UNIVERSITY AT ALBANY

State University of New York

TOPIC IN BUSINESS ANALYTICS AND TEXT MINING - BITM 609

Electricity Consumption Forecasting

Author:
Krishna Kaushik Reddy, Karri
Harsh, Patel
Sai Amruth, Bathineni
Naga Satya Alaikya, Kalidindi

Instructor:
Pranay R. Jinna

May 11, 2021

CONTENTS

LIST OF FIGURES

LIST OF TABLES

Abstract—**In 2019 the energy consumption of the world is 13.9 billion toes. The world with a population of about 7.7 billion consumes energy of 58 kWh on an average per day per person. Aren't the numbers huge? Electricity consumption in the world is increasing proportionally with the raise in population. More population indicates more energy consumption. Electricity is now a necessity. The question of how to manage the electricity generated in a city or state or a county raises more questions. Will allocating a particular amount of energy to each city in the state solve the problem of energy management? Yes, it does solve the problem in a lighter note i.e allocating a particular amount of energy to a city will only be a good practice if we know beforehand the amount of energy that will be required by the city in a particular month. To achieve this forecasting techniques should be used. But even a city is divided into various counties. Few counties in the city might have a greater number of industries where as few might have only households and a few might have a very large international airports. So deep diving into how much energy does a county require based on the type of area i.e residential or industrial may help solving the problem of energy management to a greater extent.**

***Index Terms*—Active power, Reactive power, Voltage, Regression Forecasting, Alteryx, Tableau.**

I. Introduction

We as a Data Science team in an Energy Management company would like to build a model that would forecast the energy required. The application we are planning to build is similar to the weather forecasting. In weather forecasting attributes such as temperature, humidity, geographic location of the continent, etc are considered to make the accurate forecasting. In case of energy prediction more focus is put on the type of area i.e residential or industrial. Along with the type of area Gravity is also planning on observing various trends and seasonality patterns to forecast the energy required.

II. OVERVIEW OF THE DATA SET

We choose the data set that contains the energy consumption of a household community(residential area).The data ranged from 2006-2010. The units are in kilo watts. The data is per minute data i.e how much energy is consumed by the community every minute. The attributes in the data set are Global active power, Global reactive power, Voltage, Sub metering 1, Sub metering 2, Sub metering 3. The description of the features are as follows: 1.Global active power : It is the power consumed by the appliances that are not mapped to the sub meters.

2.Global reactive power: It is the power that is allocated but not utilised.This is basically the imaginary power consumption.

3.Voltage : Average voltage of the community.

4.Sub_metering_1 : The energy consumed in the Kitchens.

5.Sub_metering_2 : The energy consumed in the Laundry.

6.Sub_ metering 3 : The energy consumed in the Air Conditioners and Heaters.

III. EXPLORATORY DATA ANALYSIS

Sticking with the minute to minute data will miss lead Gravity from observing the trend patterns and Seasonality and this is because the data set ranges from 2006 to 2010 and considering data of each minute in each day will make the graphs look more compact. To get rid of this the technique of grouping is used. Initially the grouping is done on date basis.i.e all the minutes data in 2006 December 12 is added up to a single value. Later the grouping is done on the monthly and yearly basis. For Regression based analysis the grouped data is considered where as the minute to minute data is used for the forecasting analysis of Sub_metering_1(Kitchen), Sub_metering_2(Laundry) and Sub_metering_3(Air conditioner) utilisation.

To begin with Gravity first tried to understand the difference between total energy utilised and the energy not used by the community. To get the total energy utilised we just summed up all the values in the columns i.e active power + reactive power + meter readings where as energy not used is simply the reactive power.The behaviour of the total energy consumed is in Fig.1 The interesting trend observed here is that the

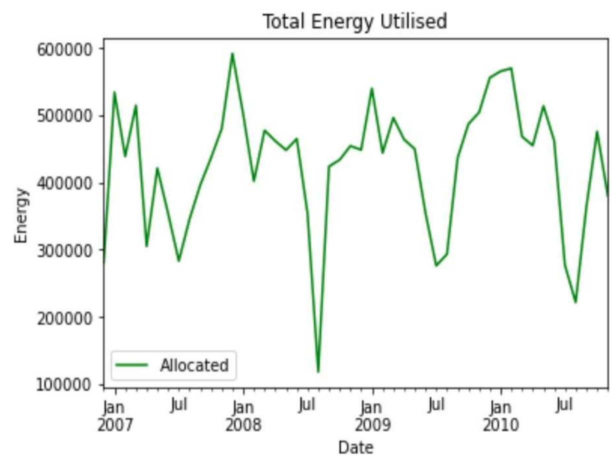

Figure 1. Total Energy Utilised

energy consumption in January is high compared to the energy consumption in July and this same pattern followed for the years of 2008, 2009 and 2010. The point to be noted here is that the month of July and it neighbours can be allocated a less quantity of Electricity Compared to the beginning and the ending of the year. Now this can be achieved only if we know how energy that is allocated is going in waste. To observe this the plot for reactive power is made(Fig.2) The high spike in July 2009 from Fig.2 is explained from the low power consumption in the month of July 2009 from Fig.1. There might be an external factor for the least consumption of electricity in the month of July 2009. Imagining that the spike in July 2009 from Fig.2 is not so high we can see an increasing trend in the month of July from Fig.2 i.e the power not being utilised in the month of July is increasing every year from 2007 to 2010. The main aim is to forecast the energy

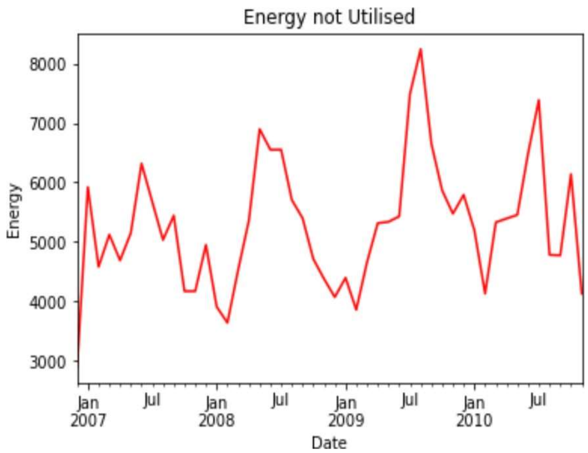

Figure 2. Total Energy Utilised

that will not be utilised. And the aim will be to reduce the energy wastage.

IV. REGRESSION BASED FORECASTING

Keeping in mind the concepts of trend and seasonality smoothening technique is applied on the Global_reactive_power. The data till 2009-06-01 is considered as training data whereas data from 2009-06-01 to 2010-11-01 is considered as validation data. From Fig.2 it can be observed that the energy not consumed follows an upward trend with additive seasonality. So keeping in mind about the upward trend and additive seasonality an Exponential Smoothing Algorithm is applied to the data and the exponential smoothing method is observed to perform well from Fig.3. From Fig.3 it is observed that the model is doing a good

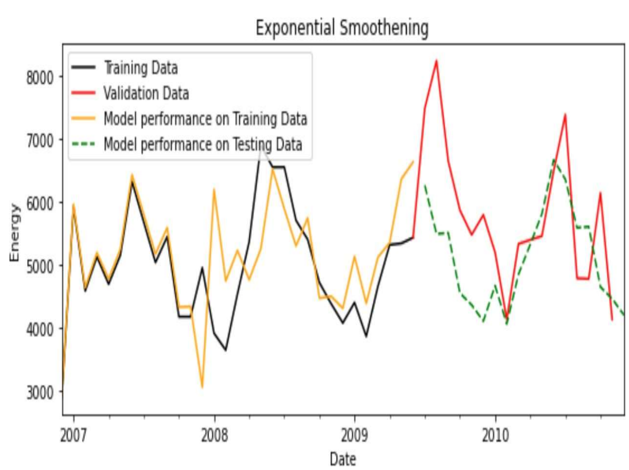

Figure 3. Exponential Smoothening on Energy not Utilised

job in forecasting the data but the gap in the month of July 2009 between predicted value and the original value is high because of the unique behavior of low power consumption in month of July 2009. In rest of the years the predictions are pretty much accurate and it is clearly evident from the Fig.4. So using this exponential smoothing technique Gravity can forecast the amount of Energy that will not be utilised. For suppose the State Government planned on Allocating 100 units of Energy to the county and with the Exponential model if the forecasted value for energy not being utilised in 15 units. Then the government can decide on allocating any power between 100 units and 85units(100-15).

Cooking, Laundry and Heaters(cold or Warm) are basic needs. What Gravity would like to observe is the energy pattern in the other appliances that are not connected to any of the sub meters. This is important because an increase or decrease in the energy consumption of the other appliances will also impact the amount of energy that is to be allocated to the residential areas. Many regression techniques incorporating trend and seasonality are performed in forecasting the amount of energy that will be utilised by the secondary appliances.But, exponential smoothing technique is found to perform well on the prediction and it can be observed in the Fig.4.

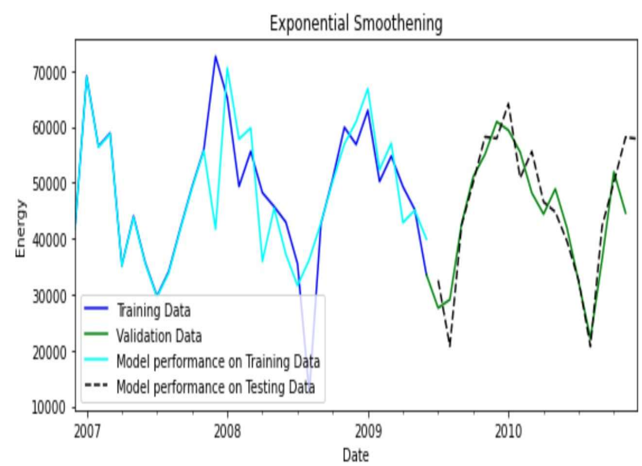

Figure 4. Exponential Smoothening on Active power

From Fig.4 it is clearly evident that the predictions are pretty accurate. Also, as discussed Sub metering 3 is the meter reading for both Air conditioning and Heaters. From Fig.5 it can be observed that the consumption of sub metering 3 appliances is high in months around January and the consumption is low in the months around June-July. Also, there consumption of energy in 2007 January is lesser than the consumption in 2008 January the same pattern can be observed in the following years as well. So, using regression techniques if we are able to capture this pattern and forecast the future consumption in various months of the year then the a lot of energy can be saved i.e for example, if the consumption of sub metering 3 appliances is 100 units in the months around January and look at the 100 units consumption if the Government tries to allocate 110 units of power in the months around July then there will be a lot of power wastage. To prevent it, Gravity applied a regression model with trend and Seasonality

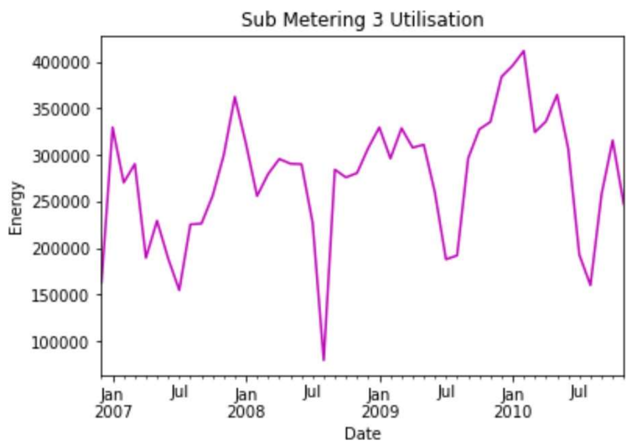

Figure 5. Sub metering 3 Consumption

on the sub metering 3 consumption and the performance of the model on the meter readings can be observed from Fig.6.

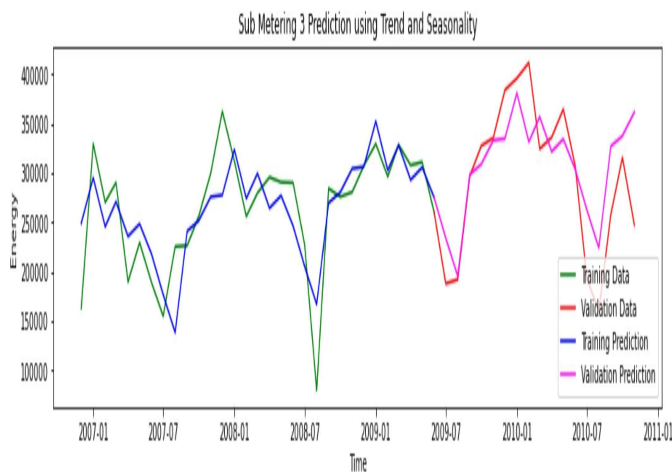

Figure 6. Sub metering 3 Regression prediction

We all know the fact that higher the r square, the better the model. The regression model applied on sub metering 3 has an r square of 0.611, which is pretty decent. So going forward with this regression model the trends and patterns are said to be captured well for sub metering 3 appliances.

V. Alteryx

Alteryx is a software that is used for data science and analytics. The software is designed to make advanced analytics accessible to any data worker. Gravity tried to forecast the active power that will be required by the household community using this software. The workflow in Fig.7 shows the compared models in Alteryx. Comparing the two models available in the time-series toolbox ARIMA and ETS(Exponential Smoothing). To see which model works better we pass the combined output to the TS Compare which compares the models and will give the different accuracy errors such as

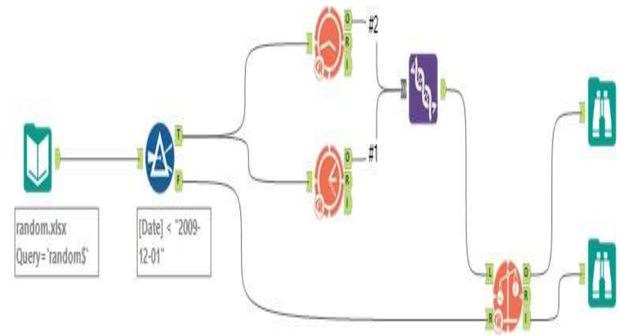

Figure 7. Comparing two models using Alteryx

the ME,MAE,etc for both models and the output is shown in the Fig 8.

Record	Model	ME	RMSE	MAE	MPE	MAPE	MASE	NA
1	Arima1	-2,452.9438	6,411.7041	5,358.4765	-4.2762	12.4981	0.5628	[Null]
2	ETS1	-1,672.9987	5,218.5348	4,251.6946	-4.3103	9.5062	0.4465	[Null]

Figure 8. Accuracy Error of the two models

From Fig.8 it is clear that the ARIMA models outweighs the ETS so we used the ARIMA to forecast, and the workflow of the chosen model can be observed from Fig.9

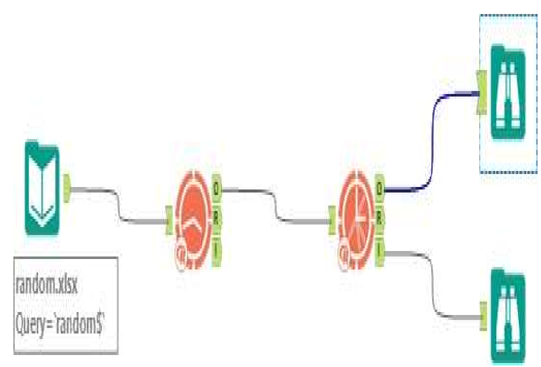

Figure 9. ARIMA model Workflow

In the workflow(Fig.9) the ARIMA model is used to forecast the 12-month energy consumption and the output from the model is passed to TS Forecast tool which basically outputs the forecast values which are shown in Fig.10 along with the confidence intervals. From the output table in Fig.10 it can be said that the ARIMA model in Alteryx is doing a good job as well in forecasting the 1 year active power consumption.

Sub_Period	forecast	forecast_high_95	forecast_high_80	forecast_low_80	forecast_low_95
1	58,089.338589	71,328.847575	66,746.188206	49,432.488973	44,849.829603
2	61,848.998708	75,088.507694	70,505.848324	53,192.149092	48,609.489722
3	51,840.944173	65,080.453158	60,497.793789	43,184.094556	38,601.435187
4	52,722.75391	65,962.262895	61,379.603526	44,065.904293	39,483.244924
5	47,765.284694	61,004.79368	56,422.13431	39,108.435077	34,525.775708
6	46,345.445217	59,584.954202	55,002.294833	37,688.5956	33,105.936231
7	36,114.895645	49,354.40463	44,771.745261	27,458.046028	22,875.386659
8	29,008.142216	42,247.651202	37,664.991833	20,351.2926	15,768.633231
9	26,926.090775	40,165.599761	35,582.940391	18,269.241158	13,686.581789
10	40,679.894582	53,919.403567	49,336.744198	32,023.044965	27,440.385596
11	51,349.061277	64,588.570263	60,005.910893	42,692.211661	38,109.552291
12	51,839.178142	65,078.687128	60,496.027759	43,182.328526	38,599.669157

Figure 10. 12 Months Forecast

The same methodology can be followed in forecasting the values for the data variables such as Global_reactive_power and the sub metering readings.

VI. TABLEAU

Tableau being a data visualization tool helped visualize the data very well with filters and plots. The data can be filtered, re-arranged and sliced as required like an hour, minute in the Date Time feature. We have tried forecasting every hour of data for a day in every month using the 15-day time frame(Fig.11) Fig.11 shows 1-day July 16th forecasted energy consumed by the sub-metering 3 for every hour, by considering 15 days i.e; July 1st to July 15th. The filters to the right are Day, Month, Year, and the forecasted indicator Actual-Dark blue and Estimated-light blue.

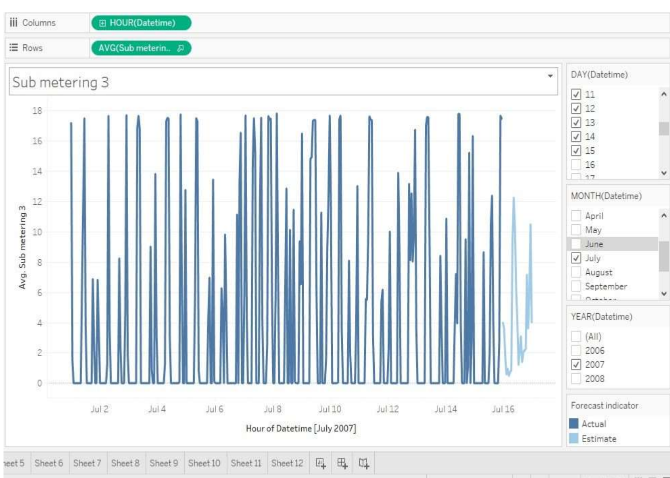

Figure 11. Developing a model using Tableau

Fig.12 shows the comparison of the forecasted values in fig(11) of July 16th- orange and the original values- blue. This shows the model captured the seasonality accurately

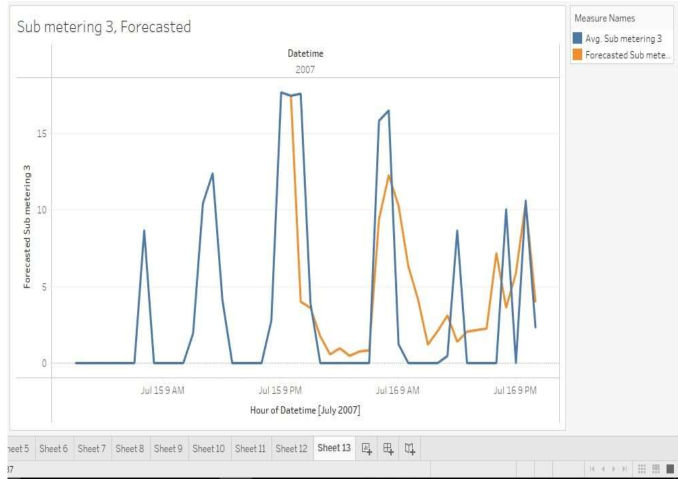

Figure 12. Forecasted Values

Fig.13 shows the quality metric of the applied model in Fig.11. This tends to forecast well with the MASE of 0.76. Tableau has a lot of features for forecasting, where we can


Describe Forecast

Summary | Models

All forecasts were computed using exponential smoothing.

Avg. Sub metering 3

Model			Quality Metrics					Smoothing Coefficients		
Level	Trend	Season	RMSE	MAE	MASE	MAPE	AIC	Alpha	Beta	Gamma
Additive	None	Additive	5.80	4.04	0.76	91.5%	1,070	0.000	0.000	0.144

Copy to Clipboard | Learn more about the forecast models | Close

Figure 13. Tableau metrics

automate the process or custom the process by adding or changing the trend and seasonality based on the data. Though tableau is great for visualizing, we didn't find it helpful to forecast minute data as the forecasted values are very uncertain and the MASE of the models obtained are very high, i.e; models aren't working good.

VII. CONCLUSION

With the help of smoothening and regression techniques the amount of energy that will not be utilised can be forecasted and new innovations in reducing the amount of energy that is not being utilised can be brought in to picture. Also with the model developed by Alteryx the amount of energy that will be utilised by the appliances that are not connected to sub meters 1,2 and 3 can be forecasted. This is important because the percentage of usage of these secondary appliances will impact the total energy consumed.