**POWER CONSUMPTION ANALYSIS FOR HOUSEHOLDS USING IBM WASTON MACHINE LEARNING**

**INTRODUCTION:**

**OVERVIEW:**

Electricity sector in India. 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. 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. This Guided Project mainly focuses on applying a machine-learning algorithm to calculate the power consumed by all appliances.  This will help you track the power consumed on regular intervals for all kinds of appliances which use heavy loads such as Air Conditioners, Oven or a washing machine etc.

**PURPOSE:**

Electricity in used in almost all homes, the typical U.S household now uses more air conditioning appliances, and consumer electronics than ever before. However, average annual income site energy use per home has declined. The reasons for this declined include:

Improved efficiencies of heating and cooling equipment, water heaters, refrigerators, lighting and appliances.

Population migration to regions with lower heating and thus lower total energy demand.

**LITERATURE SURVEY**

**Existing problem:**

Given the rise of smart electricity meters and the wide adoption of electricity generation technology like solar panels, there is a wealth of electricity usage data available.

This data represents a multivariate time series of power-related variables that in turn could be used to model and even forecast future electricity consumption.

In this, you will discover a household power consumption dataset for multi- step time series forecasting and how to better understand the raw data using exploratory analysis.

**PROPOSED SOLUTION:**

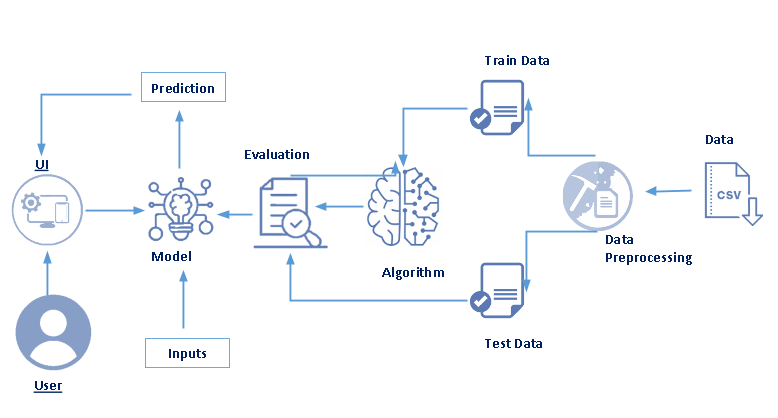
* The household power consumption dataset that describes electricity usage for a single house over four years.
* How to explore and understand the dataset using a suite of lines plots for series of data and histogram for the data distributions.
* How to understand the problem of considering different data framings of the prediction problem, ways the data may be prepared, and modeling methods that may be used.

**FEATURES OF OUR PROJECT:**

* The household power consumption dataset that describes electricity usage for a single house over four years.
* To explore and understand the dataset using a suite of lines plots for series of data and histogram for the data distributions.
* To understand the problem of considering different data framings of the prediction problem, ways the data may be prepared, and modeling methods that may be used.

**THEORITICAL ANALYSIS**

**Block Diagram:**

****

**Hardware / Software Designing:**

**Software Designing:**

* Numpy and Pandas: Open source data analysis and manipulation tool, built on top of the Python programming language.
* Matplotlib and Seaborn: Used for visualization with python.
* The finalized model is now to be saved, we will be saving the model as a pickle or pkl file.
* HTML pages “pca.html” for our home page and “result.html” which comes to use when we print out the final predictions made, both of these are stored in the templates folder.
* Let us build app.py flask file which is a web framework written in python for server-side scripting. Let’s see step procedure for building the backend application import required libraries.
* Configure app.py to fetch the user inputs from UI, process the values, and return the prediction.

**EXPERIMENTAL INVESTIGATION:**

The household power consumption dataset is a multivariate time series dataset that describes the electricity consumption for a single household over four years.

The dataset was collected between December 2006 and November 2010 and observations of power consumptions within the household were collected every minute.

It is a multivariate series comprised of seven variables (besides the data and time), they are:

* **global active power**: The total active power consumed by the household (kilowatts).
* **global reactive power**: The total reactive power consumed by the household (kilowatts).
* **Voltage:** Average voltage(volts).
* **global intensity:** Average current intensity (amps).
* **sub\_metering\_1:**Active energy for kitchen(watt-hours of active energy).
* **sub\_metering\_2:**Active energy for laundry(watt-hours of active energy).
* **sub\_metering\_3:**Active energy for climate control systems(watt-hours of active energy).

Active and reactive energy refer to the technical details of alternative current .In general terms, the reactive energy is the real power consumed by the household, whereas the reactive energy is the unused power in lines.

We can see that the dataset provides the active power as well as some division of the active power by main circuit in the house, specifically the kitchen, laundry and climate control. These are not all the circuits in the household.

The remaining watt-hours can be calculated from the active energy by first converting the active energy to watt-hours then subtracting the other sub-metered active energy in watt-hour.

**Python(source code):**

from flask import Flask,request,render\_template

import numpy as np

import pandas as pd

import pickle

import os

app = Flask(\_\_name\_\_)

model = pickle.load(open('PCA\_model.pkl', 'rb'))

@app.route('/')

def home():

return render\_template("pca.html")

@app.route('/predict',methods=["POST","GET"])

def predict():

input\_features = [float(x) for x in request.form.values()]

features\_value = [np.array(input\_features)]

features\_name = ['Global\_reactive\_power', 'Global\_intensity', 'Sub\_metering\_1',

'Sub\_metering\_2', 'Sub\_metering\_3']

df = pd.DataFrame(features\_value, columns=features\_name)

output = model.predict(df)

return render\_template('html code.html', prediction\_text=output)

if \_\_name\_\_=="\_\_main\_\_":

#port = int(os.getenv('PORT', 8080))

#app.run(host='0.0.0.0', port=port, debug=False)

app.run(debug=False)

**Jupyter notebook (source code):**

1. Importing Libraries:

import numpy as np

import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns

1. Importing Dataset

dataset=pd.read\_csv(r"C:\Users\Acer\Desktop\project\Book1 con.csv")

3.Understanding dataset

dataset.shape

dataset.head()

dataset.tail()

print(f"The Dataset has {dataset.shape[0]} rows and {dataset.shape[1]} columns")

dataset.columns

dataset.isnull().any()

dataset.isnull().sum()

dataset.drop(["Date"],axis=1,inplace=True)

dataset1 = dataset.drop(["Time"],axis=1,inplace=True)

percent\_missing = dataset.isnull().sum() \* 100 / len(dataset)

missing\_value\_df = pd.DataFrame({'percent\_missing': percent\_missing})

missing\_value\_df

dataset.loc[dataset.Sub\_metering\_3.isnull()].head()

dataset.replace('?',np.nan, inplace=True)

dataset.loc[dataset.Sub\_metering\_3.isnull()].head()

dataset = dataset.dropna(how = 'all')

for i in dataset.columns:

dataset[i] = dataset[i].astype('float64')

dataset.shape

values = dataset.values

dataset['Sub\_metering\_4'] = (values[:,0] \* 1000 / 60) - (values[:,4] + values[:,5] + values[:,6])

sns.distplot(dataset['Global\_active\_power'])

sns.distplot(dataset['Global\_active\_power'],kde=False,bins=30)

sns.distplot(dataset['Global\_reactive\_power'],kde=False,bins=30)

sns.distplot(dataset['Voltage'],kde=True,bins=30)

sns.distplot(dataset['Global\_intensity'],kde=True,bins=30)

dataset.corr()

pearson = dataset.corr(method='pearson')

mask = np.zeros\_like(pearson)

mask[np.triu\_indices\_from(mask)] = True

sns.heatmap(pearson, vmax=1, vmin=0, square=True, cbar=True, annot=True, cmap="YlGnBu", mask=mask);

sns.jointplot(x='Global\_reactive\_power', y='Global\_active\_power', data=dataset, kind='scatter')

sns.jointplot(x='Voltage', y='Global\_active\_power', data=dataset, kind='scatter')

sns.jointplot(x='Global\_intensity', y='Global\_active\_power', data=dataset, kind='scatter')

sns.jointplot(x='Sub\_metering\_1', y='Global\_active\_power', data=dataset, kind='scatter')

sns.jointplot(x='Sub\_metering\_2', y='Global\_active\_power', data=dataset, kind='scatter')

dataset.columns

dataset.drop(["Voltage"],axis=1,inplace=True)

dataset.columns

dataset.drop(["Sub\_metering\_4"],axis=1,inplace=True)

dataset.shape

#independent and dependent variables

x=dataset.iloc[:,1:]

y=dataset.iloc[:,0:1]

x.shape

y.shape

x.head()

x.tail()

x=dataset.iloc[:,1:].values

y=dataset.iloc[:,0:1].values

x

x.ndim

x.shape

y

y.ndim

y.shape

from sklearn.model\_selection import train\_test\_split

x\_train,x\_test,y\_train,y\_test=train\_test\_split(x,y,test\_size=0.2,random\_state=101)

x\_train.shape

x\_test.shape

y\_train.shape

from sklearn.linear\_model import LinearRegression

lr = LinearRegression()

lr.fit(x\_train,y\_train)

y\_pred=lr.predict(x\_test)

y\_pred

y\_test

from sklearn.metrics import r2\_score

accuracy = r2\_score(y\_test,y\_pred)

accuracy\*100

from sklearn import metrics

print('MAE:',metrics.mean\_absolute\_error(y\_test,y\_pred))

print('MSE:',metrics.mean\_squared\_error(y\_test,y\_pred))

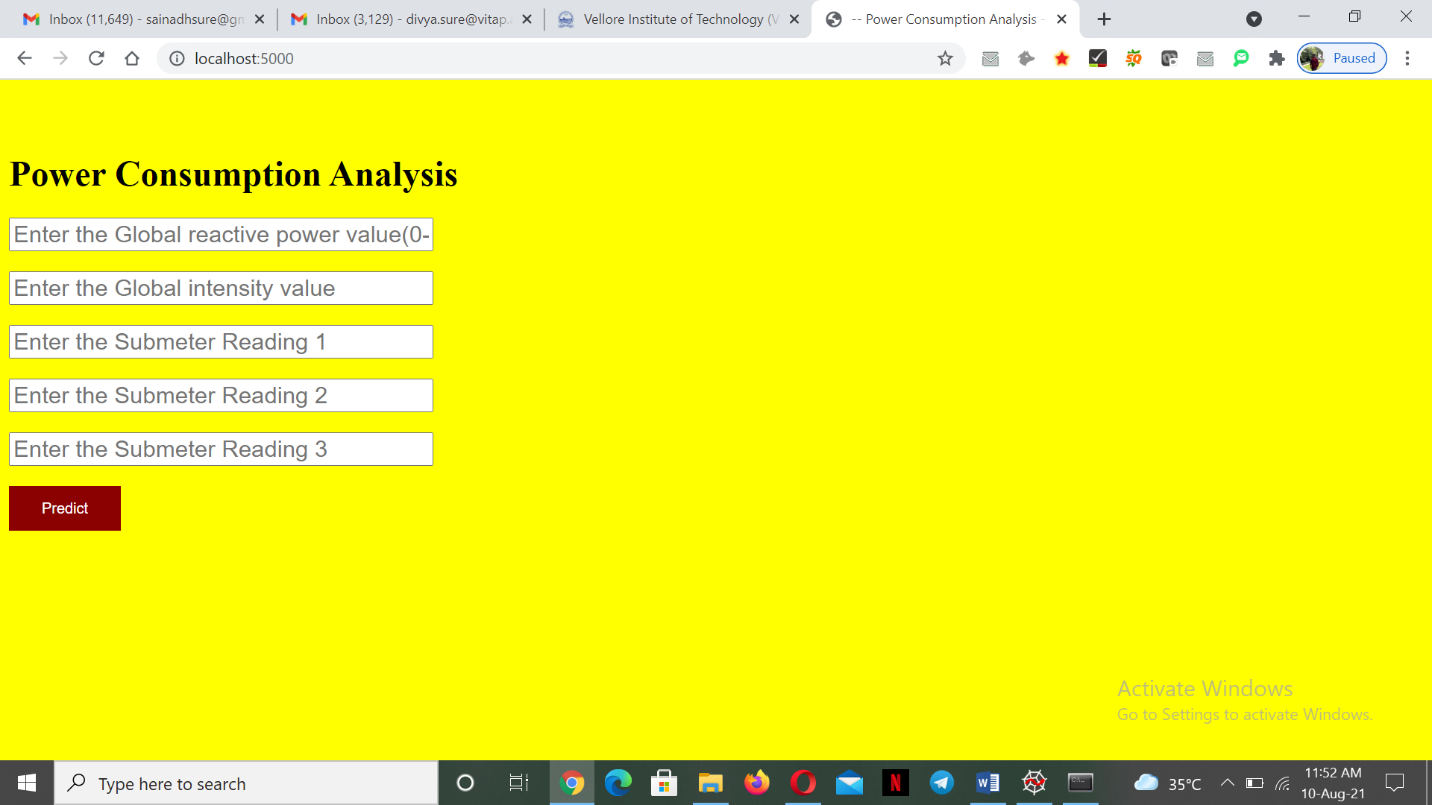
print('RMSE:',np.sqrt(metrics.mean\_squared\_error(y\_test,y\_pred)))

print('R squares value:',metrics.r2\_score(y\_test,y\_pred))

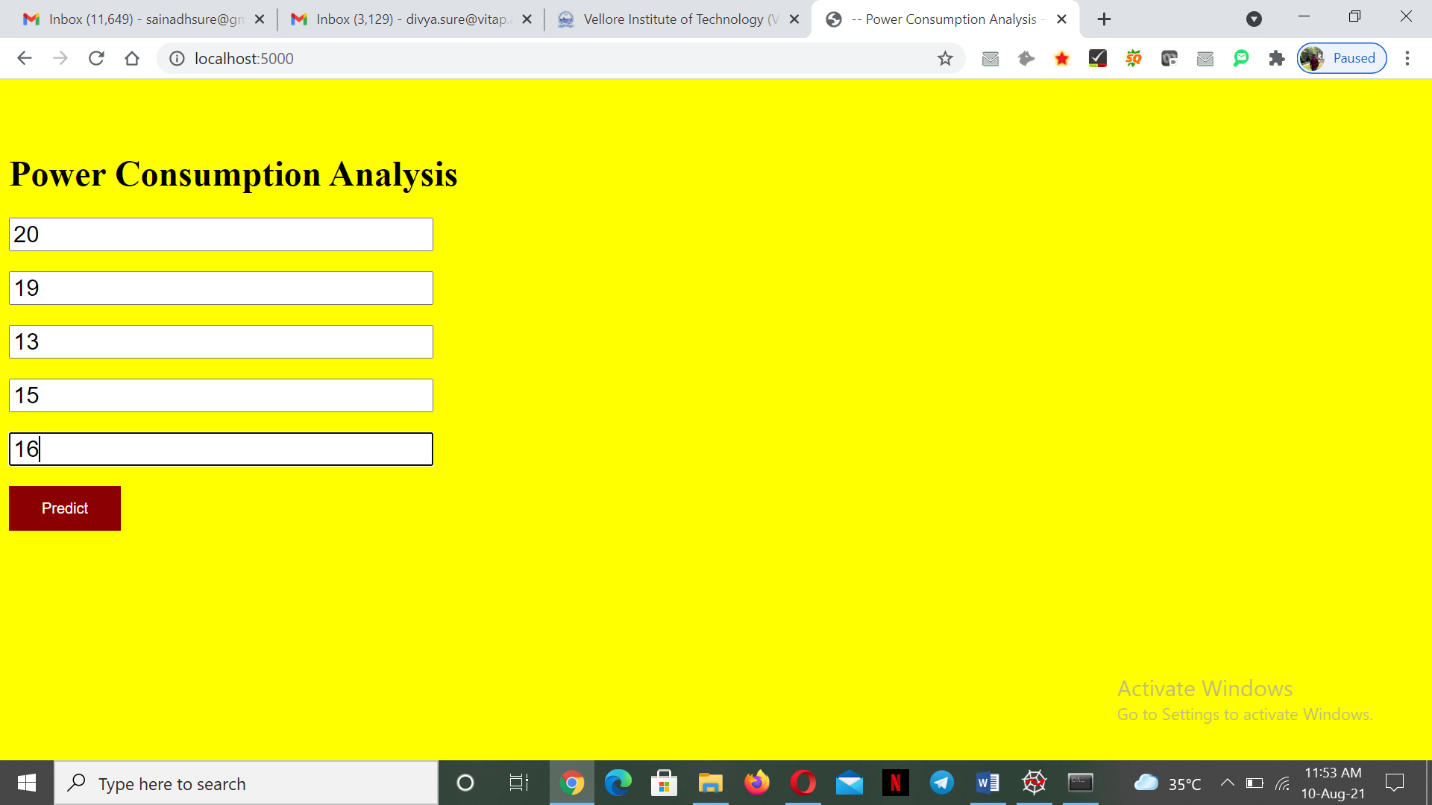
import pickle

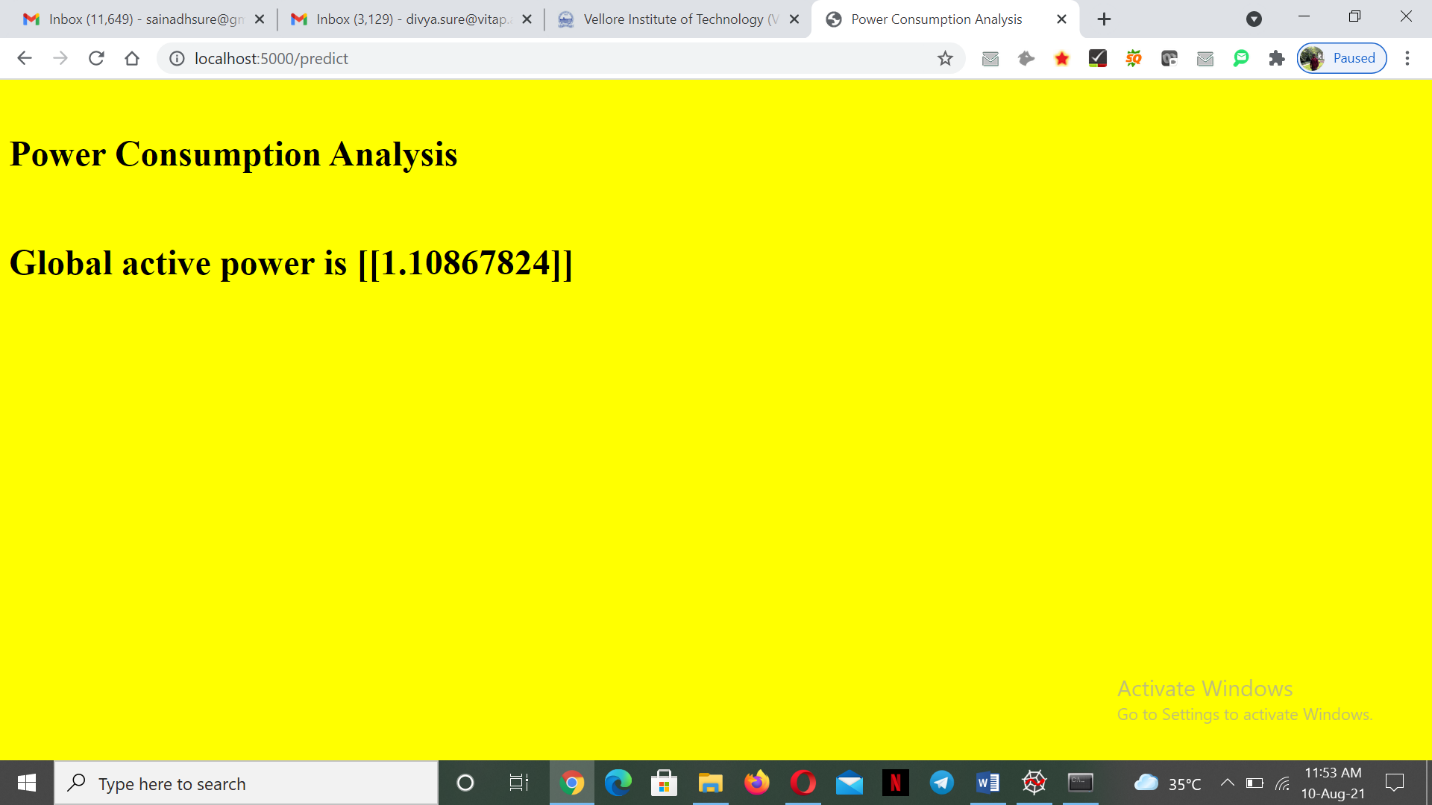
filename='PCA\_model.pkl'

pickle.dump(lr,open(filename,'wb'))

**Output:**

**Enter the values**

****

****

**Result:**

* Open the anaconda prompt from the start menu.
* Navigate to the folder where your app.py resides.
* Now type “python app.py” command.
* It will show that local host where your app is running on <http://127.0.0.1:5000/>
* Copy that local host URL and open that URL in the browser. It does navigate me to where you can view your web page.
* Enter the value, Click on the predict button and see the result prediction on webpage.

**Applications:**

1. To assess the energy efficiency and energy usage of their homes.
2. To use the new understanding of the problem to consider different framings of the prediction problem, ways the data may be prepared and modeling methods that may be used.
3. The household power consumption dataset is a multivariate time series dataset that describes the electricity consumption for a single household over years.

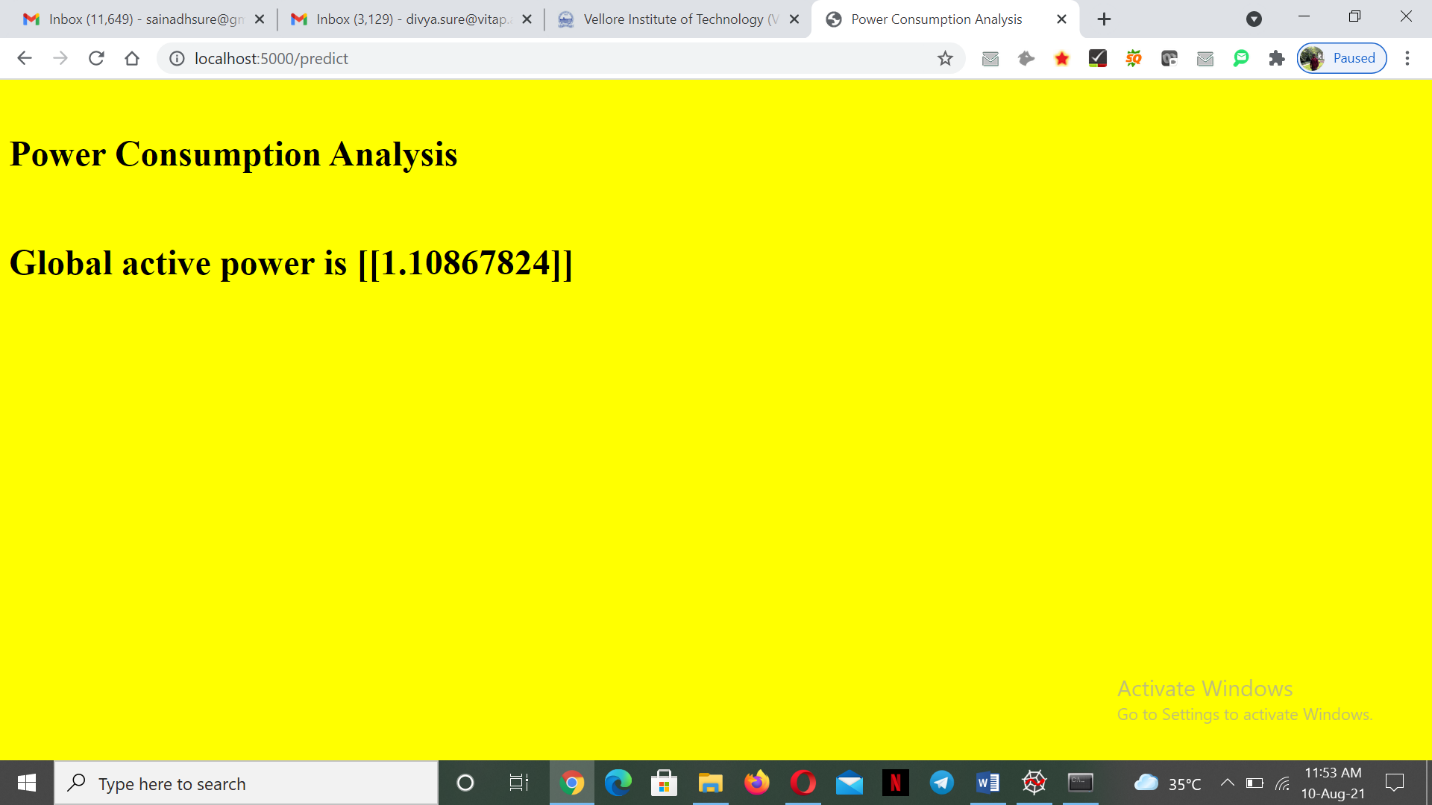
**CONCLUSION:**

Finally, total power consumption by all the appliances is calculated and displayed.

**BIBLOGRAPHY:**

<https://machinelearningmastery.com/how-to-load-and-explore-household-electricity-usage-data/>

**OUTPUT:**

****