1. Background

With the development of smart electricity metering technologies, huge amount of consumption data can be retrieved on daily and hourly basis. Energy consumption forecasting facilitates electricity demand management and utilities load planning. Most of the prior researches are focused on commercial customers or residential building-level energy consumption, or use behavioral and occupancy sensor data to experiment on individual household's electrical consumption.

we will look for an approach to forecast the hourly electricity loads of individual consumers for 24 hours by taking into account historical electricity consumption and the household's behavioral data

2. Your Understanding

Forecasting usage provides customers with the possibility of linking current usage behaviors with future costs. Therefore, customers may benefit from forecasting solutions through greater understanding of their own energy consumption and their future projections, allowing them to better manage the costs of their usage. By making energy consumption and future projections more transparent, it would be easy to understand how much we are actually using and how it would affect our budget in the future.

3. Scope

Analyze the meter reading data and forecast the top 3 households based on hourly time readings.

4. Out of Scope

1) Issue with data
2) Invalid meter readings

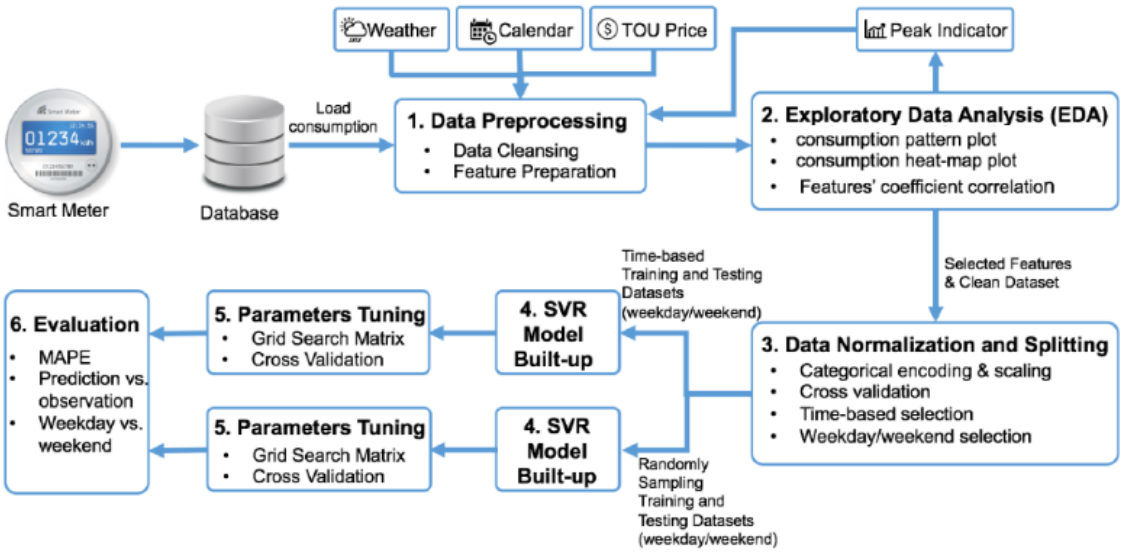
5.Assumptions

1) Download the dataset
2) Inspect the data file.
3) Remove redundant sets to ensure it has processed information. (may have missing values)

6.Solution Approach

1) High Level Solution Approach

Load Consumption data->Data preprocessing ->Data analysis->Evaluation based on parameters->obtain model->based on trained datasets-> finally normalization of data



2) Models/ Algorithms proposed

Data analysis using Pandas-Python

1.) We can start-off by loading the data file as a Pandas Data Frame and summarize the loaded data.

2.) We can use the read_csv() function to load the data.

```
from pandas import read_csv

from matplotlib import pyplot

# load the new file

dfrom matplotlib import pyplot

# load the new file

dataset = read_csv('our csv', header=0, infer_datetime_format=True, parse_dates=['datetime'], index_col=['datetime'])
```

```
# plot active power for each hour
days= ['1', '2', '3', '4']
pyplot.figure()
for i in range(len(hours)):
        # prepare subplot
        ax = pyplot.subplot(len(hours), 1, i+1)
        # determine the hours to plot
        hour = hours[i]
        # get all observations for the hours
        result = dataset[str(day)]
        # plot the active power for the hours
        result['Global_active_power'].hist(bins=100)
        # zoom in on the distribution
        ax.set_xlim(0, 5)
        # add a title to the subplot
        pyplot.title(str(year), y=0, loc='right')
pyplot.show()
```

Running the example creates a single plot with four figures, one for each of the 4 days

There is a long tail on the distribution to higher kilowatt values. It might open the door to notions of discretizing the data and separating it into peak 1, peak 2 or long tail. These groups or clusters for usage on a day or hour may be helpful in developing a predictive model.

It is possible that the identified groups may vary over the seasons of the year.

so we have to take months into consideration

```
result = dataset[month]
```

7. Implementation Framework

Hardware and Software Details:

General Implementation Approach

Using Python Pandas data frame

Software:

Anaconda

Hardware:

CPU: 2 x 64-bit 2.8 GHz 8.00 GT/s CPUs.

RAM: 32 GB (or 16 GB of 1600 MHz DDR3 RAM)

Storage: 300 GB. ...

Internet access to download the files from Anaconda Cloud or a USB drive containing all of the files you need with alternate instructions for air gapped installations.

8. Appendix

The data is only for a single household, but perhaps effective modeling approaches could be generalized across to similar households.

Perhaps the most useful framing of the dataset is to forecast an interval of future active power consumption.

Four examples include:

Forecast hourly consumption for the next day.

Forecast daily consumption for the next week.

Forecast daily consumption for the next month.

Forecast monthly consumption for the next year.

There are different modelling methods available to explore the problem(machine learning -supervised algorithm, nonlinear and ensemble methods could be explored for fitting and evaluation of these models.

Specifically we gone through The household power consumption dataset that describes electricity usage for a house over daily hour basis

How to explore and understand the dataset using a suite of line plots for the series data and histogram for the data distribution using pandas feature.

10. References

https://towardsdatascience.com/analyzing-time-series-data-in-pandas-be3887fdd621