International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 3s
DOI: https://doi.org/10.17762/ijritcc.v11i3s.6192
***Article Received**: 14 December 2022 **Revised**: 22 January 2023 **Accepted**: 04 February 2023*

Predicting Power Consumption of Individual Household using Machine Learning Algorithms

Sobhana M[1], Smitha Chowdary Ch[2], D.N.V.S.L.S. Indira[3], Gaddameedi Dinesh Kumar[4]
[1]Senior Assistant Professor
Department of Computer Science and Engineering
V R Siddhartha Engineering College
Vijayawada, Andhra Pradesh, India
sobhana@vrsiddhartha.ac.in
[2]Professor
Department of Computer Science and Engineering,
Koneru Lakshmaiah Education Foundation
Vaddeswaram, Guntur 522302, A.P., India
smithacsc@gmail.com
[3]Associate Professor
Department of Information Technology,
Seshadri RaoGudlavalleru Engineering College
Gudlavalleru, Andhra Pradesh, India.
indiragamini@gmail.com
[4]Department of Computer Science and Engineering,
Koneru Lakshmaiah Education Foundation
Vaddeswaram, Guntur 522302, A.P., India
dineshzee@gmail.com

Abstract—Climate change, as known, is the dangerous environmental effect we are going to face in the near future and electricity contributes the majority of its part in overcoming climate change as per the trends. Usage of electricity is widely increasing all over the world mainly as an alternative to the use of fossil fuels. In households the usage is rapidly increasing day by day, owing to the increase in the number of devices running on electricity. As we have observed mainly after the relaxation of the lockdown the bills received by households, especially in cities were unhappy and have left most of the people aghast. It is evident that users have no idea about the power they consume. In this work, a model to forecast the electricity bill of household users based on the previous trends and usage patterns by making use of machine learning techniques has been proposed. The historical data of the user is studied and the learning is done iteratively to improve the accuracy of the model. The model can then be used to forecast the consumption beforehand.

Keywords-— Power consumption, Time series forecasting, Bill forecasting, ARIMA, Seasonality.

I. INTRODUCTION

As human beings evolved through early ages, they have invented something or the other to make lives easier day by day. One such invention that first came into the limelight is the concept of ionization (light) approximately 13 billion years ago. Over time, the needs of humans have increased and increased which led to the invention of electricity for various kinds of purposes in daily life, in such a way that in the present situation there is no mankind if there is no electricity [1]. The household consumer is not aware of the usage of electricity that in turn affects him economically and the climate [2]. Especially households have been our area of interest because the disruptions caused in the power (outages, swells, sags, and unbalanced) due to the non-uniform usage pattern of residential users are huge [3]. Especially after the relaxation of lockdown in a country such as India where the population is high, every household was worried mostly about a common situation i.e. electricity bills. Our idea is to help people have a long-term analysis on the usage of electricity to help cut down unnecessary costs in bills. So, a time-series model has been implemented to accurately predict the future month's bill. Initially, each module makes use of the historical data of the consumer, then it constructs a prediction system to forecast the future value. The predicted value is displayed to the user giving them an idea what will be the usage of their electricity in the next month. Electricity is also non-storable, whatever is produced needs to be consumed immediately [4]. Proper production planning is required for this purpose. Forecasting methods are being employed by various agencies to identify the demand and to produce only what is required

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 3s
DOI: https://doi.org/10.17762/ijritcc.v11i3s.6192
***Article Received**: 14 December 2022 **Revised**: 22 January 2023 **Accepted**: 04 February 2023*

[5]. Every industry strives towards leaving a low carbon footprint and conventional power consumption techniques lead to a lot of pollution [6]. With proper demand in mind, the agencies can limit themselves to produce only the required amount. Identifying demand also helps these agencies to provide quality power without swells and outages. ARIMA is a time series forecasting technique widely in use. It considers the previous history of the data and tries to identify a pattern within the distribution. This ensures nearly accurate predictions on test data. We have used the ARIMA model to train the data and to predict the future consumption. The rest of the paper is divided into literature survey (Section II), Proposed System (Section III), Methodology (Section IV), Results and Analysis (Section V), Performance Evaluation (Section VI) and Conclusion (Section VII).

II. LITERATURE SURVEY

Consumers are interested in knowing where surplus consumption is occurring and always tries to avoid these if possible. This problem has been there for a long time.

Several studies have been found to predict electricity consumption. This has become a prime area of research and new improvements and techniques to predict not only the consumption but also the future demand is being introduced.

The problem of energy consumption prediction is a type of multivariate time series Deb C, et al. [7]. Therefore, the times series has to be segmented into periods within which there is similarity in consumption pattern Arghira N, et al. [8]. Every time series can be divided into three components: Trend, Seasonality and Random components identifying and removing the trend and seasonality components is important for analysis Chujai P, et al. [9].

Forecasting of power consumption requires the identification of useful features. Proper features improve the forecast accuracy. These features can be obtained by finding the relation between the consumption and input features Oveis Abe, et al. [10]. Many of the present-day techniques lack efficient feature selection methods and suffer loss of accuracy Zhiyong Du, et al. proposed a method [11]. Dong, et al. [12] proposed a Support Vector machine technique for identifying the useful features of time series data and prepared a model for predicting energy demand of buildings in tropical regions. Another model was built on the same lines for predicting the annual energy demand of a building using heat transfer coefficients and SVM Xiao, Z. et al. [13].

Noise is another problem with consumption data. It gives rise to the nonlinear relations of output with input. CNN-LSTM techniques are used in the case of noisy data. CNN can be used to remove the noise and LSTM divides the time series and models the data to generate predictions. Ahmed, et al. used deep neural network for forecasting energy demand based on climatic information, date and building usage rate by eliminating noise [14]. Lee, D et al. used deep learning to predict consumption levels in environment [15]. Greff K, et al. implemented recurrent neural network to measure the consumption level using querying [16]. Sahebalam, et al. devised a comprehensive model to predict the consumption of energy using time series [17]. Munaf, R. Nafeena Abdul, et al. has developed four models for assessment of power ingestion in household using data communication and machine learning [18]. Himeur, Yassine, et al. has given recommendations to select good data set and have applied machine learning algorithms for identifying anomalous power consumptions [19]. Chou, Jui-Sheng, et al. has contributed for the formalized method for identifying anomalous patterns in large data sets in two stages consumption prediction and anomaly detection respectively [20]. Yang, Wangwang, et al. has proposed a model on combined deep learning load forecast for power consumption of a sing household user to achieve high-accuracy and stable load forecasting [21]. Casella, Enrico, et al. has suggested machine learning algorithms for power consumption in individual households [22]. Kim, Tae-Young, et al. has proposed a CNN-LSTM neural network that can extract spatial and temporal features to effectively predict the housing energy consumption [23]. Shin, Sun-Youn, et al. has worked on random forest, XGBoost and LSTM and made a comparative analysis and concluded that different machine learning algorithms wor best for identifying power consumption [24]. Yan, Ke, Xiaokang Zhou, et al. has used NLSTM neural networks for power usage prediction [25]. Tomar, Dimpal, et al. has used sliding window algorithm for energy usage in residential building [26].

Muhammad Ridwan Fathin, et al. has made a study to help electricity providers in taking decisions by doing medium and long-term forecasts using the Auto-Regressive Integrated Moving Average (ARIMA) method, ARIMA is the best for tactical decisions (medium-term) regarding electrical energy consumption [27]. Ni Guo, Wei Chen, et al. has propoaed a model ARIMA-SVR in comparision to ARIMA, ARIMA-GBR, LSTM and GRU for power consumption prediction of an office building [28].

III. PROPOSED SYSTEM

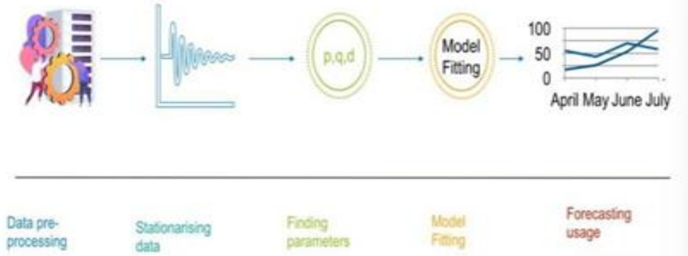

Fig 3.1: Proposed System

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 3s
DOI: https://doi.org/10.17762/ijritcc.v11i3s.6192
***Article Received**: 14 December 2022 **Revised**: 22 January 2023 **Accepted**: 04 February 2023*

Our proposed system consists of the following steps:

A. Data Preprocessing

The entire data is studied by visualizing as graphs and potential noise elements and missing values are identified and removed.

B. Stationarising the data

To present the data to the ARIMA model it is essential that the data needs to be distributed stationarily. Stationary data is the data distribution in which the mean is constant throughout the distribution. In general, any time series data either follows an increasing or a decreasing trend. To properly predict the value, it is important to stationarise and remove the trend component. To stationarize the dataset differencing is done, where each value is replaced by the difference of itself and its predecessor. Differencing beyond a point may lead the data to overfit the model which gives false results. Therefore, it is important to difference only till what is required.

C. Finding parameters (p, d, q)

The number of times differencing is done is called the order of differencing and is denoted by d. ARIMA model. ARIMA model also requires the order of AR model (p) and the order of MA model (q). The method employed to obtain the p, q values for the model here is ACF and PACF graphs. The ACF and PACF graphs of the stationary data are observed.

The AR model identifies the number of predecessors which can be used to accurately predict the present value. The graph is observed to find the predecessors or lags. Only p no. of lags is sufficient to obtain present value [17].

$$y_t=\delta+\phi_1 y_{t-1}+\phi_2 y_{t-2}+\ldots+\phi_p y_{t-p}+\varepsilon_t$$

Here, $\Phi 1$, $\Phi 2 \ldots \Phi p$ denote constant coefficients and e denotes the error.

The MA model identifies the pattern in the random or noise term and tries to apply the same for future value. Its equation is

$$Y_t=\beta_2+\omega_1\varepsilon_{t-1}+\omega_2\varepsilon_{t-2}+\ldots+\omega_q\varepsilon_{t-q}+\varepsilon_t$$

The ε terms represent the errors observed at respective lags and the weights (ω_1, $\omega 2$ …ω_q) are calculated statistically depending on the correlations [17]. q denotes the moving window denoting the number of previous error instances which can be significantly used for the prediction of present error.

D. Model Fitting

Once the parameters are obtained, they are provided to the ARIMA model which fits itself to the training data. Graphs showing the results of the ARIMA model as a comparison between actual and predicted values are shown in this phase.

E. Forecasting

The trained model is then made to predict for the next months, we have made the model to predict the bill for the next five consecutive months. This generates a forecasted value and a 95% confidence interval within which the value may lie.

IV. METHODOLOGY AND DATASET

The dataset provides power consumption information about asingle household.

1) It has observations of consumption with data recorded every month over a period of 24 months.
2) Temporal coverage is January 2019 to December 2020. The dataset contains the attributes: bill_date, meter_status, due_date, openimg_reading, closing_reading, net_units and bill_amount. We performed data cleaning and transformation; the missing data was filled with the previous month's data.

Time-series model is made by using ARIMA as it shows forecasted values over a confidence interval. Time-series data cannot be split randomly into train and test. The first 85percent of the observations were used for training themodel and the remaining 15 percent for testing.

A. Data Visualization

The dataset containing the observations was plotted with x axis as the date and y axis as units consumed in kWh. Upon carefully analyzing the plot, it can be observed that there is a seasonal component which repeats every year. There is an increase till the month of June and a slow decrease from the month of June to December. This clearly indicates the nonlinear distribution of the data.

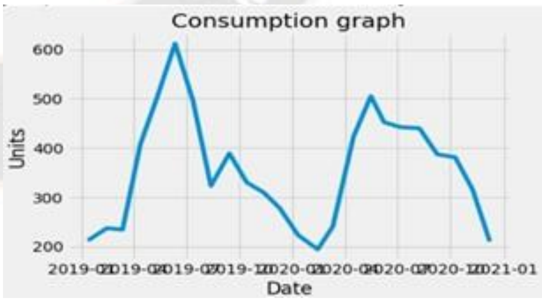

Fig 4.1: Data distribution

B. ADF Test

Before providing the data to the model for training, the data needs to be stationary. To check for the stationarity of the data the trend and seasonality components of the distribution wereobserved.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 3s
DOI: https://doi.org/10.17762/ijritcc.v11i3s.6192
***Article Received**: 14 December 2022 **Revised**: 22 January 2023 **Accepted**: 04 February 2023*

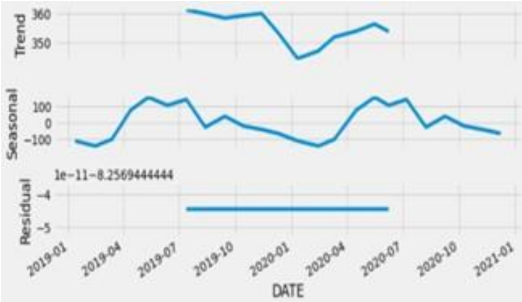

Fig 4.2: Components of the data

To stationarise the data we remove the trend and seasonal component. After removing these we perform an ADF test to see if the null hypothesis can be rejected to infer that the data is stationary. To reject the null hypothesis the p value should be less than 0.05.

The p value for the data is of high significance and hence cannot be rejected. Therefore, the data is not stationary. Differencing is performed to stationarise the data. The order of differencing is 1(d). The data becomes stationary after 1 order differencing. Differencing with greater order tends to overfit.

C. Autocorrelation and Partial Autocorrelation Function

The ARIMA model requires three parameters (p,q,d). The parameter d is the differencing order which is 1. Parameters p and q are the orders obtained from AR and MA models using ACF and PACF graphs respectively.

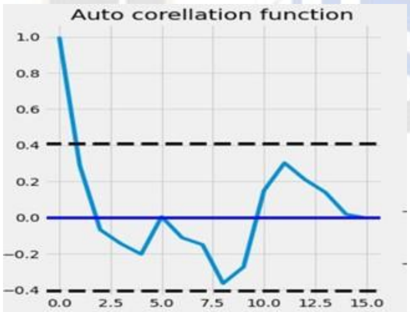

Fig 4.3: ACF Graph

The values of p and q are the values when the ACF and PACF graph moves from positive to negative. Therefore, the value of p is 2. Similarly, the value of q is 2 from the graph. The parameters (p, q, d) are provided to the model for creating the model.

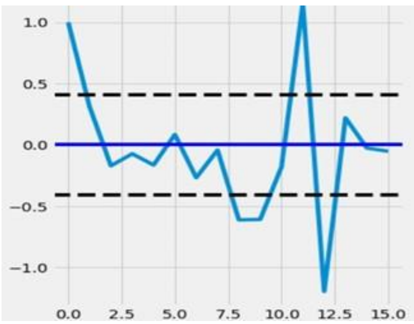

Fig 4.4: PACF Graph

D. Prediction

The Residual Sum of Squares (RSS) of the data is close to 0. This indicates that data is distributed stationarily. The obtained (p, q, d) values are provided to the ARIMA model.

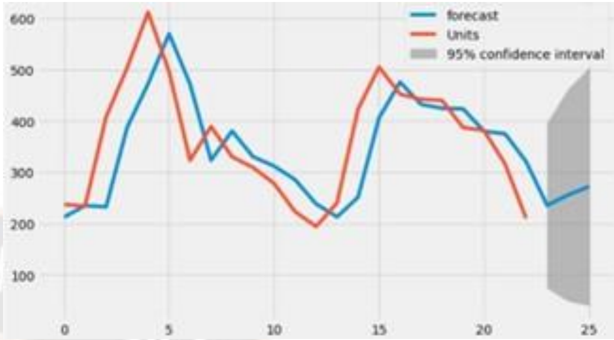

Fig 4.5: Prediction

The trained model is provided with the test data. The value for the next few months is obtained from the graph. The graph also gives the 95 percent confidence interval within which thepredicted value may lie.

V. RESULT AND ANALYSIS

The results of the ARIMA model are presented in Table I. The model transforms the data by eliminating the trend and seasonality from the data. It obtains the cyclic property of the data and tries to predict the consumption of the current month accurately. There are still differences between the actual and predicted values, but the predicted values follow the same trend as the actual values. This can be understood by considering the seasonality analysis component of the data and it also considers the lags which is the relation between previous and current data. The same model was used to predict the outcome of the next month.

TABLE I. ARIMA PREDICTED VALUES

Actual Value	Predicted Value
387	402
381	385
316	316
210	222
231	231
252	255

The table shows the actual and predicted values of the test data of the model. The MAPE value for the ARIMA modelwas found to be 0.023.

VI. PERFORMANCE EVALUATION

We have observed that time series forecasting provides more accurate results for consumption data than regression

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 3s
DOI: https://doi.org/10.17762/ijritcc.v11i3s.6192
***Article Received**: 14 December 2022 **Revised**: 22 January 2023 **Accepted**: 04 February 2023*

methods. Linear regression was applied for the same data and the results obtained are as shown in Table II.

TABLE III. LINEAR REGRESSION PREDICTED VALUES

Actual Value	Predicted Value
387	402
381	406
316	410
210	414
231	418
252	422

The line obtained using linear regression was found to have positive slope and hence the predicted values tend to increase. This is not the case with the actual data. The usage data is nonlinear and hence the linear regression model was found to be less accurate for prediction. The accuracy obtained by using linear regression was 36 percent. This provides the evidence of the nonlinear distribution of data.

VII. CONCLUSION AND FUTURE WORK

A comprehensive model has been found to predict power consumption of households. The model considers the trend and seasonality components of the usage data. Box - Jenkins method was used iteratively to improve and approach the most accurate model. We have used time series forecasting for prediction. ARIMA model not only considers the trend and cyclicity but also the previous history. It makes use of the previous models to predict the value of current month. This improves the accuracy of the model.

More work can be done in the area by including seasonal data and considering other techniques like neural networks. This requires large amounts of data. With the data in hand ARIMA model was found to provide better accuracy over other methods.

REFERENCES

[1] Sieminski A. International energy outlook. Energy Information Administration (EIA); 2017. p. 5-30.

[2] Ugursal, V.I. Energy Consumption, Associated Questions and some Answers. Appl. Energy 2014, 130, 783–792.

[3] Beckel C, Sadamori L, Staake T, Santini S. Revealing household characteristics from smart meter data. Energy 2014;78: 397-410.

[4] Tae-Young Kim, Sung-Bae Cho, Predicting residential energy consumption using CNN-LSTM neural networks, Elsevier Ltd, pp: 72-82.

[5] Streimikiene, D. Residential Energy Consumption Trends, Main Drivers and Policies in Lithuania. Renew. Sustain. Energy Rev. 2014, 35, 285–293.

[6] Zuo, J.; Zhao, Z.Y. Green Building Research- Current Status and Future Agenda: A review. Renew. Sustain. Energy Rev. 2014, 30, 271–281.

[7] Deb C, Zhang F, Yang J, Lee SE, Shah KW. A review on time series forecasting techniques for building energy consumption. Renew Sustain Energy Rev 2017;74:902-24.

[8] Arghira N, Hawarah L, Ploix S, Jacomino M. Prediction of appliances energy use in smart homes. Energy 2012;48(1):128-34.

[9] Chujai P, Kerdprasop N, Kerdprasop K. Time series analysis of household electric consumption with ARIMA and ARMA models. Proc. of the Int. MultiConf. of Engineers and Computer Scientists 2013;1:295-300.

[10] Oveis Abedinia, Nima Amjady and Hamidreza Zareipour. A new feature selection technique for load and price forecast of electrical power systems, IEEE transaction on power system, pp.1-11,2016.

[11] Zhiyong Du, Xianfang wang, Liyuan Zheng and Zhulin Zheng. Nonlinear system modeling based on KPCA and MKSVM, International colloquium computing, communication, control and management, pp 1-4, 2009.

[12] Dong, B.; Cao, C.; Lee, S.E. Applying Support Vector Machines to Predict Building Energy Consumption in Tropical Region. Energy Build. 2005, 37, 545–553.

[13] Xiao, Z.; Ye, S.-J.; Zhong, B.; Sun, C.-X. ”BP neural network with rough set for short term load forecasting”, Expert Syst. Appl. 2009, vol 36, pp 273-279.

[14] Ahmad, M.W.; Mourshed, M.; Rezgui, Y. Trees vs Neurons: Comparison between Random Forest and ANN for High-resolution Prediction of Building Energy Consumption. Energy Build. 2017, 147, 77– 89.

[15] Lee, D.; Kang, S.; Shin, J. Using Deep Learning Techniques to Forecast Environmental Consumption Level. Sustainability 2017, 9, 1894.

[16] Greff K, Srivastava RK, Koutník J, Steunebrink BR, Schmidhuber J. LSTM: a search space odyssey. IEEE Trans. on Neural Networks and Learning Systems 2017;28(10):2:222-32

[17] A. Sahebalam, Soosan Beheshti, Wael Khreich, Edward W. Nidoy, A Novel Approach in Household Electricity Consumption Forecasting, 2016 IEEE Canadian Conference on Electrical and Computer Engineering (CCECE).

[18] Munaf, R. Nafeena Abdul, et al. "An ML Approach for Household Power Consumption." 2022 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS). IEEE, 2022.

[19] Himeur, Yassine, et al. "Building power consumption datasets: Survey, taxonomy and future directions." Energy and Buildings 227 (2020): 110404.

[20] Chou, Jui-Sheng, and Abdi Suryadinata Telaga. "Real-time detection of anomalous power consumption." Renewable and Sustainable Energy Reviews 33 (2014): 400-411.

[21] Yang, Wangwang, et al. "A combined deep learning load forecasting model of single household resident user considering multi-time scale electricity consumption behavior." Applied Energy 307 (2022): 118197.

[22] Casella, Enrico, Eleanor Sudduth, and Simone Silvestri. "Dissecting the problem of individual home power consumption prediction using machine learning." 2022 IEEE

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 3s
DOI: https://doi.org/10.17762/ijritcc.v11i3s.6192
***Article Received**: 14 December 2022 **Revised**: 22 January 2023 **Accepted**: 04 February 2023*

International Conference on Smart Computing (SMARTCOMP). IEEE, 2022.

[23] Kim, Tae-Young, and Sung-Bae Cho. "Predicting residential energy consumption using CNN-LSTM neural networks." Energy 182 (2019): 72-81.

[24] Shin, Sun-Youn, and Han-Gyun Woo. "Energy Consumption Forecasting in Korea Using Machine Learning Algorithms." Energies 15.13 (2022): 4880.

[25] Yan, Ke, Xiaokang Zhou, and Jinjun Chen. "Collaborative deep learning framework on IoT data with bidirectional NLSTM neural networks for energy consumption forecasting." Journal of Parallel and Distributed Computing 163 (2022): 248-255.

[26] Tomar, Dimpal, et al. "Deep learning neural network prediction system enhanced with best window size in sliding window algorithm for predicting domestic power consumption in a residential building." Computational Intelligence and Neuroscience 2022 (2022).

[27] Muhammad Ridwan Fathin, Yudi Widhiyasana, Nurjannah Syakrani, "Model for Predicting Electrical Energy Consumption Using ARIMA Method", Advances in Engineering Research, volume 207, 2021, 298-303.

[28] Ni Guo, Wei Chen, Manli Wang, Zijian Tian, Haoyue Jin, "Appling an Improved Method Based on ARIMA Model to Predict the Short-Term Electricity Consumption Transmitted by the Internet of Things (IoT)", Hindawi, Wireless Communications and Mobile Computing Volume 2021, Article ID 6610273, 11 pages.