. (2005) cited in [2]. However, some researchers disagree with its effects on electricity consumption [19].

Time for Watching TV (TWTV): the time spent on watching television is argued to impact electricity consumption in the house [9].

Another factor that is believed to affect the consumption of electricity in a home is the social status of the family (SSF). However, social status is seen to have a different influence on electricity consumption. According to [20], a positive correlation exists between the socio-economic status of a family and household electricity consumption. On the other hand, [21] argues that there is no correlation between the family members' economic status and household electricity demand.

Additional influences such as the location of the family house, the type of car used by the family, the occupation of the family members, and the gender composition of the family are believed to influence the consumption pattern of households.

.Table 1 Household distribution of the currrent study

Community	**No. of household**	**Percentage (%)**
		(n = 350)
Area 1	33	9.43%
Area 2	67	19.14%
Area 3	39	11.14%
Area 4	55	15.71%
Area 5	65	18.57%
Area 6	68	19.43%
Area 7	23	6.57%

C. Machine Learning Models for Prediction

Support Vector Regression (SVR)

Support Vector Machine is considered as a regression method. It upholds all important features, characterizing the algorithm with highest margin. The Support Vector Regression Method uses the principles as of the SVM for classification, with little differences. The outcome is a real number. So it is difficult to predict the information with countless possibilities. In regression, tolerance or epsilon is put in an approximation to the SVM, which have been asked for the problem. A more complicated reason too exists. The algorithm is more obscured to be considered. The main objectivity is to diminish the error by individualizing the hyper plane. The hyper plane exceeds the margin by tolerating the part of error.

1. Random Forest (RF)

Another Learning methodology is known as Random Forest i.e RF. Random Forest gets operated by multiple numbers of Decision Trees. The ultimate conclusion is made considering the maximum of the trees. Here it uses the random forest method. The advantage is, it diminishes the threat of over fitting and the training time. Also, it offers a higher level of precision. Random Forest Algorithm runs efficiently and effectively in larger databases and yields highly accurate predictions by estimating the missing data. Random Forest i.e RF is an appreciable algorithm to train before time in the model development process. Its objective is quite sure to check performance measure. building might be treated "bad" by considering its simple approach and it becomes proposed as a bad random forest.

2. Category of Models Methodology Objective

The algorithm is better to be used for user, who wants its application. It also helps those, who wishes to develop a quick model. Random forests are very hard to beat performance. A model can be used to provide better performance, like a Artificial Neural Network. The only drawback is more time consumption to develop. The only thing to consider is different feature types, numeric, binary, categorical. In a nutshell, Random Forest is considered as the fast yet simple as well as flexible tool. The best part of it is, it's not restricted with any limitations.

3. Linear regression (LR)

Linear Regression [2], is a model, which finds the relationship between the response variable or energy consumption and the return or other variables. The objective of regression analysis forecasts the demand for energy from one or more independent variables. Linear regression is a method used, when trend in historical forecast data is obvious. For this reason, the application has been used for electricity consumption forecast.

4. K-Nearest Neighbors (KNN)

The K-Nearest Neighbors algorithm is a simple, supervised Machine Learning algorithm. It solves both classification and regression problems. It's easy for understanding and implementation. It has a major drawback i.e it becomes slows as the data size grows The KNN algorithm assumes, similar things exist in close proximity. This means to imply that, similar things are near to each other. KNN works by finding the distances between a query and all the examples in the data, selecting the specified number examples closest to the query, then opts for the most frequent label in the case of classification or averages the labels in regression.

5. Artificial Neural Network (ANN)

The Artificial Neural Network (ANN) is also known as neural network. It is a Machine Learning Method, evolved from the idea of simulating the human brain [38]. An ANN simulates the network of neurons, making up a human brain, so that the computer will be able to learn things and make decisions in a humanlike manner. ANNs are created by programming the Computer Systems to behave as though they are interconnected brain cells. There are several ways, ANN can be deployed to classify information, predict the outcomes and cluster data. As the networks process and learn from data, they can classify a given data set into a predefined class. It can be trained to predict outputs, from a given input and can identify a special feature of data to then classify the data by that special feature. Computers understand the world around them in a human like manner using ANN.

6. K-Means Clustering

Clustering belongs to unsupervised learning approach, where test dataset are not labeled. Hierarchical Clustering depends on building hierarchy, using two types of a clustering techniques. They are Agglomerative and Divisive. Agglomerative Clustering is paired to create big cluster in bottom up approach. The Divisive Clustering breaks a big cluster into the small clusters in a top down approach. Partitioning Clustering is a technique, which use partitioning the datasets into equal or unequal sets. Each of the set is characterized in cluster form. In K-Means Clustering, Dataset is into set of K-small clusters. Each of the cluster is represented through cluster mean [39], [40].

7. G. Extreme Learning Machine (ELM)

Extreme Learning Machine (ELM) is a new learning method, proposed by Huang et al. [14], [47]. One major drawback was caused by gradient descent based algorithms and Back Propagation. ELM overcomes out of it. ELM is based on Single hidden Layer Feed Forward Neural Network architecture. Three different layers, input layer, hidden layer and output layer are there [48]. The hidden bias and the weight for connecting input layer and hidden layer are randomly get and maintained through the whole training process. Extreme Learning Machine have no parameters to adjust the hidden neurons. These can easily be applied in regression [46] or classification [45] issues. It yields low computational cost during the process of training. ELM comes into existence for its application in time series prediction. Time sries prediction is used for predicting the sales in fashion retailing [44]. ELM is used for electricity price forecasting with fast computational ability [43]. ELM is also applications with wind power density prediction [42]. It is compared with SVM and ANN. In [41], ELM has been applied for daily dew point temperature prediction

V. SYSTEM ARCHITECTURE

The Table.2 shows the initially obtained features out of the administered questionnaires and how they were abbreviated. The qualitative response from respondents was first coded using dummy variables. Afterwards, all the 350 received responses were queued with Microsoft Excel into a comma-separated values (CSV) file format. The implementation of the proposed model was carried in Python using the Anaconda framework and Scikit-learn libraries.

Table 2 Initially selected features

Features	Abbreviation
Residential location	RL
Age of family head	AFH
Employment sector of the family head	ESFH
Type of employment of family head	TEFH
Nature of Employment	NE
Apartment Type	APT
Marital status	MS
Living with family	LWF
Family size	FS
Family male-dominated or female-dominated	FMFD
The monthly salary of the family head	MSFH
Own personal vehicle	OPV
Number of vehicles Owen	NVO
Vehicle type	VT
Electricity consumption per month	ECPM

A. RESEARCH DESIGN

The cross-industry process for data mining (CRISP-DM) modelling process [30] was adopted for this study (Fig.4). The CRISP-DM method offers an organised approach for planning data mining and forecasting studies. Its iterative process offers a continuously learning and improves communication of insight and forecasting power. Most of the tasks involved in this model can be executed in a different order, and it will often be necessary to do a volte-face to previous tasks and echo certain activities. The CRISP-DM model has six phases, namely: study objectives, data understanding, data preparation, modelling, evaluation, deployment/publish results. The first phase study objective covers ascertaining the aim and objectives, which, when well-understood, leads to the gathering of the correct data.

Phase two covers visualisation of the research data, to give a better understanding, and identification of discrepancies, and deviation in the dataset. In the third phase, the obtained dataset is pre-processed by identifying missing values, if any, and treating them appropriately by scaling or by normalisation of the dataset for better/accurate prediction. The dataset for the current study was scaled in the range of [0, 1] and by dividing every value in the dataset by the maximum value to get the new value. The feature extraction and selection techniques are applied in this same phase to the select the significant feature. In this study, the root means square error (RMSE) metric is used to select the significant feature. The next phase is the modelling phase, where the selected features serve as input (independent) variables to our MLP model to predict the expected (dependent variable) future residential electricity demand. The predicted values are then compared with the actual values (y) to evaluate the performance of the model in the evaluation phase. The error metrics used in this study are discussed in detail in section 3.4 of this paper. Lastly, the deployment or publish results phase; at the stage, the results and findings obtained from the empirical study are communicated to the scientific community.

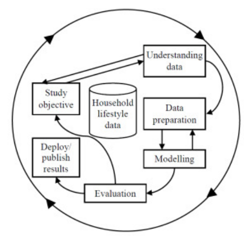

Figure 3 Flow chart of the system

Different algorithm was adopted for the current study based on its high accuracy levels in energy prediction, as discussed in section 2.5 of this paper. A supervised ML technique was used in this study, where household lifestyle served as input parameters, while actual monthly electricity consumption of participants from the supply authority served as the output target in the proposed network. Implementation was done with Python on the Anaconda framework.

Data mining and classification problems are usually expedited with optimization problem "feature selection" for choosing the essential features from various input sets while producing the same output. Although several algorithms have been developed, on the other hand, none is universally accepted to be the finest for all conditions. Therefore researchers are still trying to come up with improved solutions [31].

Artificial Neural Networks (ANNs) are a computer-based copy of the human brain composed of many "neurons" that work together to accomplish the desired purpose. They can be used for pattern recognition, classification, image matching, feature extraction, prediction, and noise reduction.

ANNs can learn and generalise, as mentioned earlier [32]. The proposed model comprises of two ANN multi-layer perceptron model. Thus, backpropagation trained ANN (BPANN_1) for feature selection and (BPANN_2) for prediction (Fig.5). The current study adopted [32] ANN algorithm for feature selection. The approach aimed at varying the weights of the different features to minimise the error in the actual value (y) and the predicted value .

So that the feature that possess lower weight was considered not important hence, rejected y

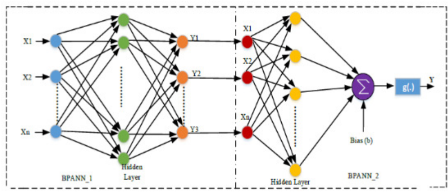

Figure 4 Proposed ANN Framework

For BPANN_1, let *N* be the number of observations *Nneuron*and *Kx* the number of features. Two classes produced (-1/+1). Where *X* is the matrix, and the rows denote the features and, the columns signify the observations, and *y* indicates the class of an observation. All features were given equal weights at the start. For each feature (*Xx*) is associated with a weight *δk*.

Let *Dδ*represent a diagonal matrix representing the weights of the features. corresponding feature Xk is of impotence. While if δk is low, then feature Xk is of no importance. A feature was considered extra

important when its weight is more significant. The aim here was

to ensure that the weight of the unimportance feature reduces nearly to zero thereby minimising the errors (E). E is computed as Finally, features with corresponding significance weight more than 0.5 were picked, while features with corresponding weights less than 0.5 were treated as unimportant features and were

rejected.

For BPANN_2, let the output of BPANN_1 be presented by dataset (DS) comprises (x1,y1), (x2,y2),…,(xn,yn) where xi ∈ Rn.

With one hidden layer and fifty hidden neuron, proposed BPANN_2 learns function f(x) is given in Eq.(8). The maximum iteration was set to 5000, optimizer = Limited-memory BFGS (lbfgs).

where W1 ∈ m□ and W1, b1, b2 ∈□ are parameters for BPANN_2,

and W1, W2 represents the input and hidden layer weights, respectively; b1, b2 represents biased at the hidden and output layer, respectively. The activation function (g(.):R→R) adopted for BPANN_2 in the current study was the hyperbolic tan is in Eq.(9). The square Error loss function (Eq.(10)) was adopted.

With an initial random weight, the loss function was optimised by repeatedly updating the weights. We compute the loss and propagates backwards pass from output to the previous layer, providing individual weight parameter with a modernise value meant to reduce the loss.

B. EVALUATION METRICS

The following error metrics discussed by [1] are used in measuring the performance of the proposed forecast model:

Root Means Squared Error (RMSE): This index estimates the residual between the actual value and predicted value. A model has better performance if it has a smaller RMSE. An RMSE equal to zero represents a perfect fit.

Mean Absolute Percentage Error (MAPE): This index indicates an average of the absolute percentage errors; the lower the MAPE, the better.

The Correlation Coefficient (R): This criterion reveals the strength of relationships between actual values and predicted values. The correlation coefficient has a range from 0 to 1, where higher R means it has an excellent performance measure. are the average values of tv and yv, respectively and tv is the actual value, yv is the predicted value produced by the model, and m is the total number of observations

VI. SIMULATION AND RESULTS

A. Predicting household electricity bills

This notebook was made as part of the course end project for a masters program in Big Data and Business Analytics.

The dataset contains information about houses and its residents and the amount paid for electricity in a month, which will be the target variable of the machine learning model

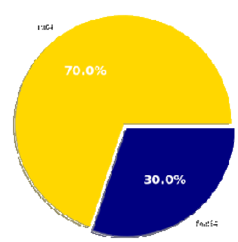

Figure 5 Composition of data types

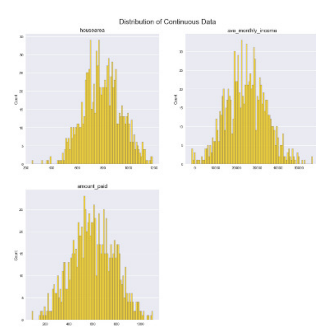

Figure 6 Distribution of continuous data

It is not logical for a house to have -1 rooms or people, so I will change those values to 0. Also, some values in ave_monthly_income are negative, these will be changed to the mean of the column.

In [12]:

df.loc[df['num_rooms'] <= 0, 'num_rooms'] = round(df['num_rooms'].mean())

df.loc[df['num_people'] <= 0, 'num_people'] = round(df['num_people'].mean())

df.loc[df['ave_monthly_income'] < 0, 'ave_monthly_income'] = df['ave_monthly_income'].mean()

Next, I will see if the dataset contains any duplicated rows or null values. The second one will be achieved by text and by visual representation showing white rectangles for null values.

In [13]:

#Check the duplicated rows

print('The dataset contains {} duplicated rows.'.format(df.duplicated().sum())).

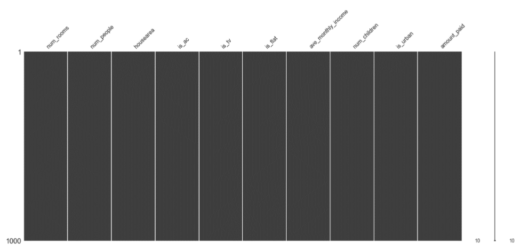

Figure 7 Correlation

Finally, show the correlation between each column

1. Linear Regression

In [20]:

#Ordinary least squares Linear Regression.

#Create the estimator

model_LR = LinearRegression()

#Train the model

model_LR.fit(x_train,y_train)

#Create predictions based on the test set

y_pred = model_LR.predict(x_test)

#Evaluate the predictions with the coefficient of determination (R^2)

results_linear = r2_score(y_test, y_pred)

#Store the scores

scores = scores.append({'Random Search' : results_linear, 'Grid Search' : results_linear},

ignore_index = True

2. Lasso

In [22]:

#Linear Model trained with L1 prior as regularizer (aka the Lasso)

#Create the estimator

model_La = Lasso(random_state = 9)

#Create the grid search inputing the estimator, parameters and cross validation value

grid_La = GridSearchCV(model_La,

{'alpha' : [.1, .25, .5, .75, 1]},

cv = 10)

#Train the models

grid_La.fit(x, y)

	param_alpha	mean_test_score
2	0.5	0.875228
1	0.25	0.875199
3	0.75	0.875164
0	0.1	0.875133
4	1	0.875010

3. Decision Tree Regressor

Starting from this model, I will start testing more parameters. I will now use a random and a grid search.

4. Random Forest

	param_n_estimators	param_max_depth	param_min_weight_fraction_leaf	param_max_features	param_max_leaf_nodes	param_bootstrap	mean_test_score
58	200	100	0.1	auto	30	True	0.710555

This time there is only one set of parameters with the highest score, the grid search will be around those values and bootstrap and max_features will be set to default values

B. Selecting Model

Now that I have tested the models, I will compare their scores to select the best ones. I will do so by visualizing the scores with a bar plot.

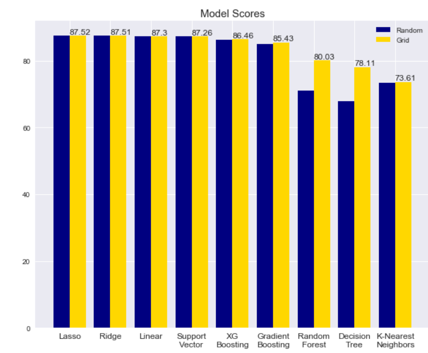

Figure 8 Model scores

The best 3 models here are Lasso, Ridge and Support Vector Regression. I will retrain these models by using the parameters that yielded the best results.

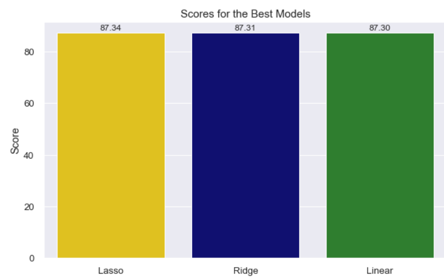

Figure 9 Scoresof the best models

Once again the best model for our data is Lasso

	Training data	a Testing data	b February	March	April	May	June	
Linear Regression	28.11	22.75	36.22	35.28	44.54	40.88	38.41	
Gradiant Boosting	26.1	18.89	38.89	34.06	42.20	35.20	33.51	
Random forest	7.00	25.09	6.63	51.32	67.59	58.94	42.82	
Lasso	14.88	13.80	16.67	30.66	45.99	41.87	55.31	
SVR	9.42	10.65	11.58	26.77	37.81	33.41	34.98	

VII. CONCLUSION

Electricity is an essential commodity in everyone's life and a conduit for economic development in every country. The growth rate of electricity demand in residential facilities keeps increasing over the years. However, several studies based on electricity forecasting overlooks this area, and the few studies that were focused in this area based their prediction on weather parameters as independent variables to their forecast models. On the other hand, residential electricity consumption is mainly based on household lifestyle. Therefore, the current study sought to forecast residential electrical energy consumption based on household lifestyle using an artificial neural network.

The experimental setup with 350 household lifestyle data randomly selected towns in the Sunyani Municipality in the Bono region of Ghana revealed that the type of vehicle used by heads of households was the most significant lifestyle feature in residential electricity load forecasting. Furthermore, the proposed model's accuracy (94.3%) as compared to the traditional machine learning such as SVR and DT shows that the proposed model can effectively forecast residential electricity consumption at an error rate of 5.7%.

Also, the use of actual electricity consumption (kWh) as the target variable compared to the amount (Ghana cedis) spent on electricity monthly by a household, makes the outcome of this study independent of the country's socio-economic factors such as gross domestic product, exchange rate, and inflation. Since the actual units (kWh) consumed by a household dependent on the household energy management lifestyle and not the country's socio-economic factors. In summary, the current study contributions are as follows: (i) An ANN-based feature-selection technique for optimal selection of the most significant independent variables. (ii) An empirical comparison of the proposed ANN model compared with other state-of-the-art forecast models. (iii) We add to the scarcity in the literature on residential load forecasting based on household lifestyle. (iv) To the best of our knowledge, this study is the first load forecasting of residential electricity consumption based on household lifestyle in India.

However, the accuracy of the forecast model is one of the evaluation metrics for a good forecast model. Therefore, despite the achieved accuracy measured in this study, we believe that there is more room for improvement. Hence, in future, we look forwarded to add diverse independent parameters such as weather data and more household lifestyle data (such as the number of electrical appliances and the average time of use) using hybridization of several state-of-the-art machine learning algorithms to optimize feature selection of parameters for better accuracy measure

REFERENCES

[1] I.K. Nti, S. Asafo Adjei and M. Agyemang, "Predicting Monthly Electricity Demand using Soft-Computing Technique", *International Research Journal of Engineering and Technology*, Vol. 6, No. 6, pp. 1967-1973, 2019.

[2] Z. Guo, K. Zhou, C. Zhang, X. Lu, W. Chen and S. Yang, "Residential Electricity Consumption Behavior: Influencing Factors, Related Theories and Intervention Strategies",

Renewable and Sustainable Energy Reviews, Vol. 81, No. 1, pp. 399-412, 2018.

[3] N.I. Nwulu and O.P. Agboola, "Modelling and Predicting Electricity Consumption using Artificial Neural Networks", *Proceedings of 11th International Conference on Environment and Electrical Engineering*, pp. 1059-1063, 2012. ISSN: 2229-6956

[4] 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 Science*, Vol. 5, No. 3, pp. 1-15, 2016.

[5] L.G. Swan and V.I. Ugursal, "Modeling of End-Use Energy Consumption in the Residential Sector: A Review of Modeling Techniques", *Renewable and Sustainable Energy Reviews*, Vol. 13, No. 8, pp. 1819-1835, 2009.

[6] Index mundi, "Electricity Consumption Per Capita-World", Available at: https://www.indexmundi.com/map/?v=81000, Accessed on 2019.

[7] IEA, "IEA's Electricity Information", Available at: https://www.iea.org/statistics/electricity/#targetText=In 2017%2C world total final, increase of 4.6%25 from 2016, Accessed on 2019.

[8] Ghana Energy Commission, "National Energy Statistics Report 2017", Available at: http://energycom.gov.gh/files/ENERGY_STATISTICS_2017.pdf, Accessed on 2017.

[9] K. Nishida, A. Takeda, S. Iwata, M. Kiho, I. Nakayama, "Household Energy Consumption Prediction by Feature Selection of Lifestyle Data", *Proceedings of International Conference on Smart Grid Communications*, pp. 235-240, 2017.

[10] A. Veit, C. Goebel, R. Tidke, C. Doblander and H. Jacobsen, "Household Electricity Demand Forecasting-Benchmarking State-of-the-Art Methods", *Proceedings of International Conference on Future Energy Systems*, pp. 1-10, 2014.

[11] P. Huber, M. Gerber, A. Rumsch and A. Paice, "Prediction of Domestic Appliances usage based on Electrical Consumption", *Energy Informatics*, Vol. 1, pp. 265-428, 2018