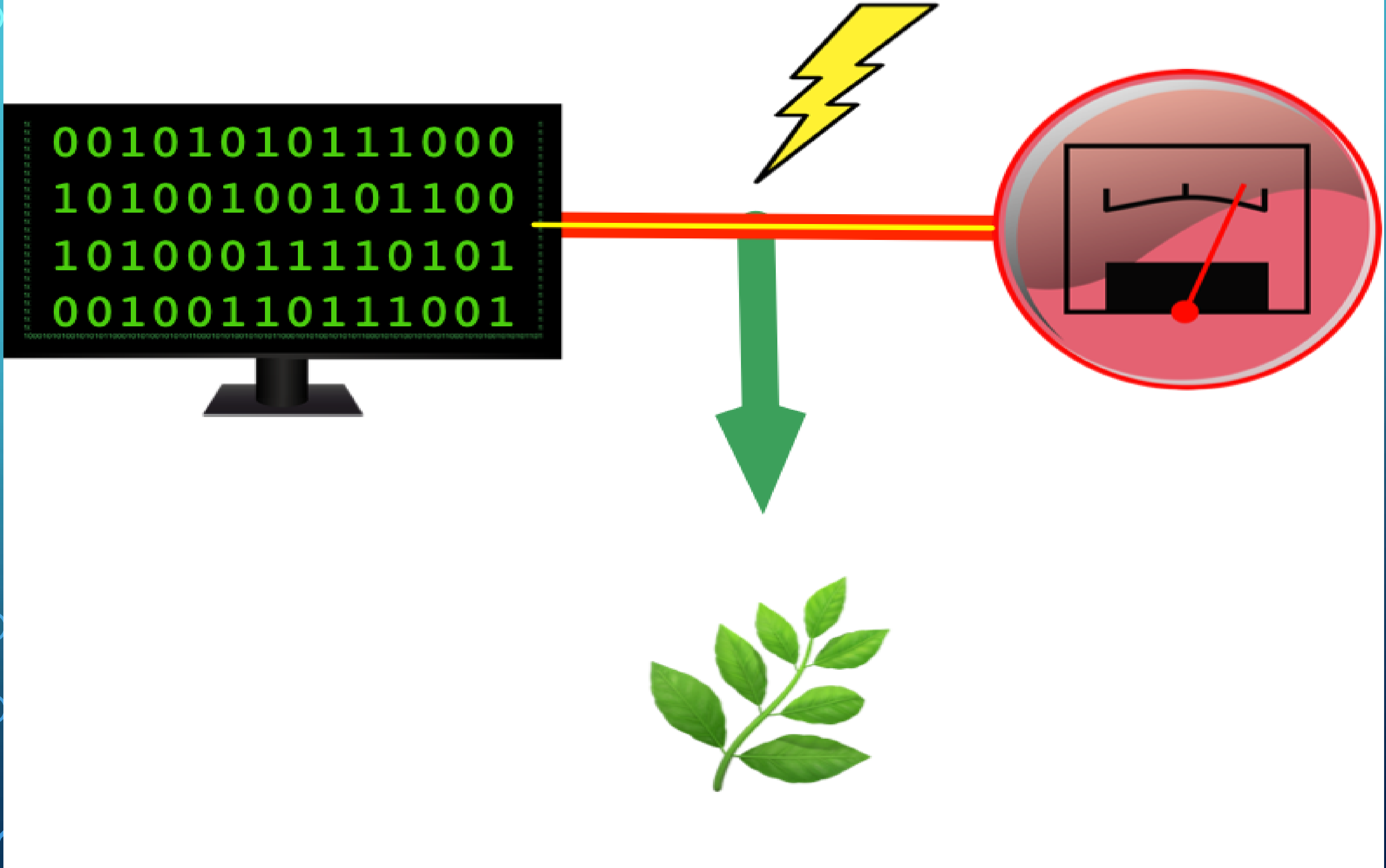
MEASURE ENERGY CONSUMPTION



Name: vedagiri S
Naan Mudhalvan ID: au513521104054
Email Id: vedagirisankar2004@gmail.com
College: ANNAI MIRA COLLEGE OF ENGINEERING AND TECHNOLOGY

INTRODUCTION

The concept of energy consumption is directly related to energy efficiency since higher consumption results in lower energy efficiency. It's estimated that during an hour about 1,000 watts are consumed, so this measure is used to calculate the consumption of homes, businesses, or any other type of building in order to issue the corresponding bills. There are various factors that directly influence energy consumption such as:

- The activity that takes place in the home or business.
- The number of people in a household or workers.
- The consumption habits of each person.
- The energy performance of household appliances.

With the right information and technology, it's possible to use energy more responsibly and efficiently. This results in a reduction in energy consumption and, therefore, in significant savings on utility bills.

DATASET

- we using dataset for the Energy Consumption analysis models the data set get from the Kaggle.
- To perform loading and preprocessing dataset get from the below link.
- Dataset: https://www.kaggle.com/c/ashrae-energy-prediction
- The data set could be in various formats. We take the dataset in csv file format.

LOADING DATASET

Let's use the electrical meter data to create clusters of typical load profiles for analysis. First we can load our conventional packages.

<u>Import relevant python packages:</u>

```
import pandas as pd
import matplotlib.pyplot as plt
import matplotlib
```

Next let's load all the packages we will need for analysis

```
import sklearn
from sklearn import metrics
from sklearn.neighbors import KNeighborsRegressor
from scipy.cluster.vq import kmeans, vq, whiten
from scipy.spatial.distance import cdist
import numpy as np
from datetime import datetime
```

DUPLICATE VALUES

Duplicate Values

Some datasets may include duplicate readings that share a common timestamp. The reason for this duplication should be investigated for each unique dataset, as reasons for data duplication can vary widely depending on the data collection methodology.

There are several approaches towards the handling of duplicate values. We have the option of discarding both or one of the duplicate values, or imputing the timestamp value using each of the available measurements. For our example, we will compute the mean energy consumption for each of our duplicate value pairs and use that value moving forward.

Once we have removed duplicate values, we manually set the frequency of the DatetimeIndex to hourly (‘H’). Normally, the date parser would be able to determine the frequency of the DatetimeIndex automatically. The presence of duplicate DatetimeIndex values prevents this from happening though, so the frequency must be set manually following removal of duplicate values. Setting the frequency now will help us avoid problems with plotting and calculations down the line.

```
# Identify Duplicate Indices
duplicate_index = duq_df[duq_df.index.duplicated()]
print(duq_df.loc[duplicate_index.index.values, :])# Replace Duplicates with Mean Value
duq_df = duq_df.groupby('Datetime').agg(np.mean)# Set DatetimeIndex Frequency
duq_df = duq_df.asfreq('H')
```

MISSING VALUES

A quick search through the dataset indicates that there are 24 missing values across our date range. We will use mean interpolation to impute these missing values. We can do this using the interpolate() method on our dataframe object:

```
# Determine # of Missing Values
print('# of Missing DUQ_MW Values:
{}'.format(len(duq_df[duq_df['DUQ_MW'].isna()])))
# Impute Missing Values
duq_df['DUQ_MW'] =
duq_df['DUQ_MW'].interpolate(limit_area='inside', limit=None)
```

VISUALIZING ENERGY CONSUMPTION DATA

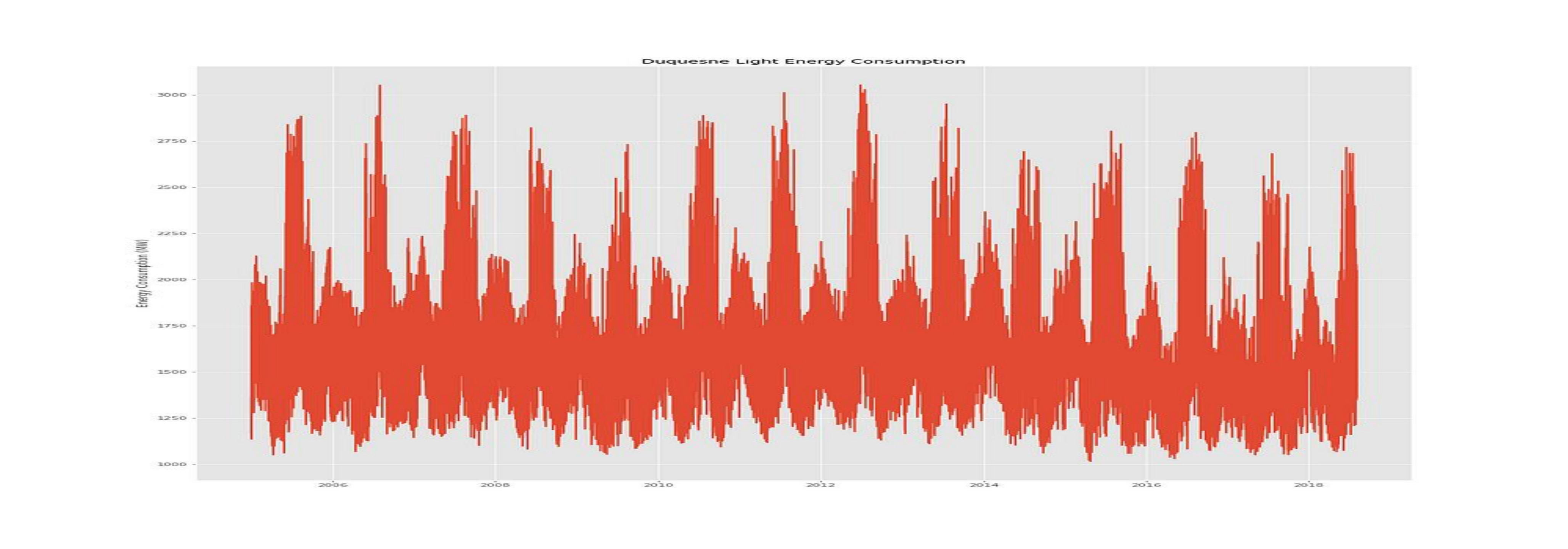
FIRST, LET'S LOOK AT A SIMPLE TIME SERIES PLOT OF OUR DATA:

```
PLT.PLOT(DUQ_DF.INDEX, DUQ_DF['DUQ_MW'])

PLT.TITLE('DUQUESNE LIGHT ENERGY CONSUMPTION')

PLT.YLABEL('ENERGY CONSUMPTION (MW)')

PLT.SHOW()
```



CREATE TRAIN AND TEST DATASETS

THE MODEL IS GIVEN A SET OF DATA THAT WILL BE USED TO **TRAIN** THE MODEL TO PREDICT A SPECIFIC OBJECTICE. IN THIS CASE, WE WILL USE A FEW SIMPLE TIME SERIES FEATURES AS WELL AS OUTDOOR AIR TEMPERATURE TO PREDICT HOW MUCH ENERGY A BUILDING USES.

1.TRAINING_MONTHS = [4,5,6]

TEST_MONTHS = [7]

2.TRAININGDATA = OFFICE_EXAMPLE_PREDICTION_DATA[OFFICE_EXAMPLE_PREDICTION_DATA.INDEX.MONTH.ISIN(TRAINING_MONTHS)]

TESTDATA = OFFICE_EXAMPLE_PREDICTION_DATA[OFFICE_EXAMPLE_PREDICTION_DATA.INDEX.MONTH.ISIN(TEST_MONTHS)]

```
TRAININGDATA.INFO()

<CLASS 'PANDAS.CORE.FRAME.DATAFRAME'>

DATETIMEINDEX: 2184 ENTRIES, 2017-04-01 00:00:00 TO 2017-06-30 23:00:00

DATA COLUMNS (TOTAL 1 COLUMNS):

 #   COLUMN                NON-NULL COUNT  DTYPE

---  ------------           ----------------------  --------

 0   PANTHER_OFFICE_HANNAH  2184 NON-NULL   FLOAT64

DTYPES: FLOAT64(1)

MEMORY USAGE: 34.1 KB
```

```
TESTDATA.INFO()

<CLASS 'PANDAS.CORE.FRAME.DATAFRAME'>

DATETIMEINDEX: 744 ENTRIES, 2017-07-01 00:00:00 TO 2017-07-31 23:00:00

DATA COLUMNS (TOTAL 1 COLUMNS):

 #   COLUMN                 NON-NULL COUNT  DTYPE

---  ------------           ------------------------  -------

 0   PANTHER_OFFICE_HANNAH  744 NON-NULL    FLOAT64

DTYPES: FLOAT64(1)

MEMORY USAGE: 11.6 KB
```

CONCLUSION

- Data-Driven Insights: Machine learning algorithms effectively analyzed vast datasets, providing nuanced insights into energy consumption patterns. This data-driven approach has the potential to uncover hidden inefficiencies and anomalies that were previously challenging to detect.
- Predictive Capabilities: The predictive models developed in this project enable real-time forecasting of energy consumption, allowing organizations to make informed decisions and adjustments proactively.
- Cost Savings: By identifying energy-saving opportunities and enabling dynamic load management, machine learning can lead to substantial cost savings, making it a valuable investment for businesses and homeowners alike.
- Environmental Impact: Reducing energy consumption not only cuts costs but also contributes to a greener, more sustainable future by lowering carbon emissions and environmental impact.
- Scalability: Machine learning models can be scaled to accommodate various industries, from smart homes to large industrial complexes, making the approach versatile and adaptable to different contexts.
- Recommendations: As a result of this project, we recommend the integration of machine learning-based energy management systems in both residential and commercial settings, along with continuous model refinement and data collection.

In essence, this project has showcased the potential of machine learning in transforming the way we understand and manage energy consumption. The application of machine learning in this domain offers not only cost-effective solutions but also a means to support environmental sustainability goals. Its scalability and adaptability make it a powerful tool for organizations and individuals seeking to optimize energy usage.