Forecast energy usage of households

CT/DT Number: DT20184336603
Contestant Name: Ayush Agarwal
College Name: D Y Patil Institute of Technology Pimpri

Background

The project is to Forecast energy usage of households, That is to forecast the electricity consumption of top 3 households with highest number of samples on an hourly basis based on the previous usage pattern. There is a csv file named Power-Networks-LCL.csv that contains six column (LcLid, stdorToU, DateTime, KWh, Acorn, Acorn_grouped). The category of the problem is time series analysis.

Time series analysis is a way to analyse time series(Time series is a series of data points taken in the timely order) and predict or provide the statics, Our goal here is to forecast consumption of electricity of each household and determine the top households on hourly basis.

The data is of the year 2012, 2013 and 2014 consisting of entry of electricity consumption of the household every half hour.

Understanding

The analysis of problem statement tells to give the top 3 households that consume that consume highest electricity at that hour. Therefore we will use Time Series Analysis to predict the output of electricity consumption of households for the next 24 hours series.(Because 1 day contains 24 hours) based on that we will find the top 3 households at each of the 24 hours.

The csv file contains six columns :

1. LcLid : It is HouseHold number
2. stdorToU : standard or dynamic time of use
3. DateTime : Reading after every Half hour
4. KWh : Meter readings
5. Acorn : Acorn is a segmentation tool which categorises the UK's population into demographic types
6. Acorn_grouped : Acorn group type

It also describes that there is a correlation of previous data and the future predicted data, the most import column in KWh which will be used to provide a forecast of the electric consumption of household in terms of KWh.

Scope

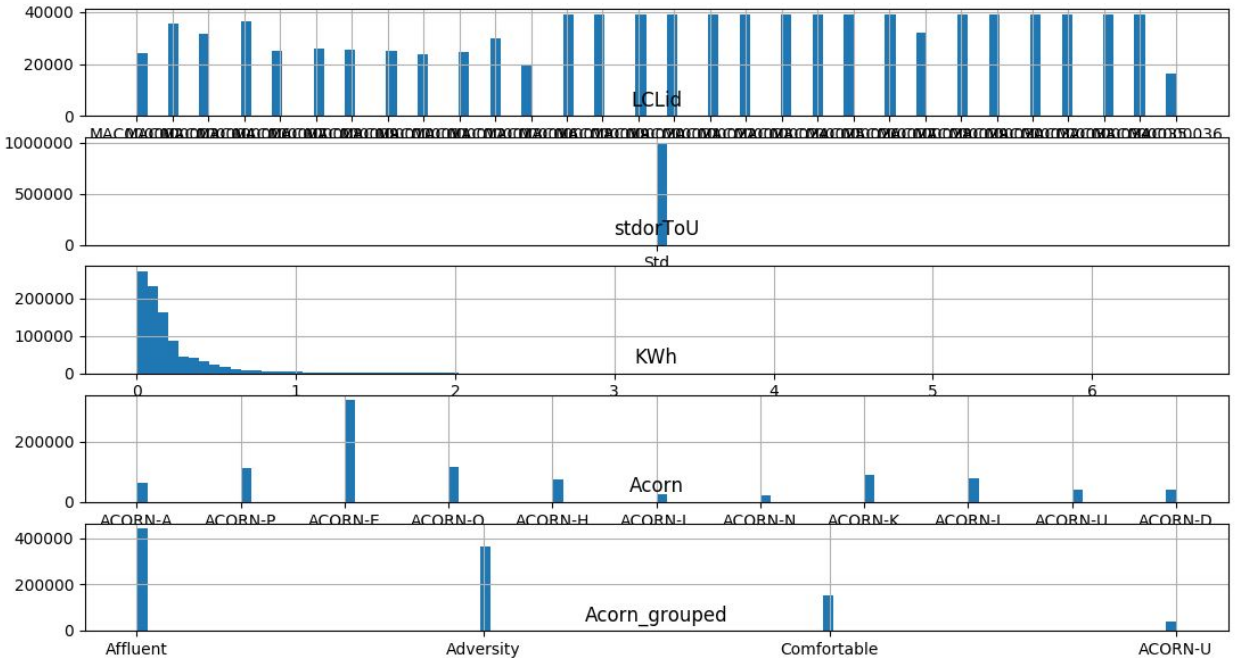

Fig. 1Analysis of DataSet

The above image shows How the data is distributed in the dataset given, as we can see that all of the data has stdorTou as standard and further analysis show that each household has its unique Acorn And Acorn_group.

So the columns that are of interest are the LcLid,DataTime,KWh. And also as the electricity consumption depend upon day(Whether it is week day of not) we have consider that days can affect the consumption.

We can also see that the Kwh column is screwed distributed, therefore we need fix or reduce it by taking log transform.

Out of Scope

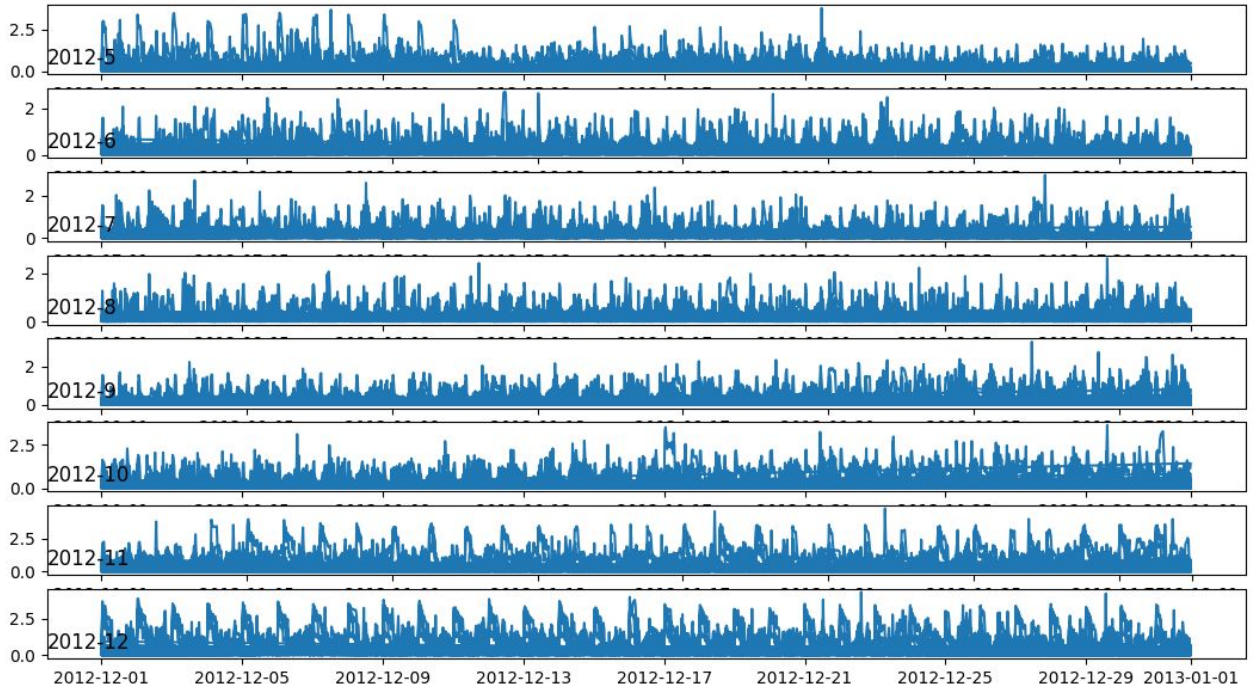

Fig 2 Monthly Kwh power consumption

The Analysis of electricity consumption shows that the week has more effect then the months, so the effect of months are not taken into consideration. And also as shown in Fig 1 all the household has standard time of use and therefore considering dynamic use is not in our scope.

Assumptions

General Assumptions:

- The most basic assumption is that the meter reading Kwh does not have wrong values.
- The second assumption is that sessions does not have any significant effect on the electricity consumption.

Technical Assumptions:

- As we are going to use auto regression the assumption is that the observations at previous time steps are useful to predict the value at the next time step i.e it is correlational.
- Homoscedasticity
- Historic timepoints dictate behaviour of present timepoints.

Solution Approach

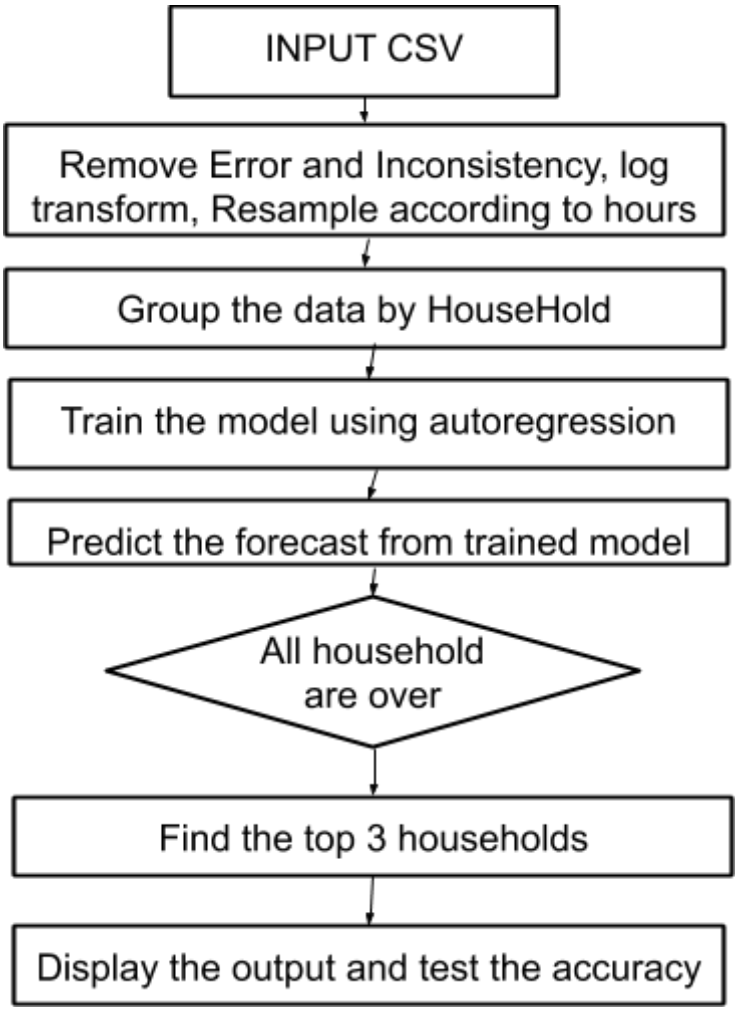

Fig 3 FlowChart

The model that is proposed to implement is the ARIMA(**A**uto **R**egressive **I**ntegrated **M**oving **A**verage) as it takes into account that the past value of time series can affect the future prediction. The features of ARIMA is that it is integrated and moving average, ARIMA uses autoregression which is a process of regressing a variable on past values of itself.

This model is proposed because our dataset depends on our past consumption of electricity and the weekdays affects our data the AMIRA is best suited for such type of time series analysis problem.

Implementation Framework

The implementation Steps are:

1. Input and error removal and log transform:
    a. Input is taken from the csv as our data consist of reading of every half hour and our prediction should be on an hourly basis we resample our data into hourly format
    b. Then there are few data that are missing we either try to fill the missing data or remove the tuple if it can't be fixed.
    c. we should also make log transformation of KWh column to reduce its skewness.
2. Group data by households:
    a. As the dataset consists of multiple households and we have to predic electric consumption of each household we group the dataset of each individual household.
3. Train the model and forecast the electric consumption
    a. The model we are joining to use is ARIMA(**A**uto **R**egressive **I**ntegrated **M**oving **A**verage) we will git it the historical data to train the model and then predict the future forecast.
4. Repeat step 3 for all the households.
5. From the predicted data find the top 3 households on hourly basis and display the output
    a. Now as all the model are run and we have predicted the consumption of every house then we can easily find the top 3 households .
    b. Display the output and accuracy.

Software and Hardware Used:

1. Linux operating system
2. Python
3. matplotlib
4. pandas
5. sklearn
6. statsmodels
7. 2GB memory
8. Intel Dual Core Processor

Solution Submission

This is the link for the github repository
https://github.com/ayushbansal323/HumanAIn

Appendix

The Github repository contains the scheenshot and output of the problem, please refer to the readme document.

References

1. The acorn structure , https://upload.wikimedia.org/wikipedia/commons/8/8c/CACI%27s_Acorn_Structure_%28Uploaded_June_2014%29.pdf.
2. Autoregression model , https://en.wikipedia.org/wiki/Autoregressive_model
3. AMIRA, https://people.duke.edu/~rnau/411arim.htm
4. matplotlib, https://matplotlib.org/
5. Improve time series forecasting, https://www.analyticsvidhya.com/blog/2018/02/time-series-forecasting-methods/
6. Forecasting electricity consumption: A comparison of regression analysis,neural networks and least squares support vector machines,M. Cengiz Taplamacioglu,https://www.researchgate.net/publication/270006584_Forecasting_electricity_consumption_A_comparison_of_regression_analysis_neural_networks_and_least_squares_support_vector_machines