Short Term Forecasting Of Household Power Consumption By Using Time Series Regression Models

Omkar V. Narhe, Kshitij S. Joglekar, Rohit S. Kolage, Maya P. Shelke

Computer Engineering Department,
Zeal College of Engineering and Research, Narhe, Pune, India.

Abstract : Due to the tremendous increase of the developing countries, the electricity supply and demand faces intense scarcity, and that has major impact on the wealth of the country and the everyday life of the residents in that country. Electricity supply has become a barrier in continuous development of nation. In this paper, we put forward a model for predicting the electricity consumption and and perspective for forecasting the consumption of electricity at the small level of house. The effect of inhabitants everyday activities and utilization of the appliances consumption of electricity and it's data is collected to improve the accuracy of the short-term forecasting model. The contribution of this project is we addressed short-term electricity forecasting for 1 week ahead on the personalized household level. We have developed and evaluate an autoregression model for multi-step forecasting household power Consumption.

Keywords: STF, Forecasting, Regresiion Models

I. INTRODUCTION

Different type of organizations are adopting information technology to accompany their jobs by providing sufficient storage units to store daily information and used that data to the most profit with numerous ways. coming up with by statement trends within the future is a method to use applied mathematics information to investigate information within the past that ar associated with this event. The results were wont to predict future events. statistic is that the order of historical information, that resembles the cluster or observation of the information that are collected over time consistent with the continual amount of your time. That collected information could already be in a very daily, weekly, monthly, quarterly, or yearly format, looking on that one is suitable to use. statistic information encompass four components: trend, seasonal result, cyclical, and irregular result. The analysis of a statistic used statement techniques to spot models from the past information. With the idea that the data can gibe itself within the future, we are able to so forecast future events from the occurred information.As electricity is important, it is more advisable for both developed and developing nations to make investments in the electrical network. It is necessary to adopt advanced scientific ways as both over and underinvestment can cause problems. Issues like electricity shortage can be caused due to underinvestment whereas overestimation can affect the investments in various biological process activities. Due to the rise in population, many countries are facing scarcity of resources. As electricity is seasonal and cannot be stored, it's demand has a distinctive feature.

II. LITERATURE SURVEY

Jin et al [1] presented a algorithm which is based on household electricity consumption. This approach included limit warning function. The main objective of this paper was to instinctively predit and modify home appliances power consumption, control overall power consumption within next thirty minutes. This approach was very time consuming process as it analyses result by comparing usage of each appliance.

Models/Algorithms used: electricity limit; load management; consumption algorithm.

Hino et al [2] proposed the system to model the household usage of electricity by using Gaussian mixture model i.e. GMM. The data of consumption of electricity is shown by a few calculated model variables, and combined to traditional consumption format with the Symmetrized Generalized KL i.e. SGKL divergence distance. To analyze electricity supply sequences and patterns by cogeneration systems and to utilize them with the traditional patterns of consumption of household electricity for developing a method which helpful for the optimal energy supply system in households. The main drawback of this paper is that it is Less reliable for huge scale data.

Models/Algorithms used:Consumption patterns of Electricity, Gaussian mixture model i.e. GMM, Hierarchical Clustering Model, KL-divergence Methodology..

Rosin et al [3] analysed household electricity consumption in country Estonia.Here the working day and holiday energy consumption of electricity is discussed and how the load transfers from day to day is reviewed. The feasible common investments are discussed in the final part. The main drawback of this paper is less accurate analysis of expected results in case of uneven pattern formation.

Models/Algorithms used: Reducing Consumption and Shifting Consumption, Pattern Analysis by comparison of various plots.

Imtiaz et al [4] approach included a further work which focuses on making forecast more accurate by using polynomial regression and polynomial curve for prediction with least standard deviation.

Models/Algorithms used: Multivariate regression analysis, Time series forecasting.

Shelke et al [5] refers Holt Winter model to classify the dataset into seasonal data for short term load forecast of electricity consumed by that institute in various seasons like winter, summer or monsoon. This paper uses k nearest neighbour algorithm to form clusters and classify how the load is getting distributed on seasonal occasions. The extra considered parameters are hard to tune with existing parameters like temperature,humidity,etc.

Models/Algorithms Used: Holt-Winters Model, General load forecasting technique, k-Nearest Neighbour algorithm,

III. MOTIVATION

Currently, people are unaware of their electricity usage. Estimation of the electricity usage will help the customers by linking their current usage with the future cost. This will contribute to better manage their cost by comprehending their own energy consumption and their future projection. Transparent electricity consumptions and future projections will help to easily comprehend in what manner are we actually consuming the resources and how it can affect our future budget. Technology alone cannot create awareness regarding energy consumption, but it will provide a way of using energy carefully.

IV. GOAL

This project aims to provide a better system than existing system which predicts household electricity consumption by using Autoregressive model (ARIMA) by providing past electricity consumption data collected every minute as input with as maximum as accuracy.

V. OBJECTIVES

- The main objective was to seek out a model to expeditiously forecast the electricity consumption in an exceedingly social unit by applying Box and Jenkins technique.
- This model may well be useful among the social unit in designing expenditures.
- It might even be useful on the availability facet for designing electricity demand for a specific social unit..
- To increase the Accuracy within the prediction of the Label: In our system we tryto increase the Accuracy within the prediction of the sort of electricity units consumed.

VI. PROBLEM STATEMENT

Given the rise of good electricity meters and additionally the broad promotion of generation of electricity technology like solar panels, there is a wealth of electricity usage knowledge accessible. Building a system supported autoregression model by victimisation the findings from autocorrelation plots for multi-step short-run prediction house power consumption.

Abbreviations and Acronyms():

RMSE stands for Root Mean Squared Error

ACF stands for Autocorrelation Function
PACF stands for Partial Autocorrelation Function
ARIMA stands for Autoregressive Integrated Moving Average

VII. METHODOLOGIES

1) *RMSE - (Root Mean Squared Error):*

The Root Mean squared Error (RMSE) (which is additionally known as because the root mean squared deviation,(RMSD) is a recently used measure of the distinction between values really ascertained from the surroundings that's being modelled and the values expected by a model.
Residuals tells how distinct the data points are from regression. It tells how intensive the data is about the line of best fit. It is generally used in weather science,forecasting, regression analysis to crosscheck the results.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(X_{obs,i} - X_{model,i})^2}{n}}$$

Taking root of the average square errors has some fascinating implications for RMSE. Since the errors are square before they're averaged, the RMSE offers a comparatively high weight to massive errors. this suggests the RMSE ought to be additional helpful once massive errors are significantly undesirable.

2) *WALK FORWARD VALIDATION APPROACH:*

In time series prediction, this analysis of models on historical knowledge is named backtesting. Walk-Forward Validation may be accustomed backtest the model on time series issues where a model is also updated on every occasion step new knowledge is received. In practice, we very likely can retrain our model as new information becomes accessible. This would offer the model the most effective chance to create sensible forecasts at on every occasion step.
There are few choices to make:

1. Minimum range of Observations. First, we should choose the minimum range of observations needed to train the model. this might be thought of because the window dimension if a sliding window is used (see next point).

2. sliding or increasing Window. Next, we'd like to make a decision whether or not the model are going to be trained on all information it has accessible or solely on the foremost recent observations. This determines whether or not a sliding or increasing window are going to be used. After a wise configuration is chosen for your test-setup, models will be trained and evaluated.

1. beginning at the establishment of the time series, the least range of samples occurring inside the window is employed to train a model.
2. The prediction of the consecutive time step is done by model.
3. The predictions are stored and are used against the best-known values
4. To incorporate the values that are best-know the window is maximized and the method is also in continual .

Because this system involves moving on the time series one-time step at a time, it is usually known as Walk Forward Testing or Walk Forward Validation. in addition, because a sliding or expanding window is employed to train a model, this methodology is additionally cited as Rolling Window Analysis or a Rolling Forecast.

3) *AUTOCORRELATION ANALYSIS:*

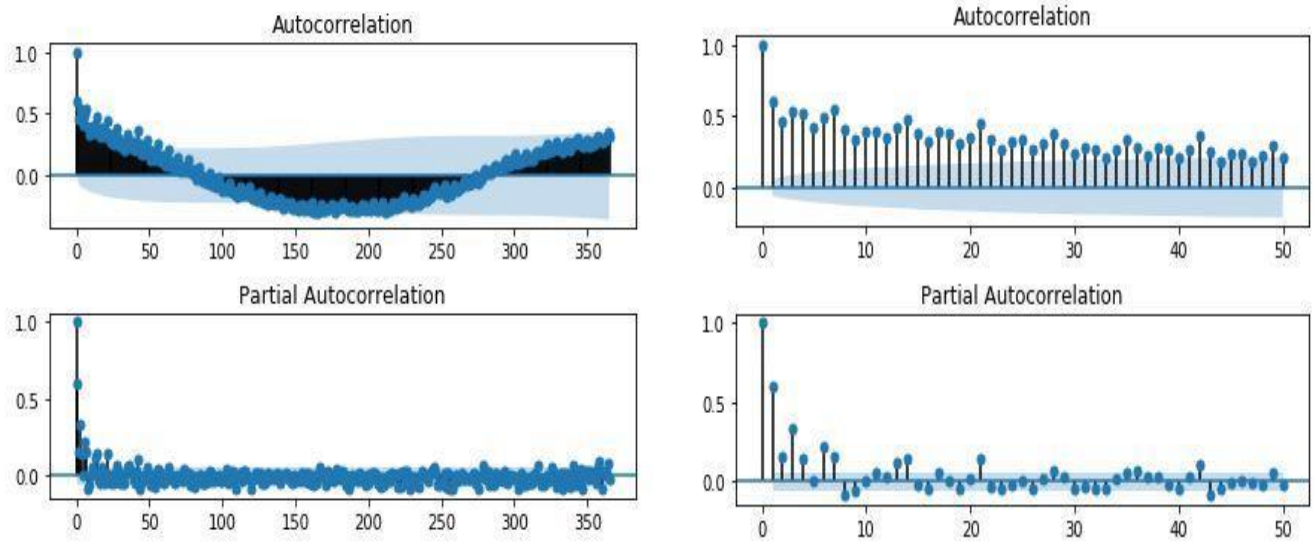
Autocorrelation could be a method used for statistical analysis. the aim is to measure the correlation of two values within the same knowledge set at completely different time steps. though the time knowledge isn't accustomed calculated autocorrelation, it slow increments ought to be equal so as to induce substantive results. The autocorrelation constant serves 2 functions. It will observe non-randomness in a very knowledge set. If the values within the knowledge set don't seem to be random, then autocorrelation will facilitate the Associate in Nursingalyst selected an applicable statistic model.

1. ACF

The correlation for statistical observations can be calculated with help of observations with preceding time steps .using the result of the correlation of the statistic observations, the values of this observations are calculated by using the values of the same kind of time series in preceding times, this can be recognised as a serial correlation, or AN autocorrelation. A plot of the autocorrelation of a statistic by lag is termed the AutoCorrelation Function, or the descriptor.ACF Autocorrelation function or the descriptor ACF are termed by a plot of the autocorrelation . This plot is typically recognized as a correlogram or An autocorrelation plot. We can plot AN autocorrelation plot mistreatment plot_acf() operate.

2. PACF

A partial autocorrelation (PACF) could be a outline of the link among an observation which are in very extraordinary or unusual statistic with observations at previous time steps with the relationships of observations which are intervening are rejected and not considered further again.The correlation that results when the effects of any correlation thanks to the terms at shorter lags is removed this relation is termed as the partial autocorrelation at lag k.Each positive correlation and indirect correlation is incorporated with an observation and an observation at a preceding time step in autocorrelation . These indirect correlations area unit a linear operate of the correlation of the observation, with observations at intervening time steps.The partial correlation function finds to get free of this that may be intuition for the partial autocorrelation this is due to these indirect correlations. We can plot a partial autocorrelation plot victimization plot_pacf() operate.



4) ARIMA Model:

An ARIMA model lies in the category of statistical models and is used for examining and calculating time series data. ARIMA which is a generalized version of the ARMA class. ARIMA can be classified into three sections as below:

• AR: stands for Autoregression. It takes the help of a dependent relationship between normal observations and lagged observations.
• I: stands for Integrated. It takes help of the difference between raw observations just to create a stationary time series.

• MA: stands for Moving Average. It takes the help of a dependent relationship between a normal observation and a residual error term.

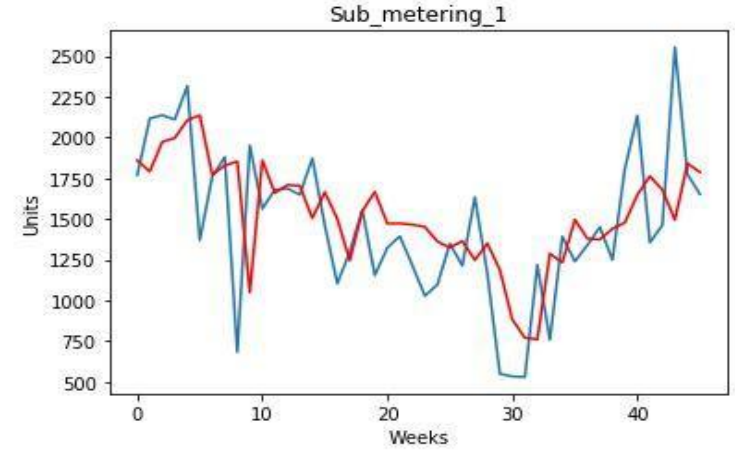
ARIMA model has three parameters i.e. 'p','d' and 'q'.

- 'p': It represents the level of autoregressive part.
- 'd': It symbolizes the level of initial differencing .
- 'q': It denotes the level of moving average .

VIII. RESULTS

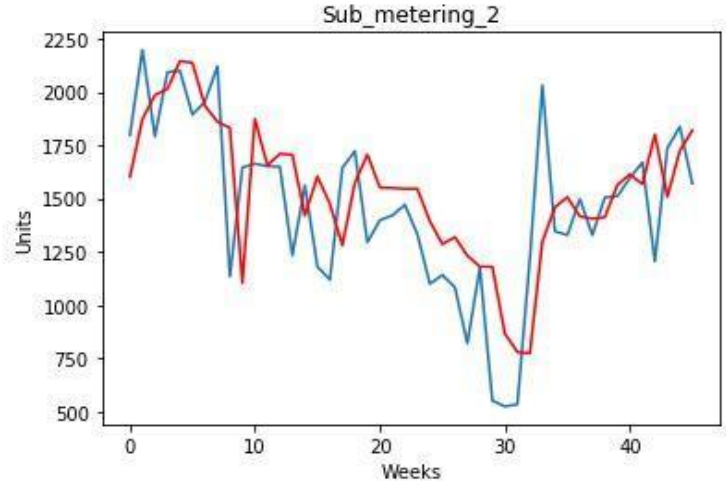
We used [6] Household Power Consumption dataset that describes the electricity consumption for single house for a period of 4 years which is obtained from UCI machine learning repository.

The dataset was divided into two parts. The model was trained using the dataset which comprised of 159 weeks records and the same model was tested using the data which comprised of remaining 46 weeks records.The model is evaluated using walk forward validation scheme.



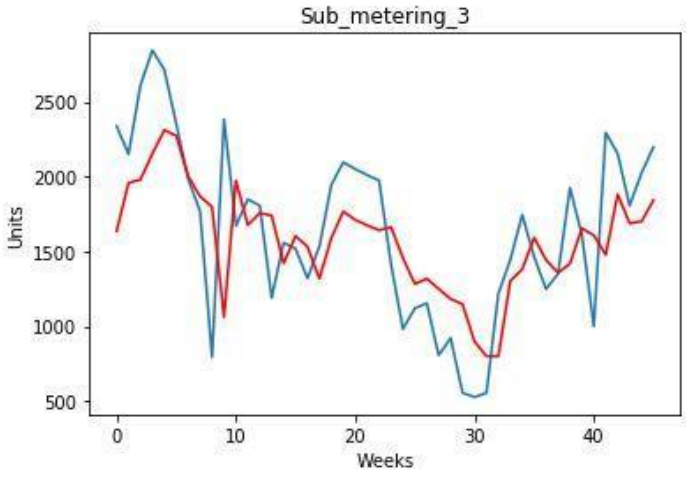
RMSE: 393.810477

Fig 1: Forecasting graph of submetering 1



RMSE: 306.024155

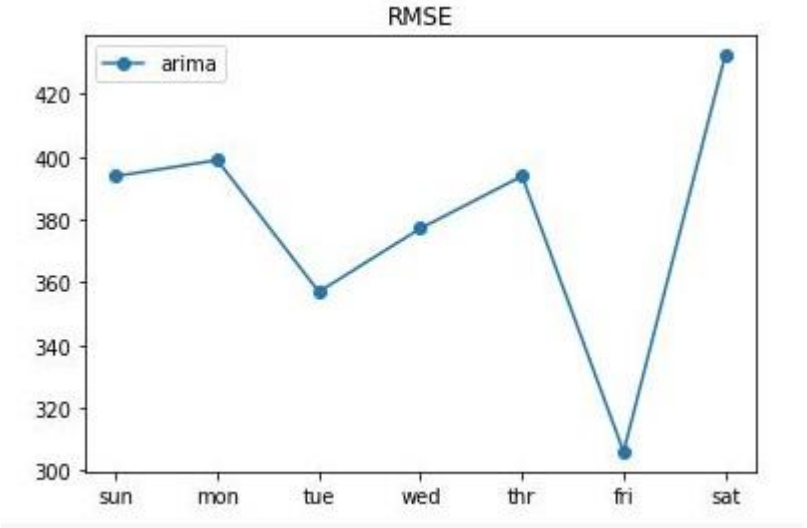
Fig 2: Forecasting graph of submetering 2.



RMSE: 432.193543

. Fig 3: Forecasting graph of submetering 3.

The model has attained the RMSE which is 381 KWatt.A line plot of the prediction which shows the RMSE in kilowatt is shown below



By observing the RMSE values in the above graph , it is clear that Friday is easy to forecast and Saturday is difficult to predict as compared with other observations.

IX. REFERENCES

[1] Jin, W.J., Feng, Y., Zhou, N. and Zhong, X., 2014, December. A household electricity consumption algorithm with upper limit. In *2014 International Conference on Wireless Communication and Sensor Network* (pp. 431-434). IEEE.

[2] Hino, H., Shen, H., Murata, N., Wakao, S. and Hayashi, Y., 2013. A versatile clustering method for electricity consumption pattern analysis in households. *IEEE Transactions on Smart Grid*, *4*(2), pp.1048-1057.

[3] Rosin, A., Hõimoja, H., Möller, T. and Lehtla, M., 2010, June. Residential electricity consumption and loads pattern analysis. In *Proceedings of the 2010 Electric Power Quality and Supply Reliability Conference* (pp. 111-116). IEEE.

[4] Imtiaz, A.K., Mariun, N.B., Amran, M.M.R., Saleem, M. and Wahab, N.I.A., 2006, November. Evaluation and Forecasting of Long Term Electricity Consumption Demand for Malaysia by Statistical Analysis. In *2006 IEEE International Power and Energy Conference* (pp. 257-261). IEEE.

[5]Shelke, M. and Thakare, P.D., 2014. Short term load forecasting by using data mining techniques. *Int. J. Sci. Res*, *3*, pp.1363-1367.

[6] https://archive.ics.uci.edu/ml/datasets/