

Analyzing and Forecasting of Electricity Consumption by Integration of Autoregressive Integrated Moving Average Model with Neural Network on Smart Meter Data

M. Suresh^a, Anbarasi. M.S^b

^aAssistant Professor, Department of Information Technology, Manakula Vinayagar Institute of Technology, Pondicherry, India.

^bAssistant Professor, Department of Information Technology, Pondicherry Engineering College, Pondicherry, India.

Article History: Received: 10 January 2021; Revised: 12 February 2021; Accepted: 27 March 2021; Published online: 10 May 2021

Abstract: Smart metering is a recently developed research area over the globe and it appears to be a remedy for increasing prices of electricity. Electricity consumption forecasting is an essential process in offering intelligence to smart grids. Rapid and precise forecasting allows a utility provider to plan the resources and also to take control actions to balance the electricity supply and demand. The customers will advantage from the metering solutions by a greater understanding of their own energy utilization and forthcoming projections, allowing them to effectively manage the cost of their consumption. In this view, this paper presents an Integration of Autoregressive Integrated Moving Average (ARIMA) Model with Neural Network (NN) for Electricity Consumption Forecasting using Smart Meter Data. As the time series data often does not hold linear as well as nonlinear patterns, ARIMA or NN models are not enough to model and predict the time series data. The ARIMA-NN model will be trained using the data and generates a model. Afterward, the generated model can be utilized to predict the electricity consumption by the application of new building data. The proposed ARIMA-NN model is evaluated and the simulation outcome strongly pointed out its superior performance over the compared methods. The presented model has obtained effective testing performance with the MAPE of 25.53, an accuracy of 48.38, and MSE of 0.21.

Keywords: Smart meter, ARIMA, NN, Load prediction, Time series

1. Introduction

At present days, significant attention has been paid to smart electricity networks, where improved control over electricity supply and utilization is going to be attained due to the development of Advanced Metering Infrastructure (AMI). Smart metering becomes more familiar and it serves as an essential process to achieve defined policies, reduces the greenhouse gas emission, improves the energy efficiency, and ensures the energy required by renewable energy sources. Smart metering models become a portion of micro-grid that comprises diverse operational and energy metrics like smart appliances, renewable energy resources, and energy-efficient resources. An important issue associated with the operation of micro-grids is the optimum energy management of housing buildings based on multiple and overlapping objectives [1]. Presently, smart grid vision and smart home concepts are commonly employed for optimizing the utilization of energy and reduces electricity bills.

The development of the smart home energy management model becomes a wide global priority for supporting the trend towards a more sustainable and reliable energy supply for smart grids [2–5]. Initially, a new metering infrastructure is defined for ensuring the automatic reading and billing depending upon the present utilization. Next to that, through the accumulation of high-frequency utilization data, the system should fulfil the needs for the design of cost-reflective prices, varying concerning the time of utilization. Afterward, these metering models aim to the contribution to the minimization of the total energy utilization by raising the energy awareness of the users.

An essential aspect of smart metering models offers encouragement to the users for the utilization of lower electricity by being effectively informed regarding the utilization pattern. The prediction offers the users with the likelihood of connecting the present utilization behaviors with future costs. So, the user gets beneficiary from the prediction solutions by thoroughly understanding the individual energy utilization and their expectations, enabling them to effectively manage the cost of utilization. Through the transparent usage of energy and future projections, it is easier to interpret how much it affects economics in the upcoming days. Numerous models have been devised for electricity load forecasting. Few of the familiar time series analysis with the autoregressive integrated moving average (ARIMA) model, fuzzy logic, neuro-fuzzy model, artificial neural networks (ANNs), and support vector machines (SVMs).

Few fascinating instances of electricity forecasting on the individual household level are provided here. [6] intended to compute which machine learning (ML) model offered optimal prediction results for the entire building energy utilization for the subsequent hour. The outcome exhibited that the LS-SVM provides the optimal outcome in the prediction of the home's future electrical utilization. Generally, the presented model offered MAPE of 1.6-13.41% for a 700kWh commercial building and around 15-32% for 3 homes with mean expenditure closer to 1.5 kWh. In [7], high-resolution data from a set of three private households gathered for a month has been examined. The experimental outcome indicated that the recent prediction models offer better forecasting. Depending upon the disaggregated data from smart homes, sensors with perseverance and smart meter benchmarks disclose-considerable estimate enhancements of 4-33% to mean absolute error.

In [8], diverse models have been employed for predicting peak demand in separate homes. It is revealed that the home's historic peak load and tenancy is an effective predictor of peak load compared to season. It is represented that Seasonal Auto-Regressive Moving Average (SARMA) can be employed for modeling the intrinsic load pattern and user movement in a home and that it has a 30% lower mean square error compared to regression-based models. In [9], Kalman filter-based prediction technique leads to load forecasting with the MAPE of 30% for a sampling period and predicts a horizon identical to an hour. Small-time duration among the reception of real-time measurement data from customers' smart meter enhanced the accurateness of the presented model and results in the MAPE of almost 13%. In [10], depending upon the power expenditure readings from 23 homes gathered over Japan, a support vector regression technique and an activity sequence-driven model are developed to infer upcoming actions and improves load prediction. The activity sequence variable turns out to be a significant characteristic that can enhance the accuracy of load prediction by 15 minutes in advance for individual houses, reaches up to 42% MAPE on average.

In [11], several prediction models have been developed with the predicting horizons ranges from 15 minutes to 24 hours. The validation of the model takes place on two datasets namely one household in Germany and 6 households in the United States. The outcome pointed out that the predictive accuracy is mainly based on the selection of prediction model and parameter setting. In [12], an electricity load prediction model based on the individual household level by the use of CART, SVM, and MLP, NNs for 24-hour short-term load forecasts has been presented. It is revealed that an integration of the past utilization data and household pattern data could mainly improve the prediction of every user load. The results of 51% MAPE are reached and 48% is obtained by SVM.

This paper presents a new load forecasting at the household level. The precise forecasting of time series data has inspired the researchers to propose new techniques for load management. Since the time series data often does not hold linear as well as nonlinear patterns, ARIMA or NNs (NN) are not enough to model and predict time series data. This paper presents a hybridization of the ARIMA and NN model to make use of the advantages of time series models and NN. The proposed ARIMA-NN model is evaluated and the simulation outcome strongly pointed out its superior performance over the compared methods.

The subsequent portions are arranged as follows. Section 2 offers the presented ARIMA-NN model and Section 3 evaluates the ARIMA-NN model. Section 4 concludes.

2. The proposed ARIMA-NN model

Fig. 1 shows the overall processes involved in the ARIMA-NN-based electricity forecasting in smart meters. Initially, the data is gathered from various sources such as datasets, actual readings, and calculations. Then, the ARIMA-NN model will be trained using the data and generates a model. Afterward, the generated model can be utilized to predict the electricity consumption by the application of new building data. The detailed operation of the ARIMA-NN model will be discussed in the following subsections.

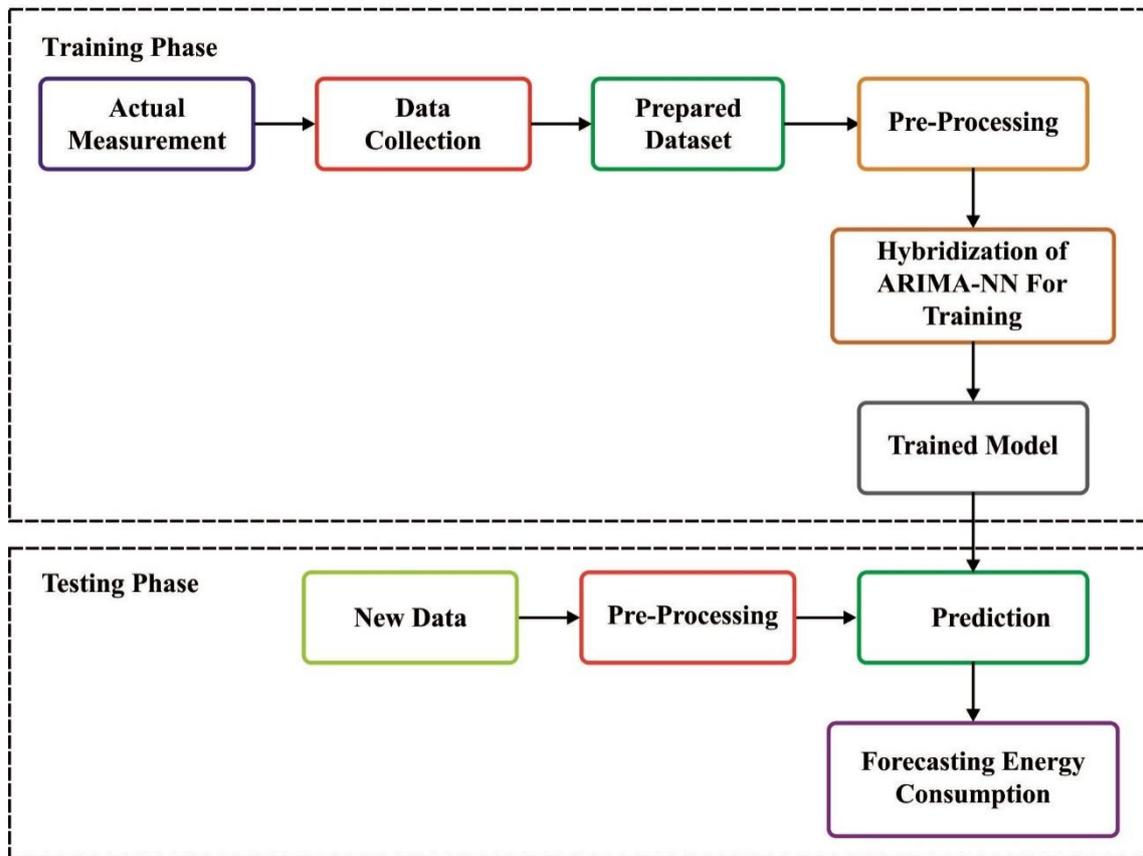


Figure 1. Block diagram of ARIMA-NN based electricity prediction

2.1. ARIMA modeling approach

$\varphi_p(B)\Phi_J(B^S)\nabla^k\nabla_S^D z_f = \theta_l(B)\Theta_L(B^S)a_t(1)$ ARIMA [13] consists of a future value with the parameter of the linear function that is obtained from various monitoring results as well as random faults. The ARIMA technique could be defined as ARIMA (j, k, l)(J, k, L)_s, where (j, k, l) denotes the non-seasonal portion of a method (P, D, Q)_s is assumed to be the seasonal part of the proposed technique that is described as follows

where j denotes the sequence of non-seasonal autoregression, k indicates the number of general differentiation, implies the linearity of non-seasonal MA, J signifies the order of seasonal autoregression, K specifies the count of seasonal difference, L is assumed to be ordered in seasonal MA, S reflects the length of periodicity, φ resembles the AR operator in order j, Φ is a seasonal AR parameter from order J, ∇^k denote the varying operator, ∇_S^K is a seasonal differencing operator, z_t implies a monitored value at time t, θ indicate the MA operator of order l, Θ is seasonal MA parameter in order L as well as a_t shows the noise of stochastic technique which is considered as $NID(0, \sigma^2)$.

ARIMA modeling scheme has 3 major levels that are given below:

- Model identification
- Parameter calculation
- Diagnostic monitoring

The process of identifying normal form has two phases are,

- The required differencing of series could be determined to attain the stationery as well as normality.

- The temporal association of transferred data might be discovered by analyzing the autocorrelation (ACF) and Partial Auto Correlation (PACF) functions.

In addition, ACF is assumed to be applicable in measuring primary values with the series of later values. PACF indicates of correlation among a parameter and lagging the correlations from the lower order. By assuming the graphs of ACF and PACF that are attained by electricity load concentration sequence, diverse different ARIMA techniques have been discovered for selecting models. Any method that offers a lower Akaike Information Criterion (AIC) can be chosen as an optimal fitting approach. Hence, the arithmetic form of AIC is presented as

$$AIC = n(\ln((2\pi RSS)/n) + 1) + 2m \tag{2}$$

where $m = (j + l + J + L)$ is several terms evaluated from the technique as well as RSS displays the combination of squared residuals.

Once the models of ARIMA have been illustrated, the attributes of such functions should be calculated. If a proper method is selected then parameters could be computed with the application of the Box-Jenkins method that analyzes the remaining model of adequate series. Various samples have been deployed for diagnostic observation in determining the residuals of chosen ARIMA from ACF and PACF graphs are not based on other frameworks. When normality metric is not given, the observations have been obtained in provided by a Box-Cox transformation. To accomplish a better forecasting method, residuals that are retained with the fitting model, should convince the needs of the white noise process. For computing, the electricity load time series is a dependent or independent, Residual Auto-Correlation (RACF) function is worked from the survey. Initially, a correlogram is obtained by segmenting the remaining ACF function in contrast with the lag value. In the case of ineffective ARIMA, the evaluated autocorrelations of residuals have been uncorrelated and distributed with the gradual value of zero.

Alternatively, Ljung-Box-Pierce statistics have been introduced to test the empty hypothesis of present autocorrelations as white noise; test statistics can be measured for a diverse number of prior lagged autocorrelations with the help of Ljung-Box-Pierce statistics ($L(r)$ test) to test the accuracy of this method. Thus, $D(r)$ values have been related to the crucial test of (χ^2) distribution in terms of the degree of freedom with minimum significance. Finally, the cumulative period gram is applied for analyzing the white noise series. While the seasonal time is modeled in the proposed work, the periodic features of electricity load concentration-time cannot be considered. Hence, the regularities present in residuals must be examined.

2.2 Structure of NNModel

From several other NN infrastructures, the three-layer-feed forward back propagation system is generally applied. The architecture is comprised of a single hidden layer of neurons along with transfer functions as well as an external layer of linear neurons and linear transfer functions. The graphical representation of the back propagation network is depicted in Fig. 2, where $X_a(a = 1,,N)$ is the input variables; $n_i(i = I,,S)$ denotes the outcome of neurons in the hidden layer and $y_b(b = I,,L)$ denotes the final result of NN. In general, NN should be trained for determining the values of intensity that produce the accurate result. From the training level, a group of input data has been applied in training the networks for several times. Therefore, the back propagation technique adapts the weights present in the steepest descent direction (SDD). This is the process of decreasing operation in a rapid manner. However, the fundamental gradient descent learning model is worse when compared with other models in terms of accuracy and efficiency. The optimization point states that training provided for NN could be assumed as similar to reduce the overall error from network intensity. To overcome the complexities, various learning techniques have been proposed namely resilient backpropagation, Levenberg-Marquardt as well as conjugated gradient backpropagation. A major technology was established named as Scaled Conjugate Gradient (SCG) approach. Hence, SCG training models are mainly used in removing the time conservation inline search. From the conjugate gradient model, and exploring operation is processed including conjugate dimensions that provides a rapid convergence when compared with SDD. Furthermore, the distinct backpropagation method is applied in NN training which measures the gradient value of global error function in terms of weights, $f(W)$, for all process k , and advanced weights based on

$$W^{u+1} = W^u - \alpha^u \nabla f(W^u) \tag{3}$$

The step size $\alpha^u > 0$ is assumed to be a user-selected training parameter that influences the operation of the learning model in a vast range. For every case, the backpropagation technique follows a zigzag mechanism in a lower, typical of steepest gradient descent approach. Subsequently, the conjugate gradient model eliminates the zigzag process for a definite point by integrating the particular association between direction and gradient vector at alliteration. When D^k points the dimensional vector for iteration K , then weight vector is upgraded as

$$W^{u+1} = W^u + \alpha^u D^u \tag{4}$$

The provided values of W^u and D^u , a specific measure of α^k minimizes objective function to a greater extent. once the minimum number of processes is carried out the searching line direction is useful in identifying the optimal size of the original minimum. By calculating the maximum step size along with the SCG learning technique improves the speed of training as well as to removes the base of crucial user-selected attributes. A major opinion of this technique is to apply the factor ρ that is increased with all process from implementing the model by the quantity δ^u , that states a Hessian matrix is not positive definite. A detailed technology of SCG in NN could be computed.

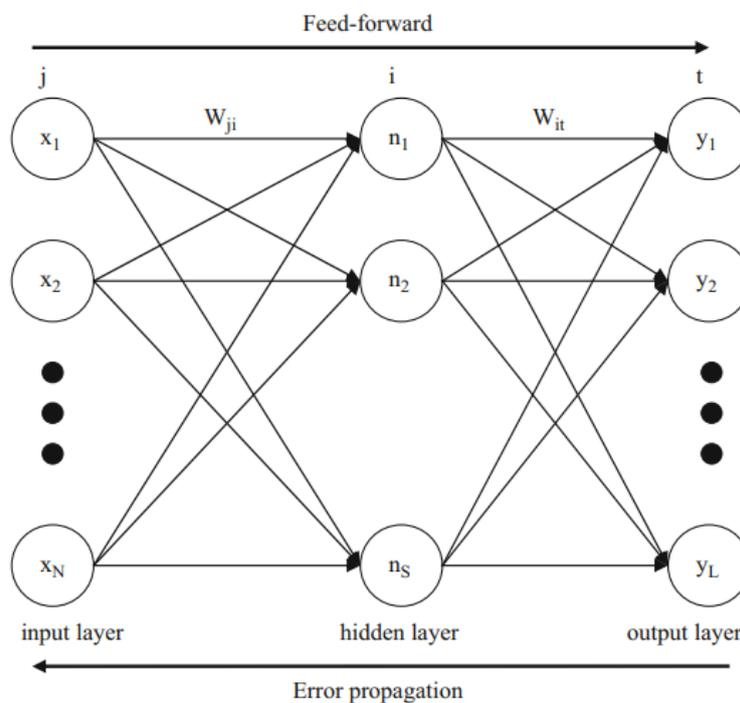


Fig. 2. Structure of feed-forward NN

2.3 Hybrid Model

The basic nature of electricity load data cannot be captured easily using the stand-alone technique due to the presence of different features like seasonality, heteroskedasticity, and so on. The approximate value of ARIMA is difficult to identify and not adequate. Besides, the application of the ANN technique is to hold the maximum result. [14] defined that it is ineffective to be applied in ANN which is not applicable for any data type. Hence, the hybrid technique with linear and nonlinear abilities is termed as the best model to predict the electricity load data. By the integration of various factors, the above criteria are obtained. As same as the hybrid model structure, an electricity load time series is comprised with a linear autocorrelation as well as a nonlinear unit as,

$$y_b = L_b + S_b \tag{5}$$

Where L_b is the series component and S_f implies non-linear element. The variable with time series data has been computed. The primary ARIMA would be deployed in capturing the linear unit and residuals are attained from the linear technique would have a nonlinear correlation. The residuals of h_b in time t is obtained from the sequential model as

$$h_b = y_b - \widehat{L}_b \tag{6}$$

Where \widehat{L}_b is assumed to be detected value of ARIMA method in time t . The analysis of residuals is mandatory in computing the accuracy of ARIMA. An ARIMA is not enough when there is a linear correlation structure that is presented in residuals. But, diagnostic checking of residuals must be pointed to for predicting the nonlinear patterns in sequential information. If the residuals are passed by diagnostic check, the method is sufficient in nonlinear relationships which are not exactly designed respectively. Hence, the residuals are capable of labeling with the help of ANN to find non-linear relationships. Using N input nodes, the ANN residuals would be provided as

$$h_b = g(h_{b-1}, h_{b-2}, \dots, h_{L-N}) + \varepsilon_r \tag{7}$$

Where g represents the non-linear notation computed by NN as well as ε_t implies the random error. At last, integrated detection would be as

$$\widehat{y}_b = \widehat{L}_b + \widehat{S}_b \tag{8}$$

3. Performance Evaluation

3.1. Dataset Description

For the validation of the presented ARIMA-NN model, it is tested using a UK Smart Meter dataset. The dataset contains a set of features comprises of a Household id, Plans used (standard or dynamic time of use), Date and Time, Meter readings (Kwh), and Acorn groups [15]. Fig. 3 shows the distribution of the household energy consumption from smart meters.

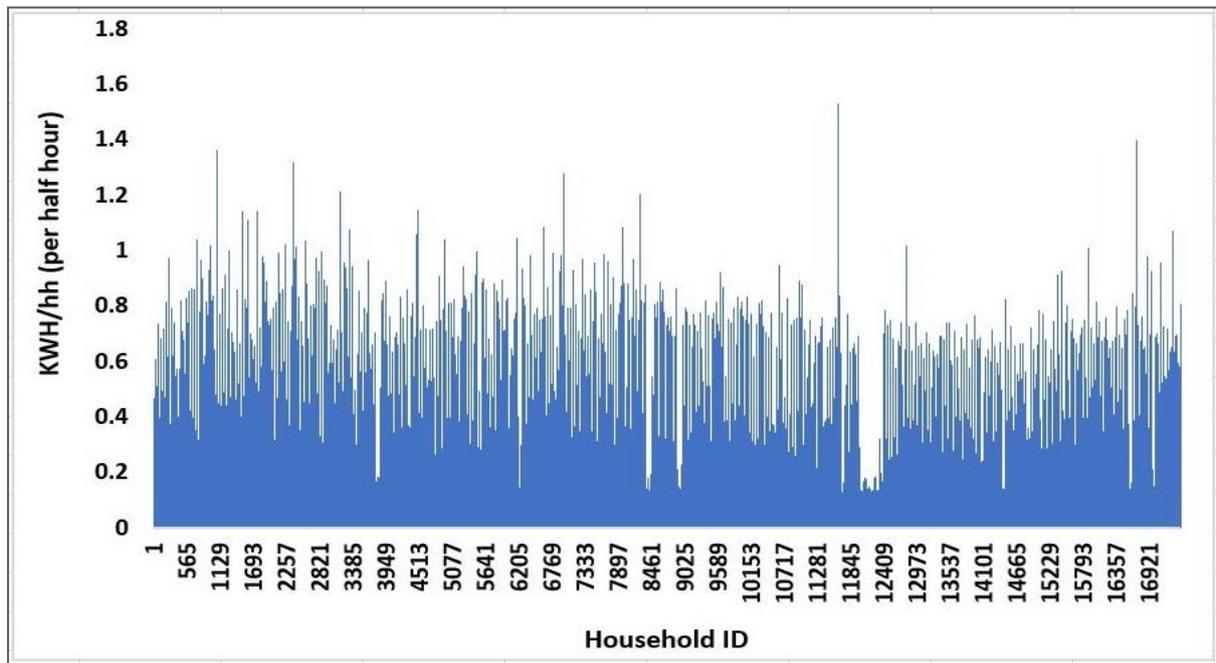


Figure 3: Smart Meter Power Consumption Distribution

3.2. Results analysis

Fig. 4 shows the results attained by the ARIMA-NN model concerning the actual and predicted energy usage of the top three household instances namely MAC000018, MAC000020, and MAC000021. The figures showcased that the ARIMA-NN model exhibits better prediction and there is only a slight variation exists between the actual and predicted energy utilization.

Table 1 shows the results offered by the presented model concerning MSE on three should dataset. From the table, it is evident that the MSE attained under the MAC000018, MAC000020 and MAC000021 are 0.018, 0.015, and 0.029 respectively. These minimal values portrayed the superior results of the presented ARIMA-NN model.

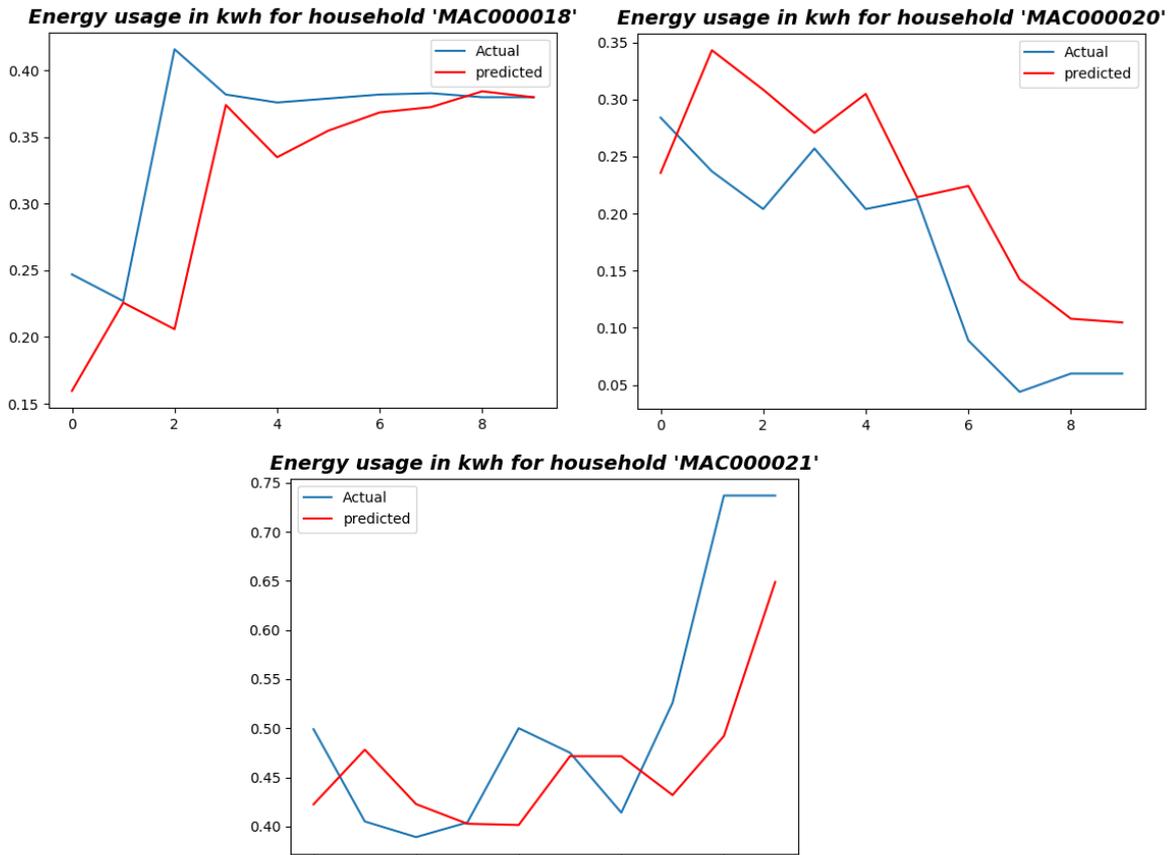


Figure 4. Top Three Households Energy Consumption Actual and Predicted

Table 1 Analysing of Expected and Predicted Energy Consumption Analysis of Different Households with Mean Square Error

S.No	MAC000018		MAC000020		MAC000021	
	Expected	Predicted	Expected	Predicted	Expected	Predicted
1	0.247	0.15963742	0.237	0.23562519	0.389	0.42244044
2	0.247	0.22582980	0.237	0.31441220	0.389	0.40467978
3	0.247	0.22251885	0.237	0.30714953	0.389	0.39447706
4	0.247	0.23229535	0.237	0.28566218	0.389	0.38810403
5	0.247	0.23769158	0.237	0.28879402	0.389	0.39204355
6	0.247	0.24168594	0.237	0.23421417	0.389	0.39106136
7	0.247	0.24699919	0.237	0.23699951	0.389	0.38899704
8	0.247	0.24699918	0.237	0.23699951	0.389	0.38899704

9	0.247	0.24699918	0.237	0.23699951	0.389	0.38899704
10	0.247	0.24699919	0.237	0.23699951	0.389	0.38899705
MSE= 0.018			MSE = 0.015		MSE = 0.029	

Table 2 shows the results attained by the ARIMA-NN model on the training data with compared methods in terms of MAPE, accuracy, and MSE. The table values indicated F_g method offered poor results over the compared methods with the MAPE of 51.86, the accuracy of 34.28, and MSE of 0.72. Then, the Z_{24} model shows a slightly better outcome over F_g with the MAPE of 42.94, the accuracy of 40.90, and the MSE of 0.61. Afterward, the ARIMA model shows moderate results with the MAPE of 38.24, the accuracy of 31.01, and the MSE of 0.35. Next to that, the SVR model shows moderate results with the MAPE of 28.73, an accuracy of 43.98, and MSE of 0.35. The NNET model attains a MAPE of 32.39, an accuracy of 39.99, and an MSE of 0.34. Simultaneously, L_f , L_b , and RPART models show somewhat manageable and near-identical results. The L_f model attains the MAPE of 36.08, the accuracy of 34.37, and the MSE of 0.32. The L_b model attains the MAPE of 36.03, the accuracy of 34.51, and the MSE of 0.32. The RPART model attains the MAPE of 37.08, an accuracy of 36.36, and an MSE of 0.32. At the same time, the KNN and proposed ARIMA-NN model offers superior and competitive results compared to other methods. The ARIMA-NN model attains a MAPE of 27.80, an accuracy of 44.32, and an MSE of 0.31.

Table 2 Result Analysis using Training Data on Proposed vs Existing Methods

Model	MAPE (%)	Accuracy (%)	MSE
Proposed	27.80	44.32	0.31
Z_{24}	42.94	40.90	0.61
F_g	51.86	34.28	0.72
ARIMA	38.24	31.01	0.35
L_f	36.08	34.37	0.32
L_b	36.03	34.51	0.32
KNN	33.36	36.15	0.31
RPART	37.08	36.36	0.32
RF	0.48	100.00	0.00
NNET	32.39	39.99	0.34
SVR	28.73	43.98	0.35

Table 3 Result Analysis using Validation Data on Proposed vs Existing Methods

Model	MAPE (%)	Accuracy (%)	MSE
Proposed	28.12	45.34	0.39
Z_{24}	40.33	43.68	0.59
F_g	45.74	37.59	0.77

ARIMA	50.60	18.57	0.49
L_f	33.72	35.81	0.43
L_b	34.19	35.07	0.40
KNN	35.01	33.28	0.39
RPART	38.69	37.00	0.42
RF	32.72	37.89	0.39
NNET	32.19	41.60	0.41
SVR	28.95	44.43	0.46

Table 3 shows the results attained by the ARIMA-NN model on the Validation data with compared methods in terms of MAPE, accuracy, and MSE. The table values indicated F_g method offered poor results over the compared methods with the MAPE of 45.74, the accuracy of 37.59, and MSE of 0.77. Then, the Z_{24} model shows a slightly better outcome over F_g with the MAPE of 40.33, the accuracy of 43.68 and the MSE of 0.59. Afterward, the ARIMA model shows moderate results with a MAPE of 50.60, an accuracy of 18.57 and an MSE of 0.49. Next to that, the SVR model shows moderate results with the MAPE of 28.95, an accuracy of 44.43 and MSE of 0.46. The NNET model attains a MAPE of 32.19, an accuracy of 41.60 and an MSE of 0.41. Simultaneously, L_f, L_b, and RPART models show somewhat manageable and near-identical results. The L_f model attains the MAPE of 33.72, the accuracy of 35.81 and the MSE of 0.43. The L_b model attains the MAPE of 34.19, the accuracy of 35.07 and the MSE of 0.40. The RPART model attains the MAPE of 38.69, an accuracy of 37.00 and the MSE of 0.42. At the same time, the KNN and proposed ARIMA-NN model offers superior and competitive results compared to other methods. The ARIMA-NN model attains the MAPE of 28.12 accuracies of 45.34 and MSE of 0.39.

Table 4 Result Analysis using Testing Data on Proposed vs Existing Methods

Model	MAPE (%)	Accuracy (%)	MSE
Proposed	25.53	48.38	0.21
Z_{24}	37.69	45.83	0.44
F_g	56.27	37.80	0.63
ARIMA	41.14	20.83	0.26
L_f	36.50	35.71	0.23
L_b	36.05	33.63	0.22
KNN	30.27	41.37	0.22
RPART	38.50	35.12	0.24
RF	32.57	41.07	0.22
NNET	30.28	42.26	0.21
SVR	26.78	47.02	0.23

Table 4 shows the results attained by the ARIMA-NN model on the testing data with compared methods in terms of MAPE, accuracy, and MSE. The table values indicated F_g method offered poor results over the compared methods with the MAPE of 45.74, the accuracy of 37.59, and MSE of 0.77. Then, the Z_{24} model shows a slightly better outcome over F_g with the MAPE of 45.74, an accuracy of 37.59 and MSE of 0.77. Afterward, the ARIMA model shows moderate results with a MAPE of 50.60, an accuracy of 18.57 and an MSE of 0.49. Next to that, the SVR model shows moderate results with the MAPE of 28.95, an accuracy of 44.43 and MSE of 0.46. The NNET model attains the MAPE of 32.19, an accuracy of 41.60 and the MSE of 0.41. Simultaneously, L_f , L_b , and RPART models show somewhat manageable and near-identical results. The L_f model attains the MAPE of 33.72, the accuracy of 35.81 and the MSE of 0.43. The L_b model attains the MAPE of 34.19, the accuracy of 35.07 and the MSE of 0.40. The RPART model attains the MAPE of 38.69, an accuracy of 37.00 and the MSE of 0.42. At the same time, the KNN and proposed ARIMA-NN model offers superior and competitive results compared to other methods. The ARIMA-NN model attains the MAPE of 28.12, an accuracy of 45.34 and the MSE of 0.39.

Figs. 5-7 shows the comparison of the results attained by diverse methods in terms of MAPE, accuracy, and MSE. The figures ensured the ARIMA-NN model shows superior outcomes over the compared methods in a significant way.

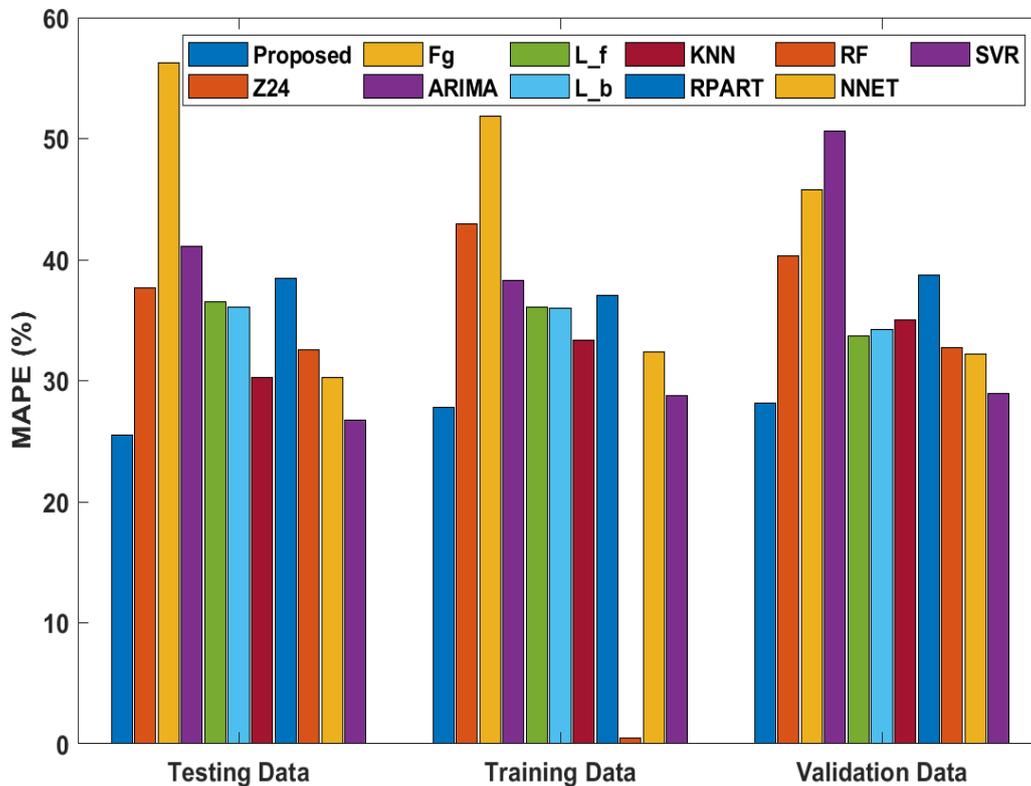


Figure 5: MAPE analysis of different methods

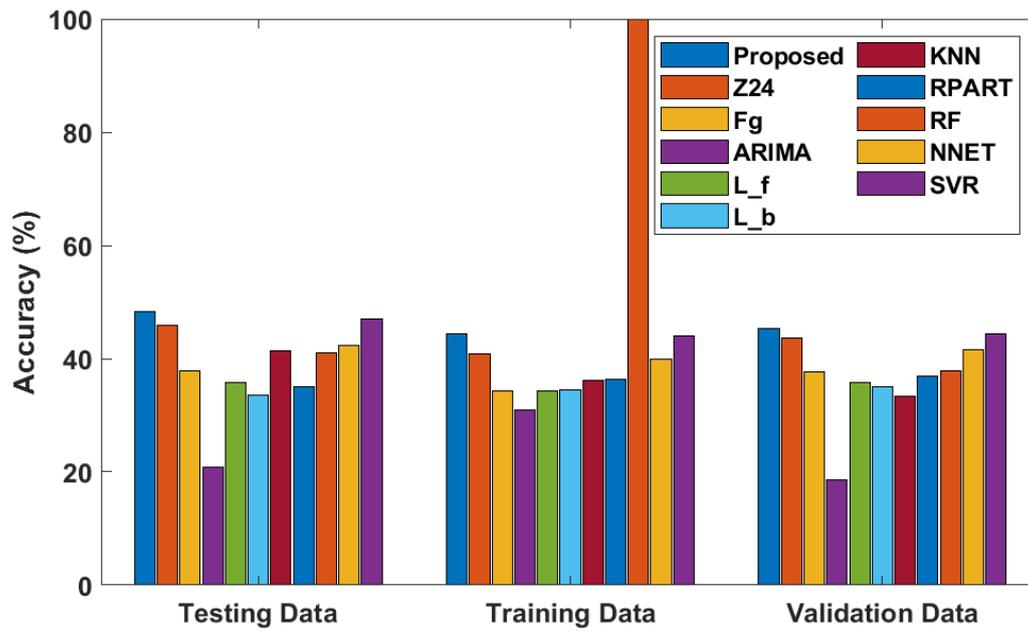


Figure 6: Accuracy analysis of different methods

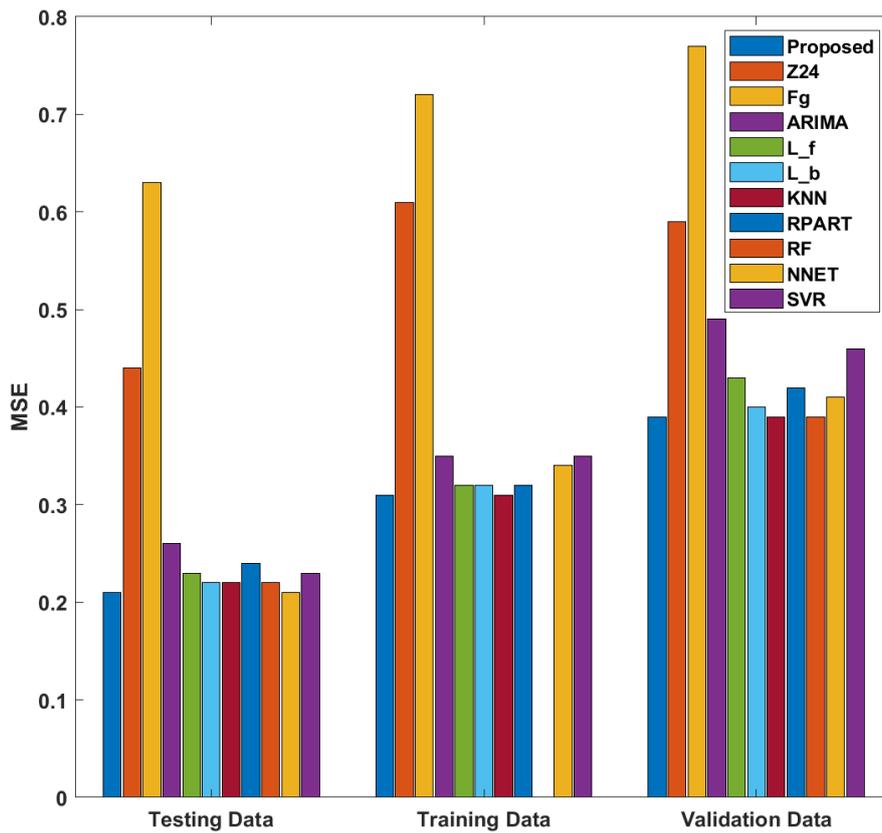


Figure 7: MSE analysis of different methods

4. Conclusion

This paper has developed an integration of the ARIMA model with the NN model for Electricity Consumption Forecasting using Smart Meter Data. As the time series data often does not hold linear as well as nonlinear patterns, ARIMA or NN models are not enough to model and predict the time series data. The ARIMA-

NN model is trained using the past data and creates a model for the validation process. Once the model is created, it can be used for the prediction of electricity consumption by the application of new building data. The proposed ARIMA-NN model is evaluated and the simulation outcome strongly pointed out its superior performance over the compared methods. The presented model has obtained effective testing performance with the MAPE of 25.53, an accuracy of 48.38, and the MSE of 0.21. As a part of the future scope, the proposed model can be deployed in real-time applications.

References

1. Anvari-Moghaddam A. Monsef H. Rahimi-Kian A , (2015). Optimal smart home energy management considering energy saving and a comfortable lifestyle. *IEEE Transactions on Smart Grid.*; 6(1): 324–332.
2. Zhou B. Li W. Chan K.W. Cao Y. Kuang Y. Liu X. Wang X, (2016). Smart home energy management systems: concept, configurations, and scheduling strategies. *Renewable and Sustainable Energy Reviews.*; 61: 30–40.
3. Wu X. Hu X. Moura S. Yin X. Pickert V., (2016). Stochastic control of smart home energy management with plug-in electric vehicle battery energy storage and a photovoltaic array. *Journal of Power Sources.* 333: 203–212.
4. Hu X. Murgovski N. Johannesson L. Egardt B. (2015) Optimal dimensioning and power management of a fuel cell/battery hybrid bus via convex programming, *IEEE/ASME Transactions on Mechatronics.* 20(1): 457–468.
5. Sun C. Sun F. Moura S.(2016). Nonlinear predictive energy management of residential buildings with photovoltaics & batteries. *Journal of Power Sources.*; 325: 723–731.
6. Edwards RE. New J. Parker LE.(2012). Predicting future hourly residential electrical consumption: A machine learning case study. *Energy and Buildings.*; 49: 591–603
7. Ziekow H. Goebel C. Strucker J. Jacobsen H.(2013). The potential of smart home sensors in forecasting household electricity demand. *IEEE International Conference on Smart Grid Communications.*
8. Singh RP. Gao PX. Lizotte DJ. (2013) On hourly home peak load prediction. *IEEE International Conference on Smart Grid Communications.*; pp. 163–168.
9. Ghofrani M. Hassanzadeh M. Etezadi-Amoli M. Fadali MS. (2011) Smart meter based short-term load forecasting for residential customers. *North American Power Symposium (NAPS).*; pp. 1–5.
10. Ding Y. Borges J. Neumann MA. Beigl M. (2015). Sequential pattern mining–A study to understand daily activity patterns for load forecasting enhancement. *IEEE First International Conference on Smart Cities.* pp. 1–6.
11. Veit A. Goebel C. Tidke R. Doblender C. Jacobsen HA. (2014) Household electricity demand forecasting: benchmarking state-of-the-art methods. *Proceedings of the 5th International Conference on Future Energy Systems.*; pp. 233–234.
12. Gajowniczek K. Ząbkowski T. (2015) Data Mining Techniques for Detecting Household Characteristics Based on Smart Meter Data. *Energies.* a; 8(7): 7407–7427.
13. Box, G.E.P., Jenkins, G.M., (1976). *Time Series Analysis Forecasting And Control.* Holden-Day, San Francisco, USA.
14. Zhang, P.G., (2003). Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing* 50, 159–175.
15. https://cc-uploads.s3.amazonaws.com/open_links/HumAIn+2019/UK_Smart_Meter.zip