B.Tech. Project Report (CSPE-40)

on

Smart Metering Analytics for Checking Energy Consumption

BY

HIMANSHU KUMAR (11710146)

POOJA (11710154)

HARSHIT GOYAL (11720001)

Group No. : 21

Under the Supervision of

Dr. Priyanka Ahlawat , Asst. Professor

DEPARTMENT OF COMPUTER ENGINEERING

NATIONAL INSTITUTE OF TECHNOLOGY

KURUKSHETRA - 136119, HARYANA (INDIA)

April-May,2021

CERTIFICATE

We hereby certify that the work which is being presented in this B.Tech Project Report (**CSPE40)** **r**eport entitled **" Smart Metering Analytics For Checking Energy Consumption ",** in partial fulfillment of the requirements for the award of the **Bachelor of Technology in Computer Engineering** is an authentic record of our own work carried out during a period from January, 2021 to April, 2021 under the supervision of Dr. Priyanka Ahlawat, Assistant Professor, Computer Engineering Department.

The matter presented in this project report has not been submitted for the award of any other degree elsewhere.

Signature of Candidate

HIMANSHU KUMAR (11710146)

POOJA (11710154)

HARSHIT GOYAL (11720001)

This is to certify that the above statement made by the candidates is correct to the best of my knowledge.

Date:

Signature of Supervisor

Dr. Priyanka Ahlawat

Asst. Professor

TABLE OF CONTENTS

ABSTRACT

Smart grids have evolved as the next generation power Grid systems that transform the traditional/new ways of functioning of present electrical grids. Advanced metering infrastructure (AMI) is the key component in smart grids. A typical AMI System comprises of systems and networks, that are responsible for the collection and analyzing of the data received from installed smart meters. In addition, AMI also monitors and manages the different applications related with power and services based on the data collected from smart meters. Thus, Advanced Metering Infrastructure (AMI) plays a significant role in the smooth functioning of smart grids. In this project, a variety of smart meter machine learning analytics algorithms have been proposed and implemented, mainly in the smart grid literature, to predict home electricity consumption enabling accurate planning and forecasting, extracting consumption profiles of various houses in the apartment to provide personalized energy saving tips to the consumers, and design targeted engagement programs to clusters of similar consumers working in this field. However, the research focus has been on the insight that can be obtained from the data rather than performance and programmer effort. At first, we discuss the main features of AMIs, the deployment scenario of smart grids and identify the relationship between smart grid and AMI. Then, we explore the collected data and understand how machine learning algorithms can be applied for the analytics purpose. After that, the analytics was performed using different machine learning models. At last, the comparison between the models is done according to their error and accuracy percentage which gives us the best model to for the prediction.

The main content of this project is covered in seven sections. First section provides the introduction about the topic of our project research and the problem that we are trying to overcome. The second section surrounds the main theoretical concepts and motivation for this research. We thoroughly have discussed the other similar research endeavors' in this section and their respective linkages to our research. The third and fourth section describes our methodology and practical implementation steps respectively. The fifth section provides the details of the data analysis and the observations/results that we produced. Sixth section presents a critical analysis of our approach and results. The last section provides a summary in a nutshell of what we have achieved from our implementation efforts and how it might be used in the real world applications.

INTRODUCTION

In the modern world, we have seen a phenomenal and huge increase of human depending on Information and Communication Technology (ICT), especially in metropolitan areas. ICT enabled products and services have transformed the way how humans are living life on this planet. Every time we use the services which contains some king of digitalism , we produce digital traces that can be recorded and analyzed. Data Analytics refers to analyzing these digital traces of human activities and deriving some meaningful information that can somehow serve as an important asset in our daily life. Businesses are not only using this vast amount of data for business purposes but also for enabling growth in revenue. To support this transfiguration, we have seen a rapid development in machine learning department. Industry giants such as Google and Yahoo have open sourced technologies and tools to facilitate these advancements. In this project, we have tried to formulate models for an end-to-end Data analytics platform based on these technologies that can ingest data from different sources, process it in an efficient manner, dig the data to generate insights based on business logic and then present the information using interactive manner using graphical visualizations. This practical part of the research includes the development as well as implementation of the mentioned topic to perform the analyses on real life use cases and generate useful insights and thereby deriving meaningful information. It aims to make metropolitan life as a functional unit for improving energy efficiency. The use of pervasive constantly encountered computing is driving the smart energy solutions. The smart energy devices as part of this lifestyle generate huge amount of data. This data needs to be instantaneously transferred, stored, analyzed and visualized for knowledge discovery and improvements of services. The data from smart energy devices was analyzed to detect the usage patterns and classify buildings on the basis of energy efficiency. Evaluation of some prediction models for energy consumption of household appliances like Air Conditioner AC's, was also included in the scope of project. The insights generated from these models can also help in spreading awareness among the consumers about ongoing behavioral changes from which the society and the individuals can benefit, which in turn will lead to the development of the society. In a nutshell, this project focuses on providing a solution for collecting, storing, analyzing and visualizing data generated by smart energy devices like smart meters for generating insights about energy consumption patterns and seeing the performance of different apartment building units in terms of energy efficiency. The data analysis part of our

project provides the models for knowledge discovery that can be used to improve energy efficiency at both producer and consumer ends since AMI is a two-way communication.
However, within the scope of this project we discuss its use for energy usage patterns, detection and efficiencies in daily life.

MOTIVATION

Problem Statement:

Using Energy in an efficient way can help to satisfy production of energy to meet growth in demand which in turn would help reduce CO2 emissions. In order to achieve this goal we need to understand and improve the energy efficiency at both producer and consumer ends in an AMI system. Information and Communication Technology (ICT) enabled smart energy grids like smart meters and devices are being set up globally to measure energy consumption and improve energy efficiency, especially in metropolitan areas. These smart meter produce huge volumes of data which is in different formats. This data needs to be collected, stored and analyzed for the purpose of knowledge discovery and supporting energy efficiency. This analysis need to be presented in an interactive visualization and the extracted insights can be used further for deriving "information" which will help in effective understanding. But the challenge is to analyze this data in real time scale because of which ,then with the time, volume and veracity of data both and also scope of analysis is expected to increase their level. So a proper and efficient machine learning model needs to build and work upon.

There is also need to implement the proposed model and test it with real life data from smart meter devices. The proposed platform can be scaled according to data requirements and additional functional components can be integrated as per the scope of analysis. The data analysis within our project also provides advance analytics models to extract the information based on energy efficiency use cases from data warehouses.

Benefits

- In this project, the data collected by smart meters is onlY the primary requirement and no other hardware of personal information is required.
- The libraries in python contains in-built functions which automatically performs mining tasks such as pre-processing, pruning, normalize and many more.

Future of Smart Metering Devices

- Day by day, Smart metering is improving by new and better connectivity innovations, especially with Internet of Things (IoT). With the help of the smart meters tied to the mechanisms of the apartments that control the resources they are measuring, we can turn our decisions over to automation backed by machine learning algorithms.

How can smart metering help businesses in their growth?

- The foremost advantage of smart metering is the precious knowledge of our business's energy and other utility usage. Most of the smart meters works in such a way that they tie directly into an interface that gives you an exact information or readout of your electric usage, breaking it down by the time of day, day of the week and month. Having such kind of information within our reach can help us make better and effective decisions about our usage and keep up-to-date records of our resource usage.
- Launching initiatives to reduce energy consumption at a given point of time.
- The immediate opportunities of smart metering lie in the areas of data access, energy efficiency, billing transparency, performance and compliance. Data Analytics and incoming technologies unfold more exciting possibilities into the coming future.
- Internet of Things (IoT) and big data analytics together will pave the way for multiple devices to be connected together thus making use of energy in an efficient manner. It can improve grid intelligence capabilities and provide projection models that will combine weather forecasts and historical data to integrate the supply of renewable resources for E&U companies. Such innovative ideas might take customer experience and interest to new levels.

Solution Pros:

Using machine learning techniques to help utility companies to analyze all aspects of energy production and consumption, Consumers are able to predict the market demands, understand customer energy consumption patterns in their daily life, optimize production costs in real time, improve distribution, and predict future market behavior. Many Companies like utility companies and their customers both benefit from energy consumption data analytics. By using data analytics techniques and energy datasets, energy managers can easily monitor and reduce grid downtime, balance assets, predict changes to fluctuating market demands in real time due to the customers percepts , and track environmental objectives. Meanwhile, energy consumers such as managers of municipalities office buildings and large factories can use energy consumption data analytics to identify checkpoints for potential savings in their earnings , adjust usage to strategy, and collaborate with site management department workers on energy consumption issues. In all cases, energy consumption data analytics helps users solve complex problems in the ecosystem and to make better operational and financial business decisions which will lead to the development of the society.

LITERATURE SURVEY

Background

Smart grids

The energy industries across the globe are facing numerous challenges. To reduce their carbon prints, expand their renewable energy portfolios, and to take energy conservation measures , there are huge pressures from regulatory authorities and environmental organisation to reduce their carbon footprint, and to take energy conservation measures. Demand response and its impact on consumers behavior require rapid adaptation in energy service providers(ESP) business models. As per United States Federal Energy Regulatory Commission:

To reduce wholesale power prices or to increase awareness of energy use or to provides more efficient operations of market or to enhances reliability demand responses provide competitive pressure and in combination with certain modern technologies, it can support the use of renewable energy resource, distribute generations, and advance metering. Thus, enabling of demand-side resource or supply-side resource, improves the economic operation of electric power markets by aligning prices more closely with the value customers place on electric power.

Power system participants are giver or taker of electricity. Demand responses and reliabilities issue along with distribution models having conventional electric power on the taker side are causing a great trend in motivate takers to produce supply at a local level mostly using these renewable energy production method. “Consumer” is an essential term for an economical motivated entity i.e

- take, produce, and power storage,
- Operate or own a small or large power grid
- Maximize the economic decision about it.

Currently uses energy grid supports one-direction distribution model and centralised in nature. Very less ability to cop-up with the consumer needs. Line losses and hierarchical topology makes it very much less reliable. They become bottleneck when rapid adaptation are required for the demand response.

Smart grids as:

Next-generation electricity grid expected to define the major losses of the present grid. In similar, the smart grid has to provide the utility company with full visibility and pervasive control over their goods and service. The smart grid has to be self-adjusting. And last but not last, a smart grids needs to empower its hare takers to define and find new ways to engage with each other and perform transaction of energy across the systems.

Energy efficiency and eco-efficiency

A lot of increasing dependency these days on conventional energy resources which leads to unprecedented challenges are part of a global phenomena. Countries like European Union has started to focus on energy efficiency methods, techniques so that they can ensure by their sides security of energy supply by making people understand to reduce primary energy consumption and their government has given orders to decrease energy imports which indirectly reduces emissions of greenhouse gases which then as a chain thereby mitigates the climate changes. The boards and states have agreed for reducing at least 20% of EU's major consumption of primary energy by 2020 in meeting of European Union in 2007. The Energy Efficiency Directive, EU (2012) which defines efficiency of energy the ratio between the input of energy and output of performance, goods or energy. In 2006 year, the above definition was discussed for efficiency of energy in the action plan of European Commission's. All major aspects of the efficiency of energy which are distribution, consumption, production and the value which is created as compared to resources which during the whole process gets consumed. Therefore, to measure the efficiency of energy and also to develop a methodology for the same, "Measuring and potentials of energy efficiency (EPO)" project was started back in year 2008 and to evaluate the savings. Being one the aim of this above mentioned project, a report was published named "Measuring efficiency of energy with Indicators and potentials in buildings, communities and energy systems". A model so as to calculate efficiency of energy was presented by this report and also same report highlighted the correlation with environmental factors.

The research that was presented in above report by VTT's considers efficiency of energy with the larger eco-efficiency as the subset. Some ecological factors are there that effects efficiency of energy for ex- CO2, NOx, Temp,SO2 etc. Efficiency of ecological itself a method to measure sustainable development. This whole system is summarized in the fig. below by VTT.

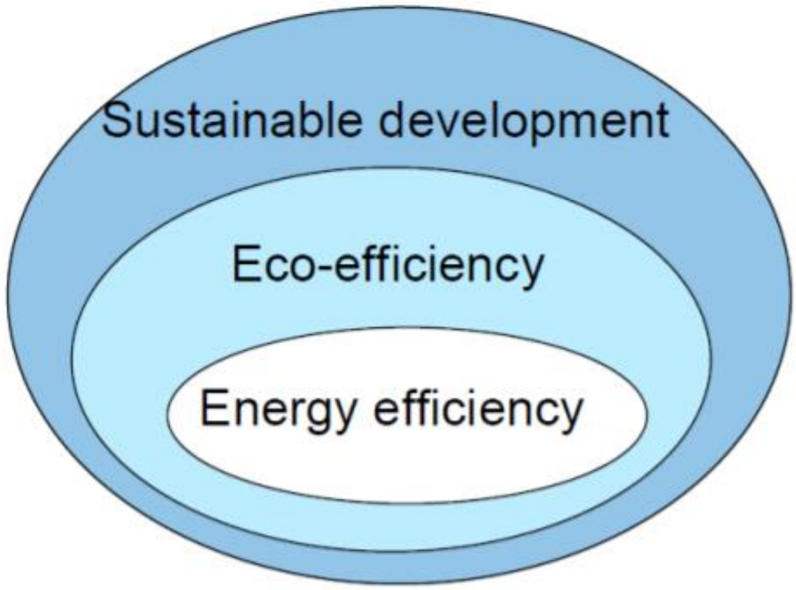


Fig: Energy efficiency, eco-efficiency and sustainability

Concept of eco-efficiency provide us basis for analysis of data in research. We have to apply basic and advance technique for analysis of data set collect from building unit consist of VTT's Green Campus initiative pilot project with motive of the efficiency models put forward in VTT's report.

Energy efficiency of a building = Energy consumed /Built area

In case of a Specific Energy Consumption (SEC) equation can be written as :

$$SEC = Q / A$$

In which Q denotes the consumption for a single type energy example electricity and A is the area built in square meters. In subsequent section, we have to refer these equations when we want to identify the pattern of use at building level, describe the relevance of energy efficiency and then discuss a model for classify building by efficiency of energy.

Daily consumption patterns, Base load and User load

Daily consumption pattern of a building unit refer to the energy usage of the building. Understanding daily usage patterns helps in identify the maximize or minimize points for improve the energy efficiency of that building unit. Base load of a building is also one of the important metric that we find from observing the daily consumption. The base load consumption that takes place without care of the original use of the building and energy consumption of user. It is basically the fix minimum load that a system of power supply has to be require to deliver. The base load is caused by the continuous consumption for building maintenance like Ac, ventilations, or night street lightings.The base load also consists energy consumptions by functionals component within building like server(computer), equipment of labs, and refrigerators. Eg. An office building has the maximum load during the running hours bcz users are using several additional appliances like pc, makers of coffee, extra lighting as compare to the off hours of the company. Below figure tells the concept of Base load and User load.

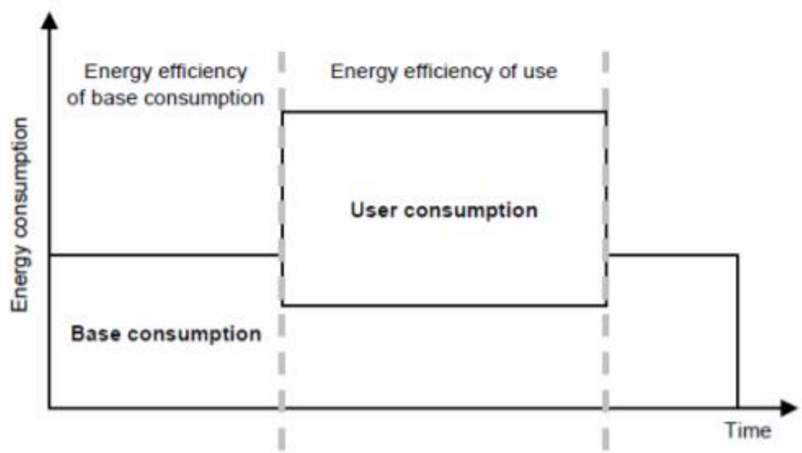


Fig: Base load, user load and energy efficiency

In above figure, the Base load is written as Base consumption. It provide a weighted metric that can be benchmark and compare. Basically help us to narrow down the scope of research by referring to building that are having problem and have various issues.

Time Series Analysis

It is the sequence of a random variable collected as time passes. From other example of series time data, meter devices energy consumption can collect periodically hence it constitute a time series.

Time series analysis is defined as comparison of a single time series at different point in time. Time series consists of a deterministic component and a random component.If X(t) -> time series data ,

Then we can say

$$X(t) = d(t) + E(t)$$

Where d(t) -> deterministic component and t is the random component. And E(t) is the amount of Energy consumption at time t.

Technology Stack:

Python: Python is a popular programming language. With the help of Python we are able to load the libraries like pandas and then handle big data to load and train then test the large data set .Also it allows complex mathematical operations and various graph and pictorial representations..

Machine Learning: it is basically data analysis method that automate the analytical model building. Part of artificial intelligence works on the idea that systems phases as learning of data, identification of patterns and take decisions with minimization of human interaction. After all that we can even work on more complex data , result delivery is faster that usual and results are more on accuracy side(even for large data).

Pandas: It is python library basically works on data sets. It has various functions like analyzing, exploring, cleaning, and manipulating of data.It basically allows us to analyze even large size data and make decisions based on statistical theories.

- **Sklearn:** Scikit-learn is the most important and robust library for ML in Python language. It provides a selected efficient tools for statistical modeling including regression, classification, clustering and dimension reduction via a better interface in Python. It built on the top of Numpy, Scipy and Matplotlib.

Jupyter Notebook: It is python based user interface in which user work with ordered list of input or output cell to achieve web server task and code solution deposit.

Binder: It is repository full of Jupyter notebooks, notebooks can be opened in executable environment, code can be reproduce by anyone and anywhere. You can run notebooks directly from web browser without install anything. It is not persistent i.e all changes will be lost after some time automatically.

CONTRIBUTION

1.)Analysing the dataset for energy consumption:-

Here first of all we loaded data set (includes energy consumption of houses for a particular home appliance) into the model using pandas library and then implemented models like

1.)Linear Regression

2.)SVM

3.)Random forest

Using scikit library where inbuilt function is provided to implement these models .

We trained the model using pre-processed dataset and then tested it for a particular instance. This result can further be used so as to predict the value of future instance and business companies can earn more revenue by estimating an idea about the future needs.

2.)Finding the best suited model as per accuracy parameters:-

Here , we tested all models for a particular instance and checked the efficiency of each model using parameters:

1.)Coefficient of Variance

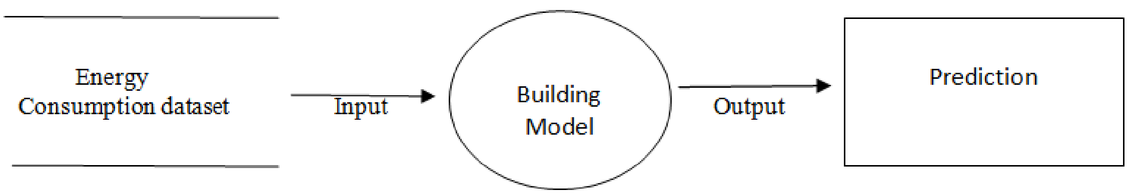
2.)Mean-Biased error

3.)R-squared error

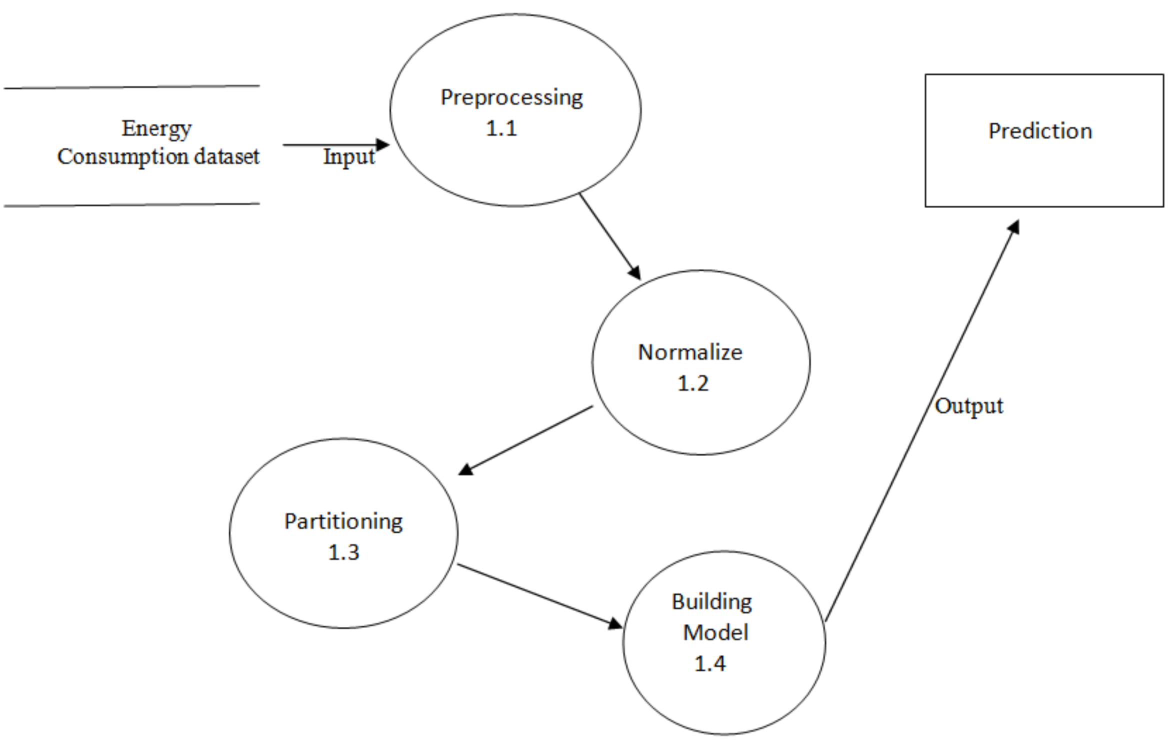
Our aim was to find a solution which has less error and more accuracy so that the forecasting done is correct as much as possible.

DATA FLOW DIAGRAMS

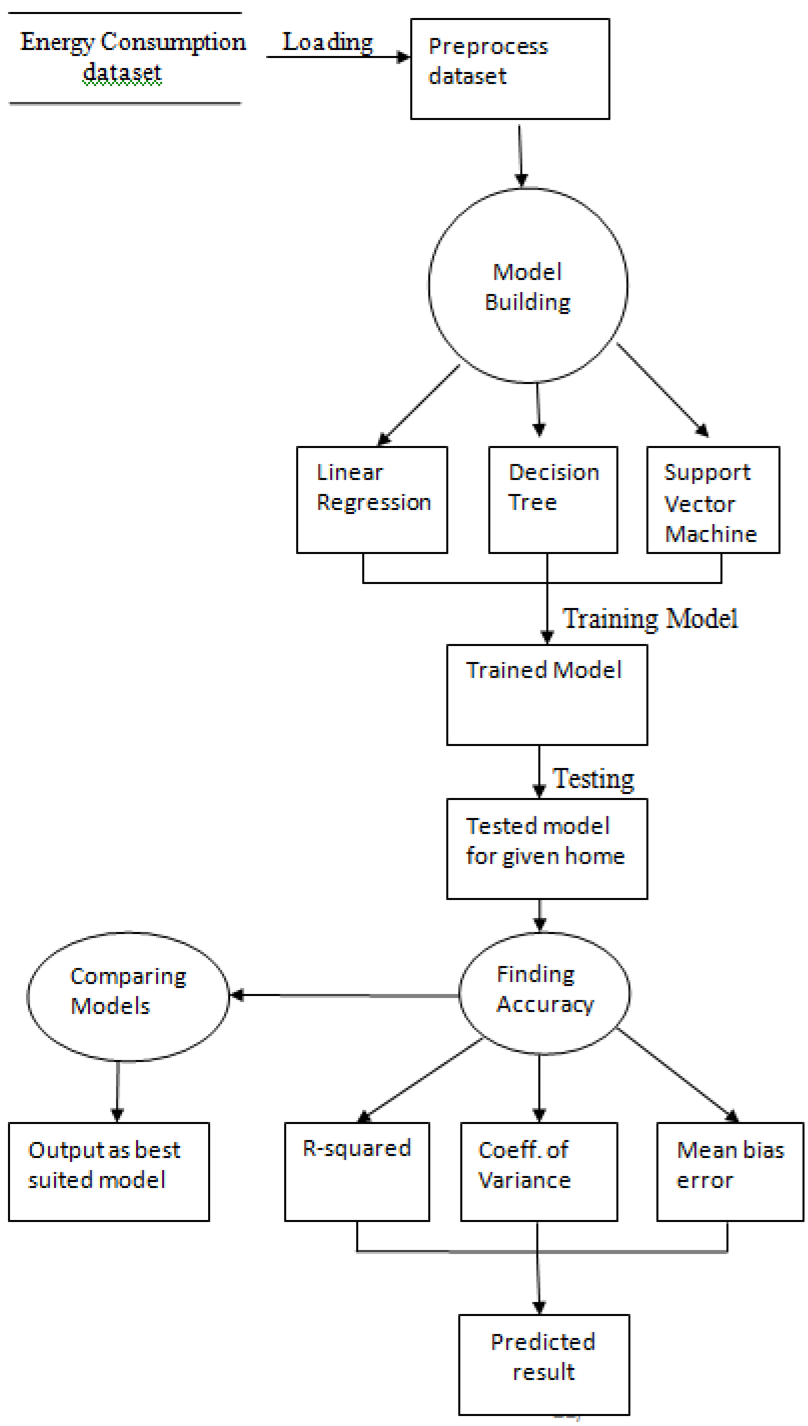
Context Level/Level 0 DFD:



Level 1 DFD:



Level 2 DFD:


Energy Consumption dataset
Loading
Preprocess dataset
Model Building
Linear Regression
Decision Tree
Support Vector Machine
Training Model
Trained Model
Testing
Tested model for given home
Finding Accuracy
Comparing Models
Output as best suited model
R-squared
Coeff. of Variance
Mean bias error
Predicted result

IMPLEMENTATION DETAILS

Our main goal has been to work on different machine learning models and compare them on a given large data set (which includes energy consumption based on time , weather and irradiance) to actually analyze the energy consumed by a house or a building in given time so that we can further predict the same. The models on which we have worked are-:

1) Loading the dataset
2) Linear regression
3) Random forest
4) Support Vector Machine
5) Compared these models based on:-
   5.1) Coefficient of variance
   5.2) Mean bias error
   5.3) R Squared

1**.) Loading the dataset**

First of all we loaded the data given using panda library . Now when the dataset is loaded to the model,we can see in the graph below ,the correlation of time.irradiance and temperature upon consumption of energy. However we filter temperature values less than 30 and consumption values less than 10 to remove most of the outliers. Also the temperature was manually shift by 2 hours and irradiance by 5 hours, so that a linear correlation was achieved.

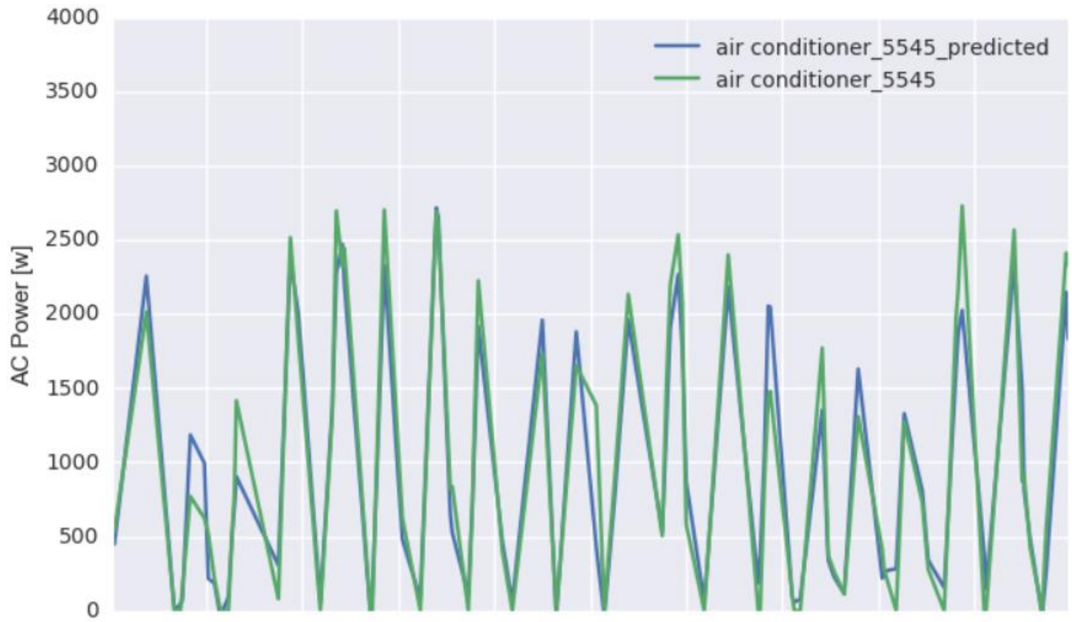
2.) **Linear regression**

As we know, Linear Regression is a machine learning algorithm .It a supervised learning approach and performs a regression task. It leads to target prediction value based on independent variables. It is mostly used for finding out the relationship between variables and forecasting. So here the variables are time(T),temperature(tp),irradiance(I). It is implemented using sklearn library. It is the most simple model for predicting and analyzing the large dataset. The syntax used is:-

```
from sklearn import linear_model
linear_reg = linear_model.LinearRegression()
```

```
linear_reg.fit(X_train, y_train)
```

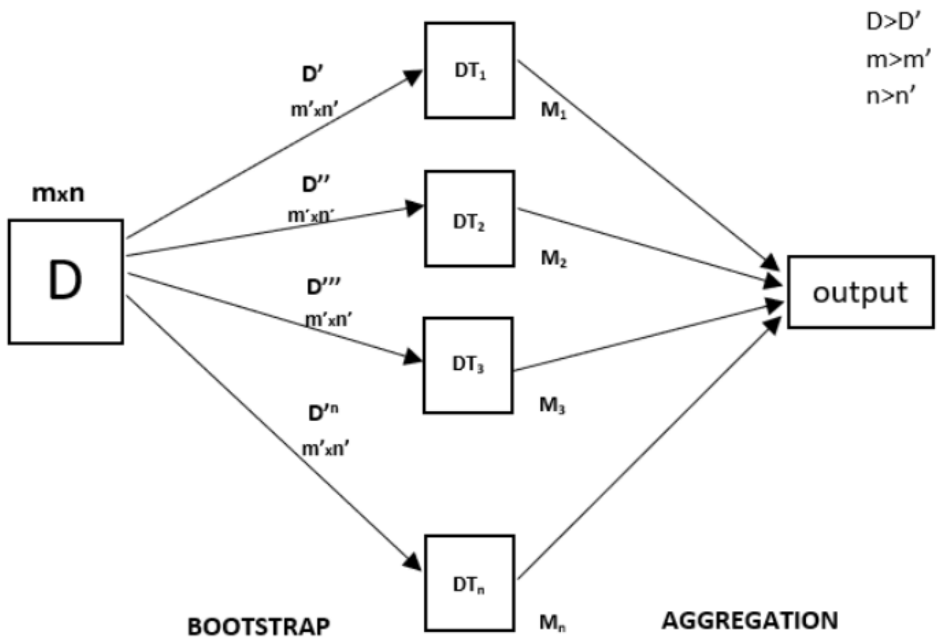
so for the model we tested it for house no =5545 and got the following outcome



3.) **Random Forest Decision tree**

A Random Forest is a technique used for both regression and classification purpose and it works with the help of multiple decision trees and a technique called Bootstrap and Aggregation, known as bagging. So here we combine multiple decision trees to find the final result instead of relying on individual decision trees. It uses multiple decision trees as base to learn models. We randomly perform row sampling and feature sampling from the dataset forming sample datasets for every model and it is called Bootstrap. Pseudo Code used here is:-

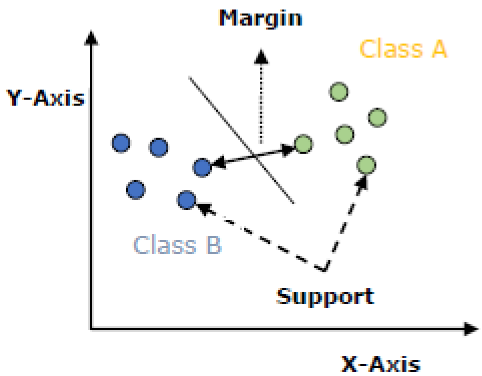
```
from sklearn.ensemble import RandomForestRegressor
reg = RandomForestRegressor()
```



4.) **Support Vector Machine**

An SVM model is generally a representation of different classes in a hyperplane in multidimensional space. Here, the hyperplane is generated in a repeated manner so as to minimise the error .Our aim here is to divide the datasets into classes so that we can find a maximum marginal hyperplane (MMH).Here in the below graph , we can see the margin between class A and class B. Code used is:-

```
from sklearn.svm import SVR
reg = SVR(kernel='rbf',C=10,gamma=1)
```



5.) **Compare models**

Here we compared the models so as to find the best suited model for predicting the energy consumption on the basis of :

5.1) Coeff of variance

It is also called relative standard deviation and it defines **variation** of the sample data on different runs.in our project we used this so as to compare models and find which suits best for our dataset .

$$\mathrm{CV} = \frac{\sqrt{\frac{\sum_{i=1}^{N}\left(y_{\mathrm{pred},i} - y_{\mathrm{data},i}\right)^2}{N}}}{\bar{y}_{\mathrm{data}}} \times 100,$$

$$\mathrm{MBE} = \frac{\frac{\sum_{i=1}^{N}\left(y_{\mathrm{pred},i} - y_{\mathrm{data},i}\right)}{N}}{\bar{y}_{\mathrm{data}}} \times 100$$

5.2) Mean biased error:-

Mean bias error is used for finding the average bias in built model and to take any step to further correct bias in any run.

5.3) R-squared error:-

It is defined as the fraction by which the variance of the errors is smaller than the variance of the dependent variable. We used this to evaluate the performance of models .

So , after comparing we got the results defined in result section.

RESULTS & OBSERVATIONS

1.) **Accuracy results for all models during testing on a particular test set**:-

1.1) For linear regression:

coefficient of variance = 61.70

Mean bias error = 35.16

R Squared = 0.894

1.2) For decision tree:

coefficient of variance = 57.87

Mean bias error = 26.10

R Squared = 0.907

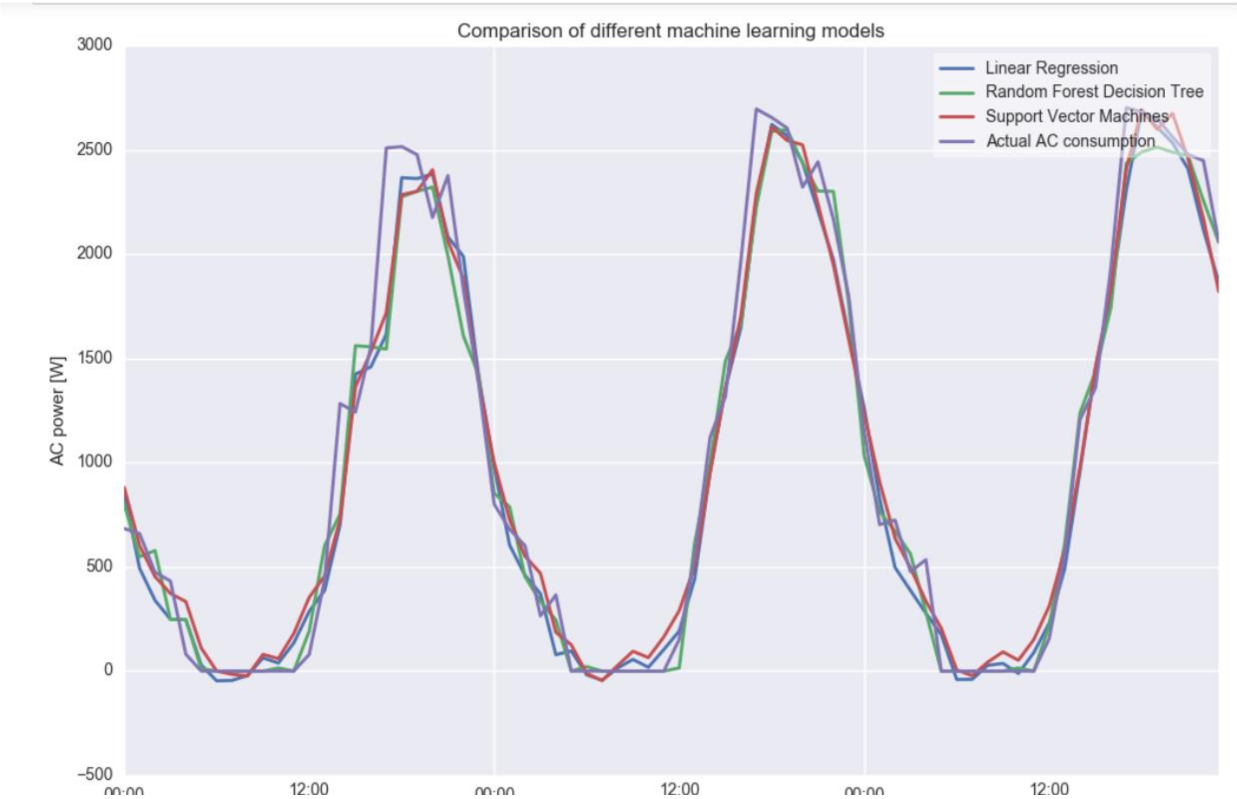
1.3) For Support Vector Machine:

coefficient of variance = 62.41

Mean bias error = 43.10

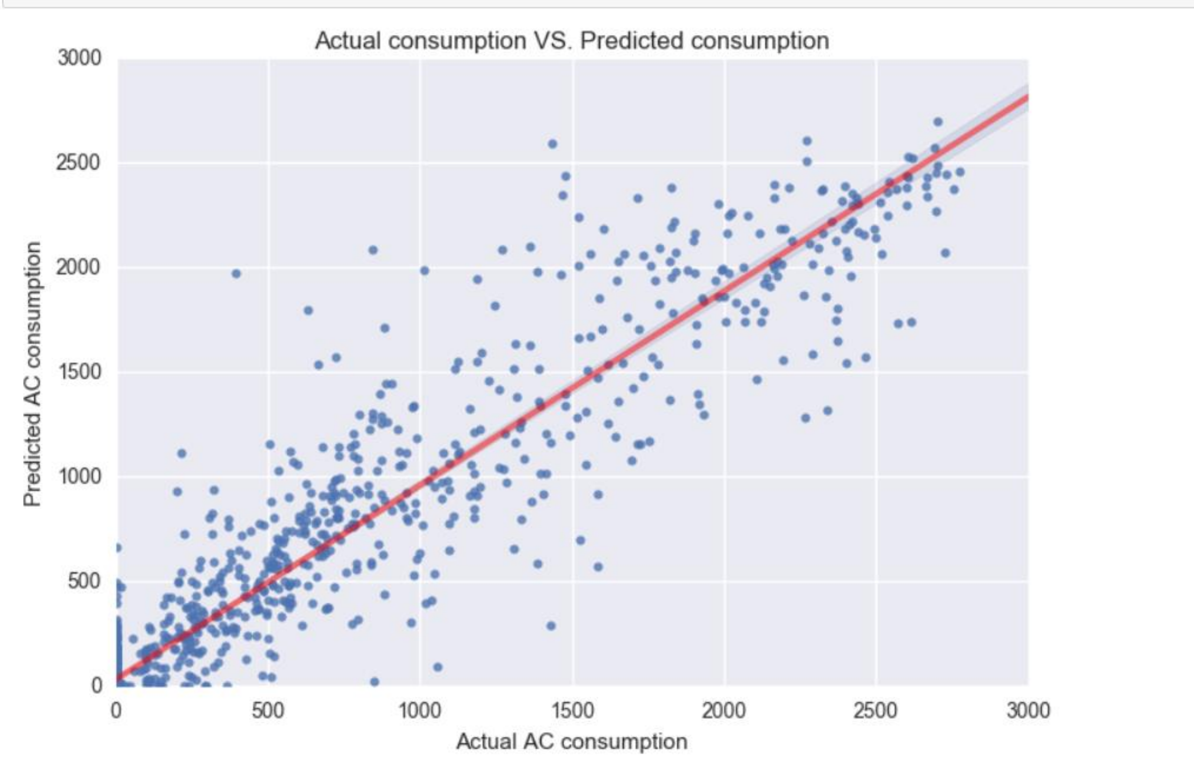
R Squared = 0.892

2.) **Graph representing the prediction of all models and actual value**:-



The graph above shows the final output and comparison of models generated and this result is for a particular test data i.e. a house and at given time .We did the comparison of different prediction models for forecasting the future energy consumption by A.C. or other different home appliances based on their previous usage.

3.) **For Random Forest Tree** we can see the actual value and predicted as in below graph:-


Actual consumption VS. Predicted consumption
Predicted AC consumption
Actual AC consumption
0
500
1000
1500
2000
2500
3000

CONCLUSION

Global needs for energy consumption are rapidly enhancing. The basic or traditional methods to generate energy doesn't seem a better option these days. Hence to find efficient energy generation and minimize the consumption is most important so as to match with the increase in demand . So, now our main concern is to understand the pattern of consumption for a large data set , and to predict the future demands. The technical achievement in sensors, grid computing acts as a base to collect the data of energy consumption for further predictions. Hence this data should be processed to find parameters to improve energy efficiency. Now we can use different types of data set either large or small but large data set leads to better result because number of rows are large in this and training set increases in size. That is why we used large data set in our project. Also , we should prefer the recent data set instead of historic one. Now as we know the volume for data keeps increasing day by day . Now here in our project , we have firstly loaded the data in relation with radiance and weather . When we trained the model and then tested it we Checked the efficiency of all models on the basis of

1.) coefficient of variance

2.) mean bias error

3.)R-squared error.

Here in our project we analysed the actual value and the values from different models so as to pro vide flexibility in visualization of consumption.

The above prediction can act as a first step for decision making by consumers and companies pro ducing the power e.g. TORRENT,ADANI etc .The decision support system thus can predict the P attern and then find the best option either it may be green energy resources or any other source.

We presented and demonstrated the concept of Data analytics platform and applied it to solve the real-life use cases from the energy industry. The output of our research is a working smart meter analytics with every model and the results generated from advance analytics techniques applied specifically to solve energy efficiency problems.

REFERENCES

[1] Arundel, Kemp, . United Nations University Working Paper Series, 2009/017, 2009.

[2] Civis project - european union 7th framework programme. http://www. civisproject.eu/.

[3] Directive 2012/27/eu of the european parliament 25 october 2012 on energy efficiency. http://eur-lex.europa.eu/LexUriServ/ LexUriServ.do?uri=OJ:L:2012:315:0001:0056:EN:PDF.

[4] Balijepalli, V. M., Pradhan, V., In Innovative Smart Grid Technologies-India (ISGT India), 2011 IEEE PES (2011), IEEE, pp. 236–243.

[5] Boehm, B., SIGSOFT Softw. Eng. Notes 11, 4 (Aug. 1986), 14–24.

[6] Jeff Atwood and Joel Spolsky Stack Overflow Inc. https://stackoverflow.com/documentation cc by-sa. rev 2020.7.3.37177, 2008

[7] Refsnes Data, w3 schools, www.**w3schools**.com, 2008-04-21

[8] Sandeep Jain, GeeksForGeeks www.geeksforgeeks.org , 2008

[9] Wes McKinney, Pandas, https://pandas.pydata.org/pandas-docs/version/1.0.0/development/developer.html, 2015

[10] David Cournapeau, Scikit-Learn, https://scikit-learn.org/stable/developers/index.html, 2007

[11] Fernando Pérez, Jupyter, https://jupyter.org/, 2014

APPENDIX

1.) Loading the data:-

```
pip install pandas
pip install matplotlib
pip install sklearn

import pandas as pds
from os.path import join
%matplotlib inline
from matplotlib import pylab as plb
plb.rcParams['figure.figsize'] = (10, 6)
disaggregated_path = '../data/consumption/'
path_of_data = '../data/'
combined =pds.HDFStore(join(disaggregated_path,'all_houses_disagg.h5'))
key = combined.keys()[0].replace('/','')
dataframe = combined.get(key)
combined.close()
dataframe.head()
df = dataframe[dataframe.filter(regex='conditioner1').columns]
%run functions.ipynb
buildings_dict = builds_dictionary(path_of_data) #contains basic details of each house, you may
explore this variable if you want
start = '2014-06-10'
end = '2014-06-12'
building = '5545' #selecting one building from the list
consum_col = 'air conditioner1_'+building
```

```
consum_col_raw = 'air conditioner_'+building
consum_col_mains = 'air conditioner_'+building+'_mains'
plb.figure()
df[consum_col][start:end].plot()
plb.xlabel('Time')
plb.ylabel('AC Power [W]')
plb.show()
df_mod_amb = df_mod.join([weather,irradiance]).dropna()
df = df_mod_amb[start:end]
fig = plb.figure()
ax1 = fig.add_subplot(111)
ax2 = ax1.twinx()
ax3 = ax1.twinx()
rspine = ax3.spines['right']
rspine.set_position(('axes', 1.1))
colors = ['b','g','r']
df.plot(ax=ax1, y=consum_col_raw, legend=False,color=colors[0])
ax1.set_ylabel('Consumption',color=colors[0])
ax1.tick_params(axis='y', colors=colors[0])
ax1.set_xlabel('Time')
plb.show()
df = df_mod_amb.loc[(df_mod_amb['temperature']>30) & (df_mod_amb[consum_col_raw] >
10)]
```

2.) SVM model:-

```
pip install pandas
pip install matplotlib
pip install sklearn
import pandas as pds
from os.path import join
```

```
%matplotlib inline
from matplotlib import pylab as plb
plb.rcParams['figure.figsize'] = (10, 6)
root_path = '../'
disaggregated_path = root_path+'/data/consumption/'
path_of_data = root_path+'/data/'
combined_raw = pds.HDFStore(join(disaggregated_path,'all_houses_raw_AC.h5'))
key = combined_raw.keys()[0].replace('/','')
dataframe_raw = combined_raw.get(key)
combined_raw.close()
df_raw = dataframe_raw.copy()
%run functions.ipynb
building = '5545' #selecting one building from the list
consum_col = 'air conditioner1_'+building
consum_col_raw = 'air conditioner_'+building
consum_col_mains = 'air conditioner_'+building+'_mains'
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(features, target,
test_size=0.2,random_state=41234)
from sklearn.svm import SVR
reg = SVR(kernel='rbf',C=10,gamma=1)
from sklearn.model_selection import cross_val_predict
predict = cross_val_predict(reg,features,target,cv=10)
predictions = pds.Series(predict,index=target.index).rename(consum_col_raw+'_predicted')
predictions_frame = pds.DataFrame(predictions).join(y_test).dropna()
from IPython.display import Image
Image(filename=root_path+'/notebook/images_on_notebooks/accuracy_metric_from_S_Karatas
ou_et_al.png', width=400, height=400)
#denormalize the dataframe
predictions_frame_denorm = (predictions_frame*(df_lagged[consum_col_raw].max()-
df_lagged[consum_col_raw].min()))+df_lagged[consum_col_raw].min()
```

```
# Accuracy metrics
metrics = accuracy_metrics()
print( "coefficient of variance =
{:.2f}".format(metrics.coeff_var(predictions_frame_denorm,consum_col_raw,consum_col_raw+
'_predicted')*100))
print( "Mean bias error =
{:.2f}".format(metrics.mean_bias_err(predictions_frame_denorm,consum_col_raw,consum_col_
raw+'_predicted')*100))
print( "R Squared =
{:.3f}".format(metrics.r2_score(predictions_frame_denorm,consum_col_raw,consum_col_raw+'
_predicted')))
predictions_frame_denorm['2014-08-01':'2014-08-14'].plot(legend=False)
plb.xlabel('Time')
plb.ylabel('AC Power [w]')
plb.ylim([0,4000])
plb.legend(['Predicted AC consumption','Actual AC consumption'])
import seaborn as sns
fig = plb.figure()
ax1 = fig.add_subplot(111)
plot = sns.regplot(x=consum_col_raw, y=consum_col_raw+'_predicted',
data=predictions_frame_denorm,ax=ax1,
line_kws={"lw":3,"alpha":0.5})
plb.title('Actual consumption VS. Predicted consumption')
plot.set_xlim([0,3000])
plot.set_ylim([0,3000])
plot.set_xlabel('Actual AC consumption')
plot.set_ylabel('Predicted AC consumption')
regline = plot.get_lines()[0];
regline.set_color('red')
```

3.) Comparison of models and finding efficient model:-

```
!pip install pandas
!pip install matplotlib
!pip install sklearn
!pip install tables
import pandas as pds
from os.path import join
%matplotlib inline
from matplotlib import pylab as plt
plt.rcParams['figure.figsize'] = (12, 8)
root_path = '../'
disaggregated_path = root_path+'/data/consumption/'
data_path = root_path+'/data/'
weather_path = root_path+'/data/weather/'
combined_raw = pds.HDFStore(join(disaggregated_path,'all_houses_raw_AC.h5'))
key = combined_raw.keys()[0].replace('/','')
dataframe_raw = combined_raw.get(key)
combined_raw.close()
df_raw = dataframe_raw.copy()
%run functions.ipynb
building = '5545' #selecting one building from the list
consum_col = 'air conditioner1_'+building
consum_col_raw = 'air conditioner_'+building
consum_col_mains = 'air conditioner_'+building+'_mains'
df_mod = (df_raw[[consum_col_raw]]).shift(-5)
df_mod_amb = df_mod.join([weather,irradiance]).dropna()
df_features = features_creation(df_mod_amb)
```