Power Consumption Analysis For House Holds Using ML

INDEX

1.INTRODUCTION

a.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

b.Purpose:

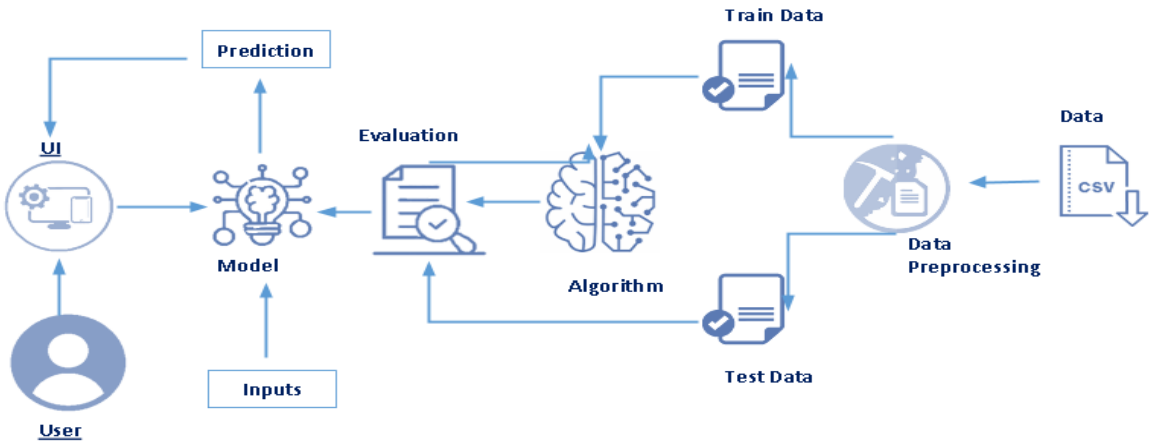
We'll be able to understand the problem to classify if it is a regression or a classification kind of problem. We will be able to know how to pre-process/clean the data using different data preprocessing techniques. You will be able to analyze or get insights into data through visualization. Applying different algorithms according to the dataset and based on visualization. We will be able to know how to build a web application using the Flask framework.

2.LITERATURE SURVEY

To calculate the power consumed by all appliances. To solve this problem, we use linear regression machine learning algorithm.

3.THEORITICAL ANALYSIS

3.1 Block Diagram:



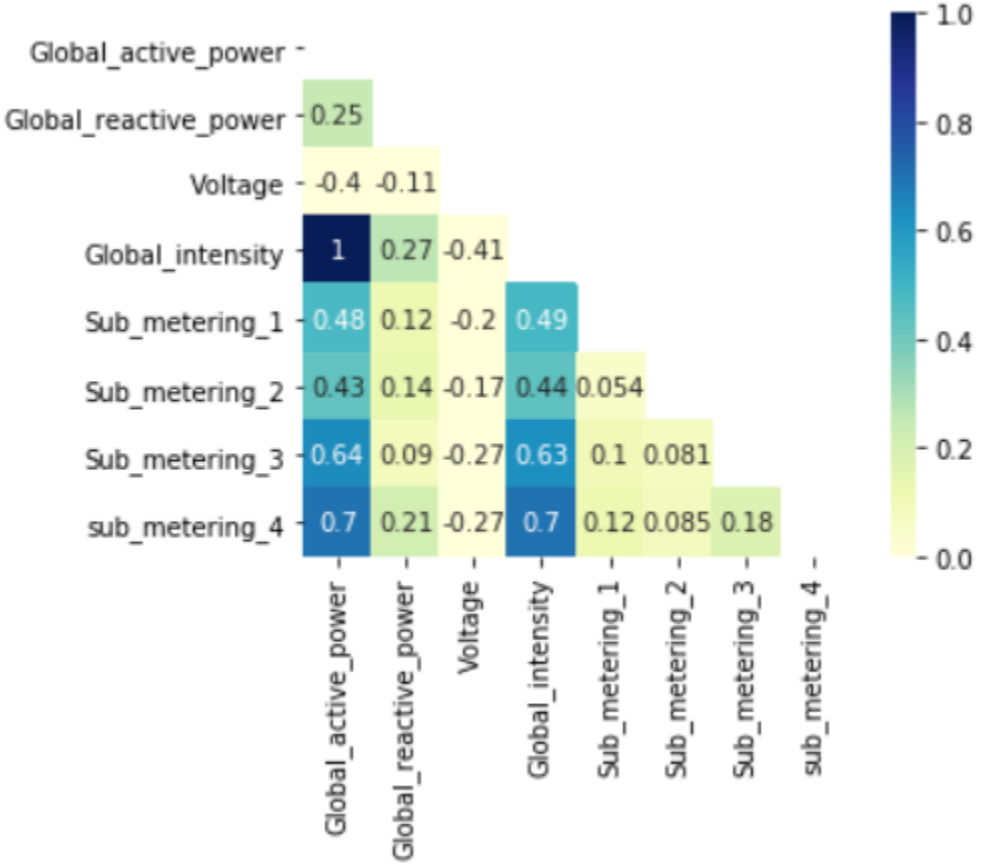
Hardware / Software designing

1. Dataset: Kaggle
2. IDE: Jupyter Notebook, Spyder,Anaconda navigator
3. Programming Languages (Back-end): Python 3.10
4. Front-End:HTML,CSS
5. Framework: Flask

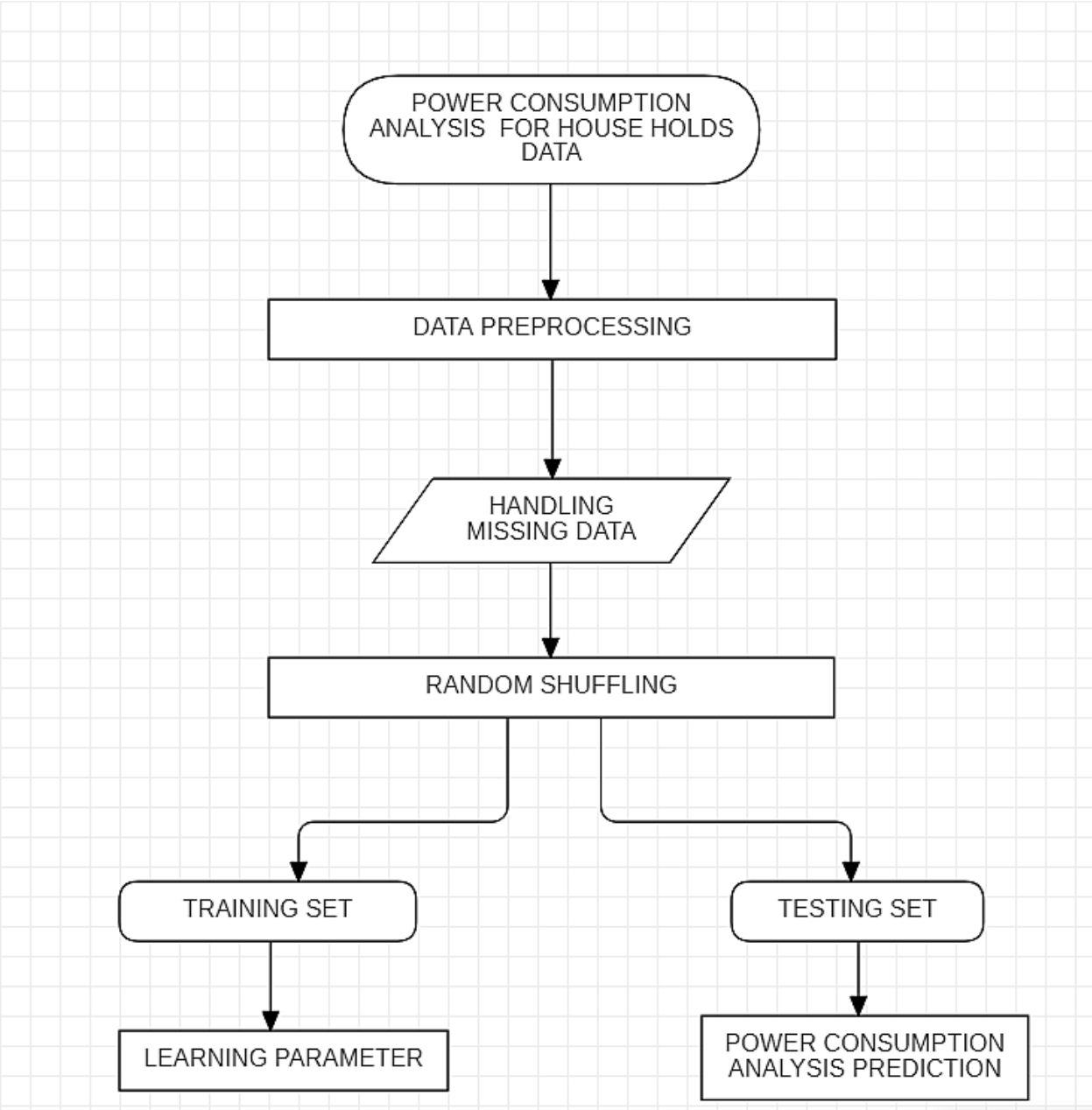
4.EXPERIMENTAL INVESTIGATIONS

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

Heatmap is the plot of values of correlation between the variables in a data set. The correlation values are plotted using heatmap.



5.FLOWCHART


POWER CONSUMPTION ANALYSIS FOR HOUSE HOLDS DATA
DATA PREPROCESSING
HANDLING MISSING DATA
RANDOM SHUFFLING
TRAINING SET
TESTING SET
LEARNING PARAMETER
POWER CONSUMPTION ANALYSIS PREDICTION

6.RESULT

Execute the python code and after the module is running, open index.html page and scroll down to find the buttons to test with.

- 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 the 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 values, click on the predict button and see the result/prediction on the web page.

7. ADVANTAGES AND DISADVANTAGES

a. advantages :

1. The interface is user friendly and easy to use and understand even to a person who has very little/no knowledge for the same.
2. The model uses a huge amount of data and the results generated are accurate and reliable.
3. Easy to use & has a user-friendly interface.
4. Results can be improved by training data to our choice of parameter.
5. The users can easily tell the amount of house holds power consumption

b. disadvantages :

1. Some complex integrations of services are required.
2. No free server available on IBM Cloud for deploying Backend

8. APPLICATIONS

Household-Power-Consumption-Analysis Project was done to understand the advantages of big data applications. Analyzed the amount of energy consumed in a household which is given to us as a timeseries, and our objective is to derive patterns from the obtained real time data. Imported data into Databricks Azure.

9.Conclusion

By the end of the project, I have understood that

- I have understood the problem to classify if it is a regression or a classification kind of problem.
- I can know how to pre-process/clean the data using different data pre-processing techniques.
- Applying different algorithms according to the dataset
- I can know how to find the accuracy of the model.

I can build web applications using the Flask framework

10.FUTURE SCOPE

1. Despite our model giving good results, we can add robustness to it by making it do the predictions for a greater time in the future.
2. We can add a feature that is the user can add a specific watt's to be consumed so that if it is reached they will get a message.

11.BIBILOGRAPHY:

Data set : https://www.kaggle.com/datasets/uciml/electric-power-consumption-data-set
Book : Python Machine Learning: Unlock deeper insights

12.APPENDIX

a.source code:

```
import pandas as pd
import numpy as np
import seaborn as sn
import matplotlib.pyplot as plt
dataset=pd.read_csv("C:\households\Household_power_consumption.csv\Household_power_consumption.csv")
dataset.head()
dataset.tail()
print(f"the dataset has {dataset.shape[0]} rows and {dataset.shape[1]} columns")
dataset.columns
dataset.describe()
dataset.info()
dataset.head()
dataset.isnull().sum()
percent_missing = dataset.isnull().sum() * 100 / len(dataset)
missing_value_df = pd.DataFrame({'percent_missing': percent_missing})
missing_value_df
dataset.describe()
sn.distplot(dataset['Global_active_power'])
sn.distplot(dataset['Global_active_power'],kde=False,bins=30)
sn.distplot(dataset['Global_reactive_power'],kde=False,bins=30)
sn.distplot(dataset['Voltage'],kde=True,bins=30)
sn.distplot(dataset['Global_intensity'],kde=False,bins=30)
dataset.corr()
pearson=dataset.corr(method='pearson')
mask=np.zeros_like(pearson)
mask[np.triu_indices_from(mask)]=True
sn.heatmap(pearson,vmax=1,vmin=0,square=True,cbar=True,annot=True,cmap="YlGnBu",mask=mask);
sn.jointplot(x='Global_reactive_power',y='Global_active_power',data=dataset,kind='scatter')
sn.jointplot(x='Voltage',y='Global_active_power',data=dataset,kind='scatter')
```

```
sn.jointplot(x='Global_intensity',y='Global_active_power',data=dataset,kind='sc
atter')
sn.jointplot(x='Sub_metering_1',y='Global_active_power',data=dataset,kind='sca
tter')
sn.jointplot(x='Sub_metering_2',y='Global_active_power',data=dataset,kind='sc
atter')
sn.jointplot(x='Sub_metering_3',y='Global_active_power',data=dataset,kind='sc
atter')
dataset.datetime=pd.to_datetime(dataset.datetime)
dataset.set_index("datetime",inplace=True)
x=dataset.iloc[:,[1,3,4,5,6]]
y=dataset.iloc[:,1]
x.head()
y.head()
from sklearn.model_selection import train_test_split
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.3,random_state=100)
print(x_train.shape)
print(x_test.shape)
print(y_train.shape)
print(y_train.shape)
from sklearn.linear_model import LinearRegression
lm=LinearRegression()
lm.fit(x_train,y_train)
predictions = lm.predict(x_test)
predictions
from sklearn import metrics
print('MAE:',metrics.mean_absolute_error(y_test,predictions))
print('MSE:',metrics.mean_squared_error(y_test,predictions))
print('RMSE:',np.sqrt(metrics.mean_squared_error(y_test,predictions)))
print('RSquarevalue:',metrics.r2_score(y_test,predictions))
import pickle
filename = 'PCASSS_model.pkl'
pickle.dump(lm,open(filename,'wb'))
```

b.UI output screenshot