POWER CONSUMPTION ANALYSIS FOR HOUSE HOLDS

Using Machine Learning(Random Forest Regression)

Developed by:Srujana reddy polaka, Shobanaboina Sai prasanna, Kadarla Ramya, Pashikanti Vyshnavi

Smart Bridge-Remote Summer Internship program

1. INTRODUCTION

Electricity load forecasting has gained sustainable 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 levels of individuals 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 household.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.

Energy Consumption of software Green and sustainable software is a software product that has the smallest possible economic, societal, ecological impact as well as impact on human beings. This has led to the introduction of various programmes and initiatives that encourages energy efficient software such as green software engineering and Eco-design software.According to the Greenhouse Gas Protocol (2012), applications are executed with an OS. They affect the power consumption of a device due to data requests and processing. Managing energy requires accurate measurement of the energy available and consumed by a system. This involves monitoring or estimating the resource and energy consumption of hardware and software. However, a device's power consumption is subjected to the type of application and the task being performed which is evident in our experimental results presented in Section 4 of this paper. In order to reduce the overall power consumption for a web-based or standalone task, it will be necessary to provide users with an insight of the power consumption of the different web-based browser applications (e.g. Google Chrome, Internet Explorer, Mozilla Firefox, Safari, etc…) and also the resource hungry nature of many applications such as movie player and games.

In this research,we are interested in time series analysis with the popular forecasting technique that I used in this study; ARIMA(Auto regressive Moving Average) I applied this method for detecting patterns and trends of the electric power consumption in the households with real time series period in daily ,weekly, monthly, and quarterly.I used

Python program for constructing the model.

1.1.**Overview**

By examine and exploring all variables of the consumption of electrical power in individual household within practically four years, they wanted us to develop an ML-based solution to collect and monitor the sequences of patterns.so,we found out some patterns,seasonal features and other significant information that allows us to do forecasting of the future demand with a certain degree of accuracy.

1.2.Purpose

Our aim is to make use of pandas,matplotlib&seaborn libraries from python to extract the libraries for machine learning for power consumption analysis .

Secondly,to learn how to hyper tune the parameters using greed across the validation for the random forest machine learning algorithm.
And finally,to predict the active energy consumption for next day,next week,next month and next year using ensemble techniques of combining the predictions from multiple machine learning algorithms and withdrawing the conclusions.

2.LITERATURE SURVEY

Data mining is the process of analyzing data from different perspectives and extracting useful knowledge from it.It is the core of knowledge discovery process.The various steps involved in extracting knowledge from raw data is depicted.Different data mining technique include classification,clustering,sequentialpatterns,neuralnetworks,regression etc,.Regression is commonly applied data mining technique ,which employs a set of pre classified examples to develop a model.weight prediction,insurance prediction comes under regression.This approach frequently empoly's Decision tree based regression algorithm.In regression,train set is used to build model and test set is used to validate model.

2.1 Existing Problem

India is the world's third largest producer and third largest consumer of electricity. The gross electricity consumption in 2018-19 was 1,181 kWh per capita. In 2015-16, electric energy consumption in agriculture was recorded as being the highest (17.89%) worldwide. Energy use can be viewed as a function of total GDP, structure of the economy and technology.The increase in household energy consumption is more significant than that in the industrial sector. To achieve reduction in electricity consumption, it is vital to have current information about household electricity use.Many households worry a lot due to the high electricity bill. Even they want to save the consumption of power. Due to the lack of information about individual appliances power consumption leads to huge wastage of power as well as money.

2.2.Proposed Solution

Machine Learning (Random Forest):

Random Forest is a popular machine learning algorithm that belongs to the supervised learning technique.It can be used for both Classification and Regression problems in ML.It is based on the concept of ensemble learning,which is a process of combining multiple classifiers to solve a complex problem and to improve the performance of the model.

As the name suggests,***Random Forest is a classifier that contains a number of decision tress on various subsets of the given data set and takes the average to improve the predictive accuracy of that data set***.Instead of relying on one decision tree , the random forest takes the prediction from each tree and based on the majority votes of predictions,and it predicts the final output.

The greater number of trees in the forest leads to higher accuracy and prevents the problem of over fitting.

3.THEORETICAL ANALYSIS

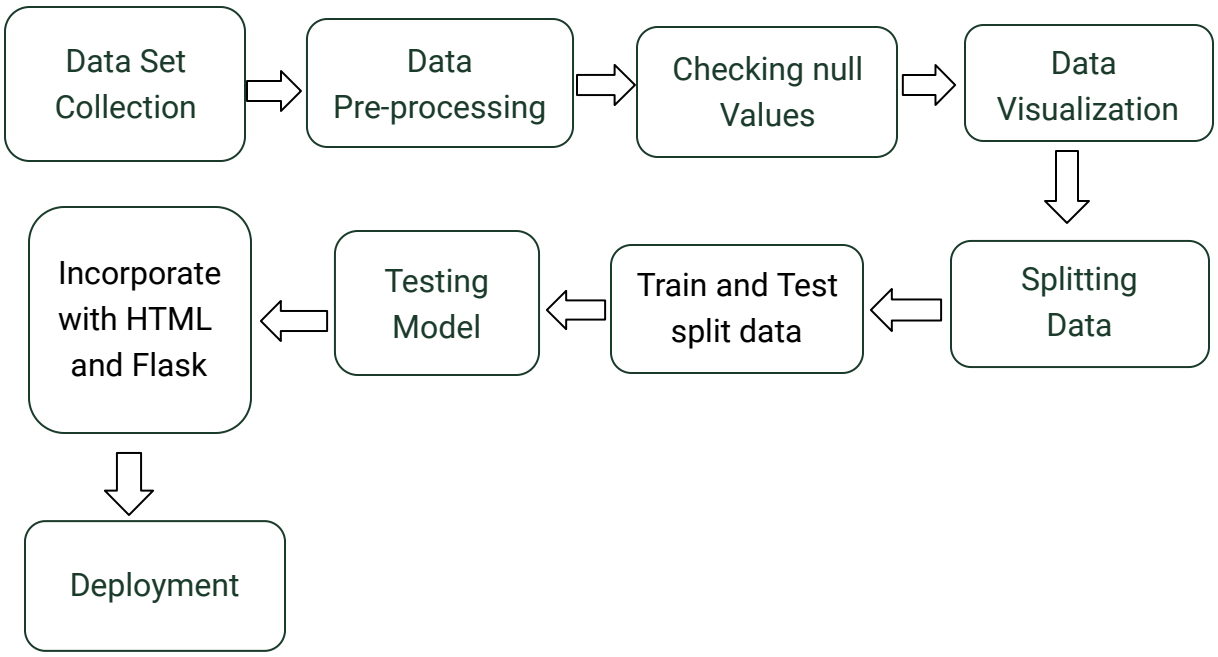
While selecting the algorithm that gives an accurate prediction we gone through lot of algorithms which gives the results abruptly accurate and from them we selected only

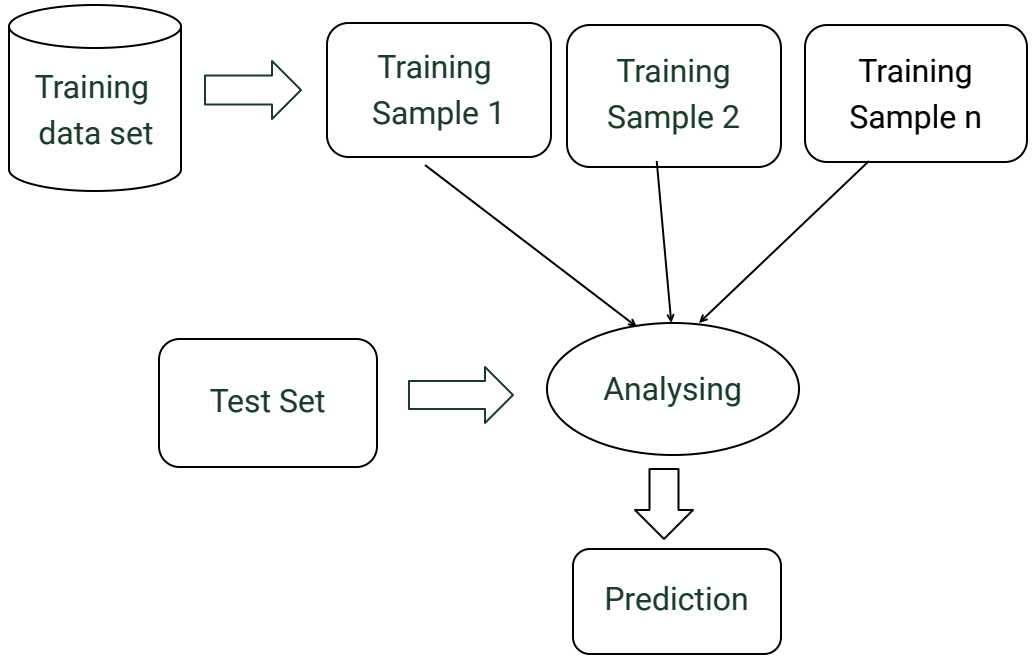
one algorithm for the problem analysis that is Random Forest Regression,it assumes that the presence of a particular feature in a class is unrelated to the presence of any other feature thats how the analysis work great with Random Forest Algorithm.

Below are some points that explain why we should use the Random Forest Algorithm:

- It takes less training time as compared to other algorithms.
- It predicts output with high accuracy,even for the large data set it runs efficiently.
- It can also maintain accuracy when a large proportion of data is missing.

3.1.Block Diagram





3.2.Software Designing

- Jupyter Notebook Environment
- Sypder Ide
- Machine Learning Algorithms
- Python(Pandas,numpy,matplotlib,seaborn,sklearn)
- HTML
- Flask

We developed this health monitoring by using the python language which is a interpreted and high level programming language and using the machine learning algorithms. For coding we used the jupyter notebook environment of the anaconda distributions and the spyder.It is an integrated scientific programming in the python language. For creating an user interface for the prediction we used flask.It is a micro web framework written in python.It is classified as a microframework because it does

not require particular tools or libraries.It has no database abstraction layer,form validation,or any other components where pre-existing third-party libraries provide common functions,and a scripting language to create a webpage in HTML by creating the templates to use in functions of the Flask andHTML.

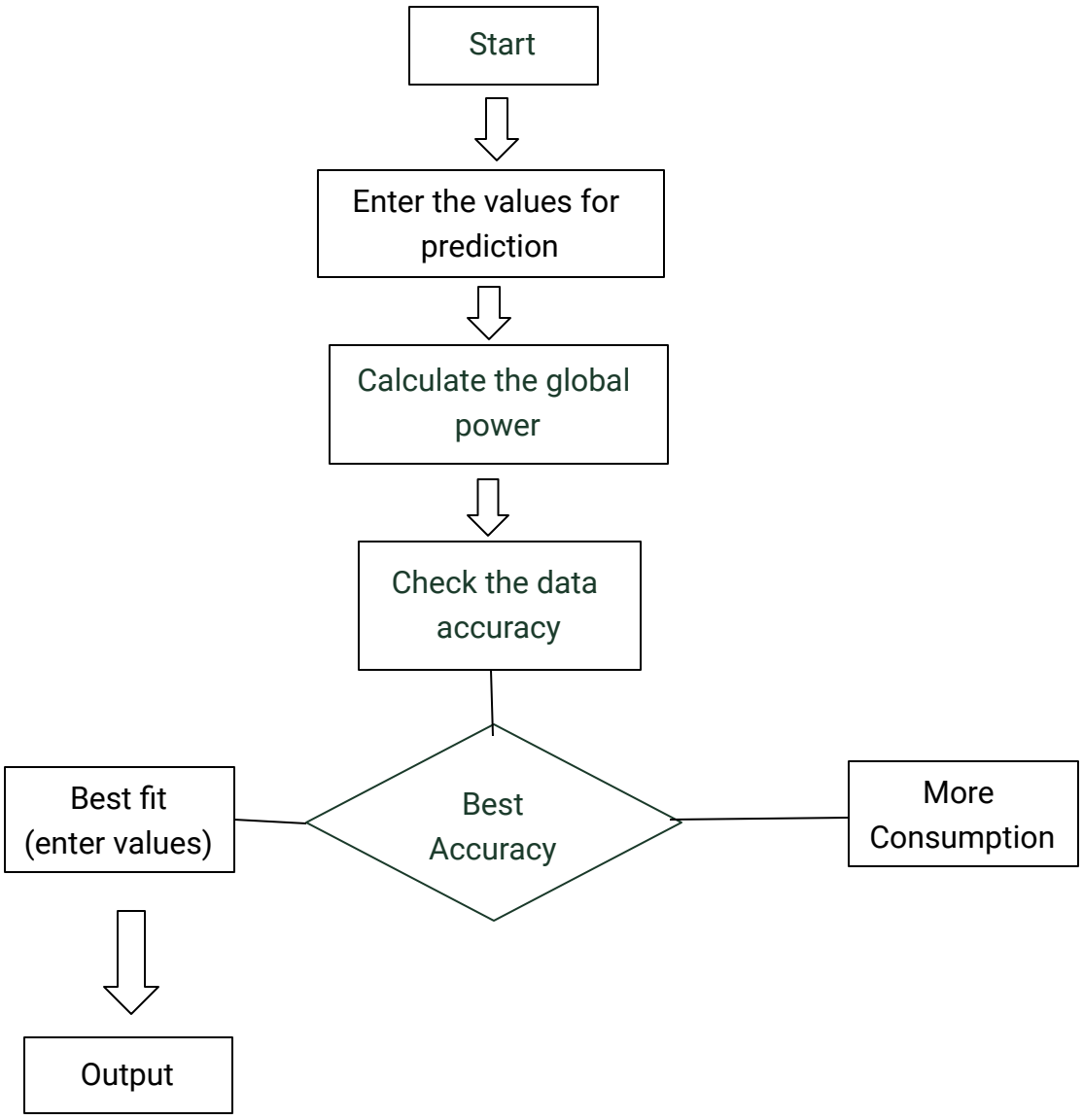
4.EXPERIMENTAL INVESTIGATION

Algorithms used were **Linear regression** , **Decision-Tree regression** , **Random forest regression**.The dataset we used is derived from www.kaggle.com. It contains plenty of datasets which are real time. We choosed power prediction dataset which contains 9 attributes and 400 rows. After that, the missing values are checked and the unwanted columns are deleted, finally we have retained to 8 attributes.Those are shown below.

	Date	Time	Global_active_power	Global_reactive_power	Voltage	Global_intensit
0	16/12/2006	17:24:00	4.216	0.418	234.84	18.
1	16/12/2006	17:25:00	5.360	0.436	233.63	23.
2	16/12/2006	17:26:00	5.374	0.498	233.29	23.
3	16/12/2006	17:27:00	5.388	0.502	233.74	23.
4	16/12/2006	17:28:00	3.666	0.528	235.68	15.
...	...	...	...	...	...	.
6066	20/12/2006	22:30:00	3.136	0.052	238.85	13.
6067	20/12/2006	22:31:00	3.126	0.052	238.51	13.
6068	20/12/2006	22:32:00	2.884	0.046	237.82	12.
6069	20/12/2006	22:33:00	2.768	0.000	237.97	11.
6070	20/12/2006	22:34:00	2.758	0.106	238.12	11.

6071 rows × 9 columns

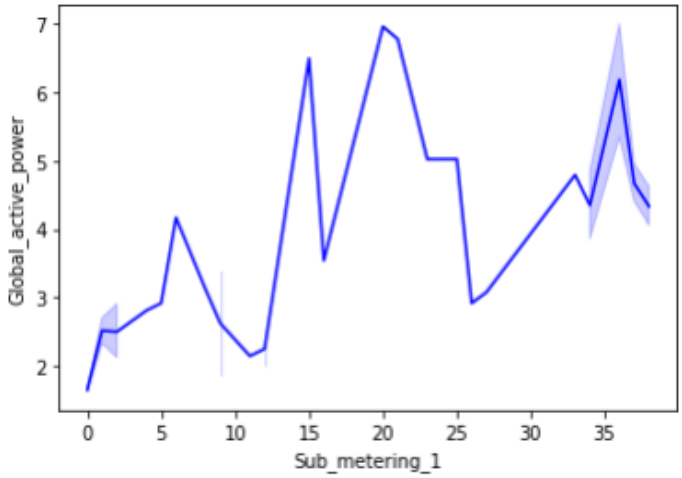
5.FLOWCHART

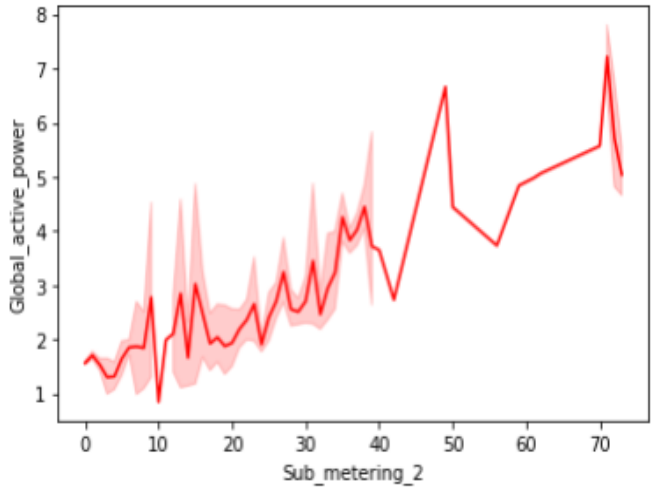


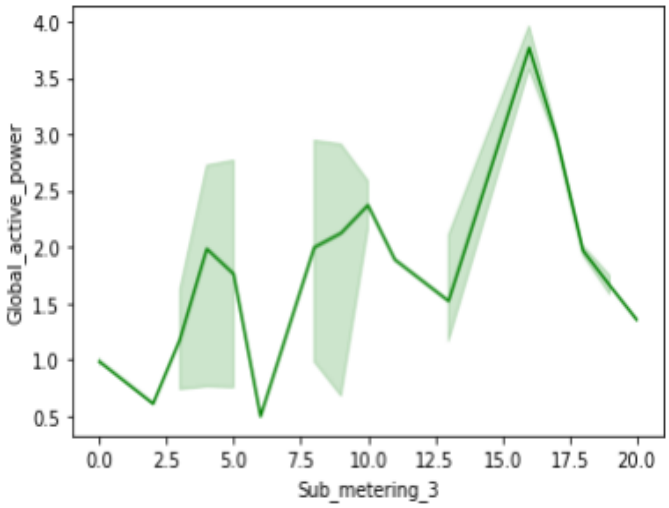
6.RESULT

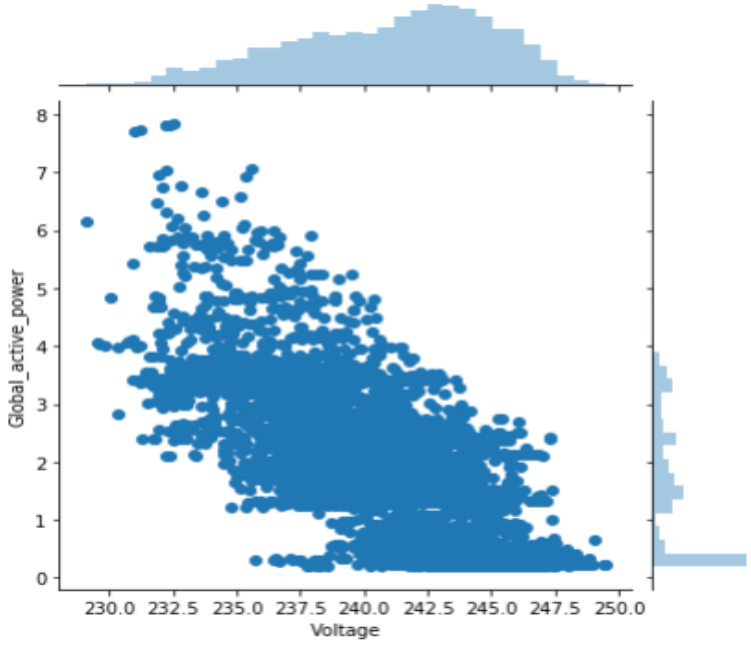
We are proposing a system that analyzes electricity consumption of individual appliances, identify the energy consumption patterns apply statistical modeling and suggest the user about reduction of power consumption.The main objective of this project is to analyze and visualize the energy consumption. The regular consumption is captured through sensor and stored in database. Applying various exploratory data analysis using various libraries, visualizing them by

using matplotlib and seaborn and suggesting the user based on consumption. Web application to visualize the Analysis. Get insights from the data. suggests the user based on insights.The algorithm is also a great choice for anyone who needs to develop a model quickly.On the top of that,it provides a pretty good indicator of the importance it assigns to your features.









Sl.no	Algorithm Used	Accuracy
1.	Random Forest	99%
2.	Decision Tree	99%
3.	Linear Regression	99%

7.ADVANTAGES AND DISADVANTAGES

ADVANTAGES

- It reduces greenhouse gas emissions which contribute to climate change. While renewable sources of energy, like wind, play an increasing role in our system, the vast majority of our electricity is still generated from fossil fuels like gas, oil and coal.
- Saving money by reducing your power bills in the long term by reducing the need for more generators and network infrastructure (costs which get passed onto customers)
- Reducing our electricity use reduces the flow of carbon dioxide into the atmosphere.

DISADVANTAGES

- Unwanted Side Effects
- Power plants that burn fossil fuel pump carbon dioxide into the atmosphere. Carbon dioxide is a greenhouse gas that causes Earth's temperature to rise. Nuclear power plants must find ways to dispose of radioactive waste safely.

8.APPLICATIONS

- The use of applications or software on our computer systems consumes energy and it also affects how various hardware components and system resources consume energy. Consequently, running web browsers applications will utilize considerable energy and battery consumption.
- Generating the Many more resources for the upcoming generation.
- Limited usage gives more prediction of power

9.CONCLUSION

Therefore in this individual house hold electricity consumption dataset**,ss** power consumption is used to build a UI model for predicting electric consumption of individual appliances and the results are compared to identify the power consumption patterns apply statistical modeling**.** The model shows the analysis of house holds energy prediction which performs best than the other two algorithms in the prediction of power consumption. There is no definitive guide of which algorithms to be used. What may work on some datasets may not work on others. Therefore, always check the accuracy and predict with the dataset values

10.FUTURE SCOPE

This linear regression model can also be used in the future predictions like weather prediction, salary prediction etc. In future we can create a web application on this type of problems so that it can be accessed from everywhere with more users.
In future random forest can be applied on other data sets available for health care. Analysis of other machine learning algorithms other than these six can also be done in future to investigate the power of machine learning algorithms for health care.In further study,we will try to conduct experiments on larger data sets or try to tune the model so as to achieve the state of art performance of the model and a great UI system making it complete web application model.

11. BIBLIOGRAPHY

● Adichie, J. N. (1967). Estimates of regression parameters based on rank test. Annals of Mathematical Statistics, 38, 894-904.
● Aia, M. A., Goldsmith, R. L., and Mooney, R. W. (1961). Predicting stoichiometric CaHP04 • 2H20.Indllstrial and Engineering Chemistry, 53, January, 55-57.
●Adcock, R. J. (1878). A problem in least squares. Analyst, 5, 53-54.
● Andrews, D. F. (1974). A robust method for multiple linear regression. Technometrics. 16. 523-531.
●Andrews, D. E, and D. Pregibon (1978). Finding the outliers that matter. Journal of the Royal Statistical Society, Series B, 4, 84-93.

●Barnett, V. D. (1967). A note on linear structural relationships when both residual variances arc known. Biometrika, 54, 670-672.
●Dobson. A. J. (1990). Introduction to Generalized Linear Models. London: Chapman and Hall.
●Marquardt, D. W. (1963). An algorithm for least squares estimation of nonlinear parameters. Journal of the Society for Industrial and Applied Mathematics, 11,431-441.
●McCullagh, P., and J. A. Nelder (1989). Generalized Linear Models, 2nd ed. London: Chapman and Hall.
●Schemper. M. (1990). The explained variation in proportional hazards regression. Biometrika, 77,216-218.

APPENDIX

HTML:

```
<!DOCTYPE html>
<html>
<!--From https://codepen.io/frytyler/pen/EGdtg-->

<head>
  <meta charset="UTF-8">
  <title>ML API</title>
  <link href='https://fonts.googleapis.com/css?family=Pacifico' rel='stylesheet'
type='text/css'>
  <link href='https://fonts.googleapis.com/css?family=Arimo' rel='stylesheet'
type='text/css'>
  <link href='https://fonts.googleapis.com/css?family=Hind:300' rel='stylesheet'
type='text/css'>
  <link href='https://fonts.googleapis.com/css?family=Open+Sans+Condensed:300'
rel='stylesheet' type='text/css'>
  <link rel="stylesheet" href="../static/css/style.css">

  <style>
    .login
{
      top: 20%;
```

```
    }
  </style>
</head>

<body>
  <h1>Global_active_power</h1>
  <div class="login">


    <!-- Main Input For Receiving Query to our ML -->
    <form action="/login" method="post">
      <label for="Global_reactive_power"> Global_reactive_power</label>
      <input type="float" name="Global_reactive_power" placeholder="Enter
Global_reactive_power" id="Global_reactive_power" required="required" />

      <label for="Voltage">Voltage</label>
      <input type="float" name="Voltage" placeholder="Enter Voltage" id="Voltage"
required="required" />

      <label for="Global_intensity">Global_intensity </label>
      <input type="float" name="Global_intensity" placeholder="Enter Global_intensity"
id="Global_intensity" required="required" />

      <label for="Sub_metering_1">Sub_metering_1</label>
      <input type="float" name="Sub_metering_1" placeholder="Enter Sub_metering_1"
id="Sub_metering_1" required="required" />

      <label for="Sub_metering_2">Sub_metering_2</label>
      <input type="float" name="Sub_metering_2" placeholder="Enter Sub_metering_2"
id="Sub_metering_2" required="required" />

      <label for="Sub_metering_3">Sub_metering_3</label>
      <input type="float" name="Sub_metering_3" placeholder="Enter Sub_metering_3"
id="Sub_metering_3" required="required" />

      <button type="submit" class="btn btn-primary  btn-large">click</button>
      <button>ACTIVE_POWER:{{y}}</button>
```

```
        <br>
        <br> {{ prediction_text }}
        </form>


    </div>


</body>

</html>
```

CSS:

```
@import url(https://fonts.googleapis.com/css?family=Open+Sans);
.btn
{
    display: inline-block;
    margin-top: 1rem;
    *display: inline;
    *zoom: 1;
    width: 40%;
    padding: 4px 10px 4px;
    margin-bottom: 0;
    font-size: 13px;
    line-height: 18px;
    color: #333333;
    text-align: center;
    text-shadow: 0 1px 1px rgba(255, 255, 255, 0.75);
    vertical-align: middle;
    background-color: #f5f5f5;
    background-image: -moz-linear-gradient(top, #ffffff, #e6e6e6);
    background-image: -ms-linear-gradient(top, #ffffff, #e6e6e6);
    background-image: -webkit-gradient(linear, 0 0, 0 100%, from(#ffffff), to(#e6e6e6));
    background-image: -webkit-linear-gradient(top, #ffffff, #e6e6e6);
    background-image: -o-linear-gradient(top, #ffffff, #e6e6e6);
```

```
    background-image: linear-gradient(top, #ffffff, #e6e6e6);
    background-repeat: repeat-x;
    filter: progid: dximagetransform.microsoft.gradient(startColorstr=#ffffff,
endColorstr=#e6e6e6, GradientType=0);
    border-color: #e6e6e6 #e6e6e6 #e6e6e6;
    border-color: rgba(0, 0, 0, 0.1) rgba(0, 0, 0, 0.1) rgba(0, 0, 0, 0.25);
    border: 1px solid #e6e6e6;
    -webkit-border-radius: 4px;
    -moz-border-radius: 4px;
    border-radius: 4px;
    -webkit-box-shadow: inset 0 1px 0 rgba(255, 255, 255, 0.2), 0 1px 2px rgba(0, 0, 0,
0.05);
    -moz-box-shadow: inset 0 1px 0 rgba(255, 255, 255, 0.2), 0 1px 2px rgba(0, 0, 0, 0.05);
    box-shadow: inset 0 1px 0 rgba(255, 255, 255, 0.2), 0 1px 2px rgba(0, 0, 0, 0.05);
    cursor: pointer;
    *margin-left: .3em;
}

.btn:hover,
.btn:active,
.btn.active,
.btn.disabled,
.btn[disabled]
 {
    background-color: #e6e6e6;
}

.btn-large
 {
    padding: 9px 14px;
    font-size: 15px;
    line-height: normal;
    -webkit-border-radius: 5px;
    -moz-border-radius: 5px;
    border-radius: 5px;
}
```

```
.btn:hover
 {
   color: #333333;
   text-decoration: none;
   background-color: #e6e6e6;
   background-position: 0 -15px;
   -webkit-transition: background-position 0.1s linear;
   -moz-transition: background-position 0.1s linear;
   -ms-transition: background-position 0.1s linear;
   -o-transition: background-position 0.1s linear;
   transition: background-position 0.1s linear;
}

.btn-primary,
.btn-primary:hover
 {
   text-shadow: 0 -1px 0 rgba(0, 0, 0, 0.25);
   color: #ffffff;
}

.btn-primary.active
 {
   color: rgba(255, 255, 255, 0.75);
}

.btn-primary
 {
   background-color: #4a77d4;
   background-image: -moz-linear-gradient(top, #6eb6de, #4a77d4);
   background-image: -ms-linear-gradient(top, #6eb6de, #4a77d4);
   background-image: -webkit-gradient(linear, 0 0, 0 100%, from(#6eb6de), to(#4a77d4));
   background-image: -webkit-linear-gradient(top, #6eb6de, #4a77d4);
   background-image: -o-linear-gradient(top, #6eb6de, #4a77d4);
   background-image: linear-gradient(top, #6eb6de, #4a77d4);
   background-repeat: repeat-x;
   filter: progid: dximagetransform.microsoft.gradient(startColorstr=#6eb6de,
endColorstr=#4a77d4, GradientType=0);
```

```
    border: 1px solid #3762bc;
    text-shadow: 1px 1px 1px rgba(0, 0, 0, 0.4);
    box-shadow: inset 0 1px 0 rgba(255, 255, 255, 0.2), 0 1px 2px rgba(0, 0, 0, 0.5);
}

.btn-primary:hover,
.btn-primary:active,
.btn-primary.active,
.btn-primary.disabled,
.btn-primary[disabled]
 {
    filter: none;
    background-color: #4a77d4;
}

.btn-block
{
    width: 100%;
    display: block;
}

* {
    -webkit-box-sizing: border-box;
    -moz-box-sizing: border-box;
    -ms-box-sizing: border-box;
    -o-box-sizing: border-box;
    box-sizing: border-box;
}

* {
    box-sizing: border-box;
}

body
 {
    background-repeat: no-repeat;
    width: 100%;
```

```
    height: 100%;
    font-family: 'Open Sans', sans-serif;
    background: #092756;
    color: #fff;
    font-size: 18px;
    text-align: center;
    letter-spacing: 1.2px;
    background: -moz-radial-gradient(0% 100%, ellipse cover, rgba(104, 128, 138, .4) 10%, rgba(138, 114, 76, 0) 40%), -moz-linear-gradient(top, rgba(57, 173, 219, .25) 0%, rgba(42, 60, 87, .4) 100%), -moz-linear-gradient(-45deg, #670d10 0%, #092756 100%);
    background: -webkit-radial-gradient(0% 100%, ellipse cover, rgba(104, 128, 138, .4) 10%, rgba(138, 114, 76, 0) 40%), -webkit-linear-gradient(top, rgba(57, 173, 219, .25) 0%, rgba(42, 60, 87, .4) 100%), -webkit-linear-gradient(-45deg, #670d10 0%, #092756 100%);
    background: -o-radial-gradient(0% 100%, ellipse cover, rgba(104, 128, 138, .4) 10%, rgba(138, 114, 76, 0) 40%), -o-linear-gradient(top, rgba(57, 173, 219, .25) 0%, rgba(42, 60, 87, .4) 100%), -o-linear-gradient(-45deg, #670d10 0%, #092756 100%);
    background: -ms-radial-gradient(0% 100%, ellipse cover, rgba(104, 128, 138, .4) 10%, rgba(138, 114, 76, 0) 40%), -ms-linear-gradient(top, rgba(57, 173, 219, .25) 0%, rgba(42, 60, 87, .4) 100%), -ms-linear-gradient(-45deg, #670d10 0%, #092756 100%);
    background: -webkit-radial-gradient(0% 100%, ellipse cover, rgba(104, 128, 138, .4) 10%, rgba(138, 114, 76, 0) 40%), linear-gradient(to bottom, rgba(57, 173, 219, .25) 0%, rgba(42, 60, 87, .4) 100%), linear-gradient(135deg, #670d10 0%, #092756 100%);
    filter: progid: DXImageTransform.Microsoft.gradient( startColorstr='#3E1D6D', endColorstr='#092756', GradientType=1);
}

.login
 {
    width: 35%;
    margin: 1rem auto;
}

h1
 {
    color: #fff;
    font-size: 4rem;
    text-shadow: 0 0 10px rgba(0, 0, 0, 0.3);
```

```
    letter-spacing: 1px;
    text-align: center;
    padding: 0;
    margin-top: 10px;
}

input
 {
    text-align: center;
    width: 100%;
    margin-bottom: 1rem;
    background: rgba(0, 0, 0, 0.3);
    border: none;
    outline: none;
    padding: 10px;
    font-size: 13px;
    color: #fff;
    text-shadow: 1px 1px 1px rgba(0, 0, 0, 0.3);
    border: 1px solid rgba(0, 0, 0, 0.3);
    border-radius: 4px;
    box-shadow: inset 0 -5px 45px rgba(100, 100, 100, 0.2), 0 1px 1px rgba(255, 255, 255,
0.2);
    -webkit-transition: box-shadow .5s ease;
    -moz-transition: box-shadow .5s ease;
    -o-transition: box-shadow .5s ease;
    -ms-transition: box-shadow .5s ease;
    transition: box-shadow .5s ease;
}

input:focus
 {
    box-shadow: inset 0 -5px 45px rgba(100, 100, 100, 0.4), 0 1px 1px rgba(255, 255, 255,
0.2);
}

@media screen and (max-width: 700px)
{
```

```
   .login
 {
      width: 60%;
   }
   h1
 {
      font-size: 2rem;
   }
}
```

APP.PY:

```
from flask import Flask,render_template,request
import pickle
import numpy as np
model=pickle.load(open('Global_active_power.pkl','rb'))
app=Flask(_name_)
@app.route('/')
def y_pred():
   return render_template("index.html")
@app.route('/login',methods=["POST"])
def func2():
   Global_reactive_power=request.form['Global_reactive_power']
   Voltage=request.form['Voltage']
   Global_intensity=request.form['Global_intensity']
   Sub_metering_1=request.form['Sub_metering_1']
   Sub_metering_2=request.form['Sub_metering_2']
   Sub_metering_3=request.form['Sub_metering_3']

data=[[float(Global_reactive_power),float(Voltage),float(Global_intensity),float(Sub_mete
ring_1),float(Sub_metering_2),float(Sub_metering_3)]]
   pred=model.predict(data)
```

```
    print(pred[0])
    output = np.round(pred[0],2)
    output = str(output[0])+'KWh'
    return render_template("index.html",y="Global_active_power is {}".format(output))
if _name=='main_':
   app.run(debug= True)
```