Individual Household Electric Power Consumption Forecasting using Machine Learning Algorithms

Aaditi Parate
Student, M.Sc. (Big Data Analytics)
MIT WPU Pune, India

Sachin Bhoite
Assistant Professor, Computer Science
MIT-WPU, Pune, India

---***************---

Abstract: Electric energy consumption is the actual energy demand made on existing electricity supply. However, the mismanagement of its utilisation can lead to a fall in the supply of electricity. It is therefore imperative that everybody should be concerned about the efficient use of energy in order to reduce consumption [1]. The purposes of this research are to find a model to forecast the electricity consumption in a household and to find the most suitable forecasting period whether it should be in daily, weekly, monthly, or quarterly. The time series data in our study is the individual household electric power consumption [4].To explore and understand the dataset I used line plots for series data and histograms for the data distribution. The data analysis has been performed with the ARIMA (Autoregressive Integrated Moving Average) model.

Keywords: Energy consumption prediction, ARIMA, AR, MA, Python.

----

1. INTRODUCTION

Electricity load forecasting has gained substantial importance nowadays in the modern electrical power management systems with elements of smart greed technology. A reliable forecast of electrical power consumption represents a starting point in policy development and improvement of energy production and distribution. At the level of individual households, the ability to accurately predict consumption of electricity power significantly reduces prices by appropriate systems for energy storage. Therefore, the energy efficient power networks of the future will require entirely new ways of forecasting demand on the scale of individual households [2]. The analysis of a time series used forecasting techniques to identify models from the past data. With the assumption that the information will resemble itself in the future, we can thus forecast future events from the occurred data. There are several techniques of forecasting and these techniques provide forecasting models of different accuracy. The accuracy of the prediction is based on the minimum error of the forecast. The appropriate prediction methods are considered from several factors such as prediction interval, prediction period, characteristic of time series, and size of time series [4].

In this research, we are interest in time series analysis with the popular forecasting technique that I used in this study; ARIMA (Autoregressive Moving Average) I applied this method for detecting patterns and trends of the electric power consumption in the household with real time series period in daily, weekly, monthly, and quarterly. I used Python program for constructing the model.

2. RELATED WORK:

How to Load and Explore Household Electricity Usage Data In this tutorial, you will discover a household power consumption dataset for multi-step time series forecasting and how to better understand the raw data using exploratory analysis.[5]

How to Develop an Autoregression Forecast Model for Household Electricity Consumption In this tutorial, you will discover how to develop and evaluate an autoregression model for multi-step forecasting household power consumption.[6]

Time Series Analysis of Household Electric Consumption with ARIMA and ARMA Models: In this research, we are interest in time series analysis with the most popular method, that is, the Box and Jenkins method. The result model of this method is quite accurate compared to other methods and can be applied to all types of data movement. There were two forecasting techniques that were used in this study; Autoregressive Integrated Moving Average (ARIMA) and Autoregressive Moving Average (ARMA).[1]

3. DATASET DESCRIPTION:

Source:
https://archive.ics.uci.edu/ml/datasets/individual+household+electric+power+consumption

The Household Power Consumption dataset is a multivariate time series dataset that describes the electricity consumption for a single household over four years.

This archive contains 2075259 measurements gathered in a house located in Sceaux (7km of Paris, France) between December 2006 and November 2010 (47 months).

It is a multivariate series comprised of seven variables (besides the date and time); they are:
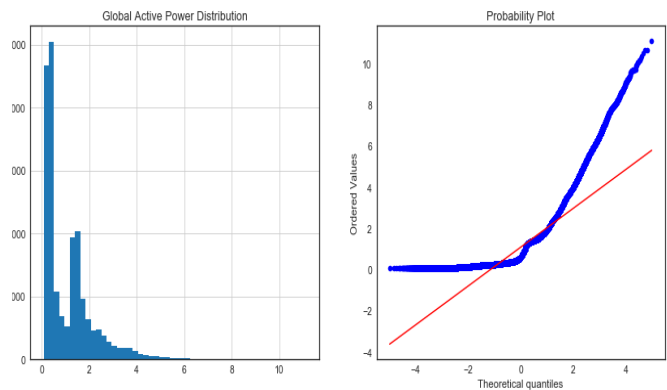
- **global_active_power**: The total active power consumed by the household (kilowatts).
- **global_reactive_power**: The total reactive power consumed by the household (kilowatts).
- **voltage**: Average voltage (volts).
- **global_intensity**: Average current intensity (amps).
- **sub_metering_1**: Active energy for kitchen (watt-hours of active energy).
- **sub_metering_2**: Active energy for laundry (watt-hours of active energy).
- **sub_metering_3**: Active energy for climate control systems (watt-hours of active energy).

4. PRE-PROCESSING:

The dataset contains some missing values in the measurements (nearly 1,25% of the rows). All calendar timestamps are present in the dataset but for some timestamps, the measurement values are missing: a missing value is represented by the absence of value between two consecutive semi-colon attribute separators. For instance, the dataset shows missing values on April 28, 2007.We cannot ignore the missing values in this dataset therefore we cannot delete the missing values. I copied the observation from the same time the day before and implemented this in a function named *fill_missing()* that will take the NumPy array of the data and copy values from exactly 24 hours ago Then we saved cleaned-up version of the dataset to a new file *household_power_consumption.csv*'[3].

5. EXPLORATORY DATA ANALYSIS:

1) Global active power distribution plots:

Normal probability plot also shows the data is far from normally distributed.

2) 1st graph represents the mean global active power by Year.
   2nd graph represents the mean global active power by Quarter.
   3rd graph represents the mean active power by the Month and the 4th graph represents the mean global active power by Day.
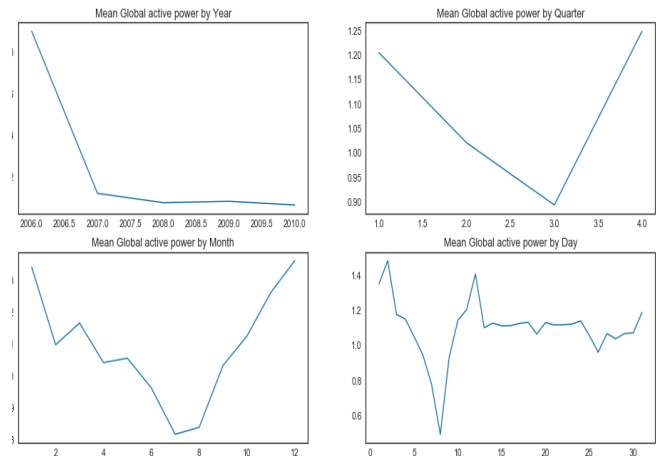

The above plots confirmed our previous discoveries. By year, it was steady. By quarter, the lowest average power consumption was in the 3rd quarter. By month, the lowest average power consumption was in July and August. By day, the lowest average power consumption was around 8th of the month

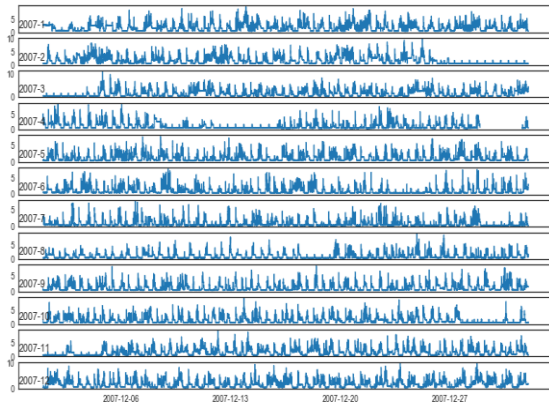3) Global active power by year. This time we removed year 2006
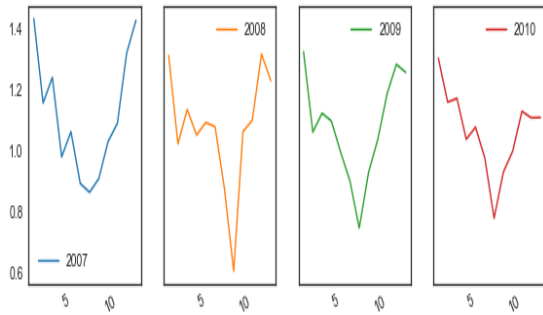
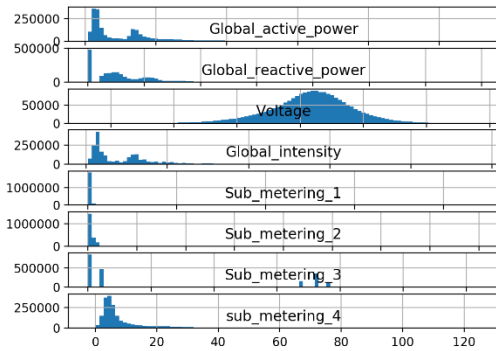The pattern is similar every year from 2007 to 2010.

4)   Line plot of Active power for years:
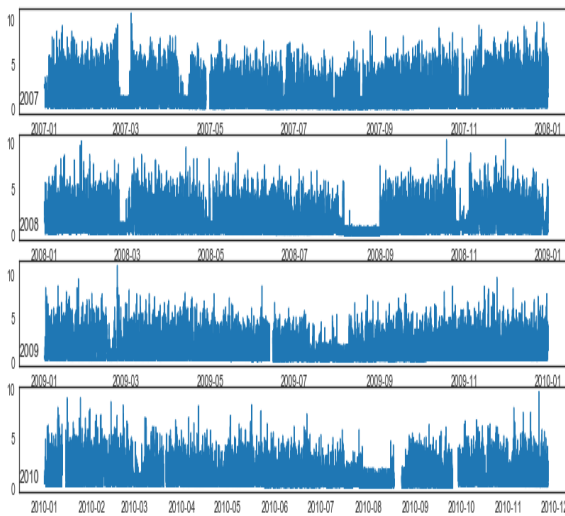
5)   Line plots for Active Power for all months in a year:

6)   Histogram plots for Each Variable in the Power Consumption Dataset

[4].

5. RESULT AND DISCUSSION:

I developed an autoregression model for univariate series of daily power consumption. I used the Statsmodels library that provides multiple ways of developing an AR model, such as using the AR, ARMA, ARIMA, and SARIMAX classes.

I use the ARIMA implementation as it allows for easy expandability into differencing and moving average.

First, the history data comprised of weeks of prior observations is converted into a univariate time series of daily power consumption. I specified an AR (7) model, which in ARIMA notation is ARIMA(7,0,0).

Running the example first prints the performance of the AR (7) model on the test dataset.

We can see that the model achieves the overall RMSE of about 381 kilowatts.

This model has skill when compared to naive forecast models, such as a model that forecasts the week ahead using observations from the same time one year ago that achieved an overall RMSE of about 465 kilowatts.

A line plot of the forecast is also created, showing the RMSE in kilowatts for each of the seven lead times of the forecast. We can see an interesting pattern. We might expect that earlier lead times are easier to forecast than later lead times, as the error at each successive lead time compounds.

Instead, we see that Friday (lead time +6) is the easiest to forecast and Saturday (lead time +7) is the most challenging to forecast. We can also see that the remaining lead times all have a similar error in the mid- to high-300 kilowatt range. [3]
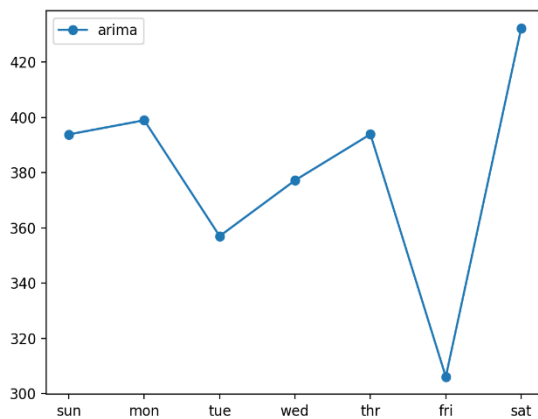
an-autoregression-forecast-model-for-household-electricity-consumption/#

[4] Pasapitch Chujai*, Nittaya Kerdprasop, and Kittisak Kerdprasop"Time Series Analysis of Household Electric Consumption with ARIMA and ARMA Models"Proceedings of the International MultiConference of Engineers and Computer Scientists 2013 Vol I, IMECS 2013, March 13 - 15, 2013, Hong Kong

 [5]   https://machinelearningmastery.com/how-to-load-and-explore-household-electricity-usage-data/

[6] https://machinelearningmastery.com/how-to-develop-an-autoregression-forecast-model-for-household-electricity-consumption/#:

6. CONCLUSION:

Many researchers wrote about the ARIMA model, AR model, MA model and also worked with these models on the consumption of electricity. I find the ARIMA model easy as compared to the other models and also the ARIMA model gives a better accuracy then the other models.

Therefore in this research I used the ARIMA model on the individual household electricity consumption dataset. Then, chose the suitable forecasting method and identified the most suitable forecasting period by considering the smallest values of RMSE. In this data set the consumption of electricity is more in the month of December and regular during the other time period of the year.The results showed that the ARIMA model represent the most suitable forecasting periods in monthly and quarterly, daily and weekly.

The ARIMA model result: arima: [465.902] 428.0, 448.9, 395.8, 522.3, 450.4, 380.5, 597.9Here the RMSE is 465.902

7. REFERENCES:

 [1] c Kamunda A Study on Efficient Energy Use for Household Appliances in Malaw

 [2] Naser Farag Abed1 and Milan M.Milosavljevic 1,2 1 Singidunum University Belgrade, 11000, Serbia 2 School of Electrical Engineering, Belgrade University, Belgrade, 11000, Serbia"Single Home Electricity Power Consumption Forecast Using Neural Networks Model"IJISET - International Journal of Innovative Science, Engineering & Technology, Vol. 3 Issue 1, January 2016. ISSN 2348 – 7968

[3]https://machinelearningmastery.com/how-to-develop-