

SB3001- PROJECT BASED EXPERIENTIAL LEARNING PROGRAM

NAAN MUDHALVAN REPORT

Date	01/11/2023
Name	J.Jancy
Team ID	Proj_212174_Team_1
Project Name	Measure Energy Consumption
Maximum Marks	

Acknowledgement:

The project Measure energy consumption have been discussed among the team members and brainstormed with different ideas. The project work were equally distributed among the team mates and each content and work are from different team members and their effort. During the phase 1 project Brainstroming and Idea Generation , Defining the problem statements and empathy map are the work done by the team members. In phase 2 project the problem codes and their output were programmed and results were obtained. In phase 3 project Data Virtualization were implemented for the given datasets and plotting graphs for the dataset used and constructing the architecture for the code used and their algorithm. In phase 4 module was developed according to the dataset and the constrains obtained for the project. In phase 5 the overall view and solutions , ideas done , abstract of the project , introduction about the project , Literature survey , problem definition , design thinking , Innovating problem solving , importing the datasets and performing data learning and analysis , data visualization , model development and evaluation , code sample, output screenshot and concluding the entire project work some reference from the researches done by the team. Team divided the project into groups so that the researchs can be done easier by separating them into groups. This project has helped us to learn more about the how the energy is consumed in each year by the countries and loss of energy resources, we have achived a idea to reduce the energy consumption in each houses and saving the energy resources. We would also like to thank all the mentors, staffs for helping us providing support and help when we had some doubts about how to start the program and whats the main idea of the project. Proper instructions and helpful ideas about the project were taken place with the help of mentors. The helpful informations uploaded by the mentors are into account and were helpful for the project development. And at last we would like to thank all the team members for all their hard work and efforts they have put in this project and for their time and dedication and for supporting the team when someone needs help , and the involvement of the team to complete the project.

Once again we thank all the people who have supported us in completeing this project.

Index:

S.No	Content	Page No
1.	List of Figures & tables	4
2.	Abstract	4
3.	Introduction	5-6
4.	Literature Survey	7-9
5.	Problem definition,design thinking,Innovation&problem solving	9
6.	Importing dataset&performing data cleaning &Analysis	10-14
7.	Data visualization	14-16
8.	Model Development&Evaluation	16-19
9.	Code Sample	19-20
10.	Output Screenshot	21
11.	Conclusion&Future Enhancement References	22

1)List Of Figures & Tables:

Figure	Page number
Fig-1	21
Fig-2	21
Table	Page Number
Table-1	9

2)Abstract:

This abstract provides an overview of the research conducted to investigate various methods for measuring energy consumption and their practical applications. Energy consumption is a critical aspect in today's world, with growing concerns about sustainability and efficient resource utilization. This study aims to contribute to the understanding of energy measurement techniques, encompassing both traditional and advanced approaches.

3)Introduction:

Measuring energy consumption using data sets is a critical component of understanding and managing energy usage in various sectors, including residential, commercial, and industrial settings. Data-driven approaches provide valuable insights into how energy is consumed, allowing businesses, individuals, and policymakers to make informed decisions, reduce energy costs, and minimize environmental impacts.

Data Collection:

To measure energy consumption, data is collected from various sources such as smart meters, sensors, and utility bills. These sources record information like electricity usage, gas consumption, and other energy-related parameters.

Data Preprocessing:

Raw data is often noisy and needs preprocessing to ensure accuracy. This involves data cleaning, error correction, and data normalization to make it suitable for analysis.

Data Analysis:

Energy consumption data sets are subjected to various analytical techniques, including statistical analysis, machine learning, and data visualization. These analyses help identify patterns, trends, and anomalies in energy consumption.

Predictive Modeling:

Machine learning models can be trained on historical energy consumption data to make predictions about future

consumption. This is especially valuable for demand forecasting and proactive energy management.

Energy Efficiency Programs:

Using data sets, energy providers and businesses can design and implement energy efficiency programs that target specific areas of high consumption. These programs can lead to reduced energy waste and carbon emissions.

Real-time Monitoring:

With the advent of smart technology, real-time energy consumption data can be monitored remotely, enabling immediate responses to unusual spikes in usage or equipment malfunctions.

Conservation Measures:

By analyzing energy consumption data, individuals and organizations can make informed decisions about energy conservation measures such as upgrading appliances, optimizing building designs, or implementing renewable energy solutions.

Policy and Regulation:

Governments and regulatory bodies can use energy consumption data to develop policies, regulations, and incentives to promote energy conservation, reduce greenhouse gas emissions, and support the transition to cleaner energy sources.

4)Literature survey:

A literature survey on the measurement of energy consumption provides valuable insights into the existing body of research and the various methods and technologies employed for this purpose. Here's a concise overview of key topics and notable works in the field:

1. Traditional Energy Measurement Techniques:

- **Utility Bills:** Traditional method based on monthly or quarterly utility bills to measure energy consumption at residential and commercial levels.
- **Submetering:** Submetering systems for tracking energy usage in specific areas within larger buildings, offering more detailed data.

2. Advanced Energy Measurement Techniques:

- **Smart Meters:** Electronic devices that record real-time energy consumption and provide insights into usage patterns. Notable works include studies on smart grid technology and its impact on energy management.

3. Energy Monitoring and Management Systems:

- **Building Energy Management Systems (BEMS):** Research on BEMS, which use sensors and controls to optimize energy usage in commercial and residential buildings.
- **Industrial Energy Management Systems (IEMS):** Studies on IEMS for optimizing energy consumption in industrial processes.

4. Data Analytics and Energy Consumption:

- Research on data analytics and machine learning techniques for analyzing energy consumption data to identify trends, anomalies, and optimization opportunities.

5. Renewable Energy Integration:

- -Studies on the measurement of energy consumption in the context of renewable energy sources, including the integration of solar panels and wind turbines.

6. Energy Efficiency and Conservation:

- Works focused on how accurate energy measurement contributes to energy conservation efforts, including demand response programs and load shifting strategies.

7. Environmental Impact:

- Research exploring the environmental consequences of energy consumption, including its role in reducing greenhouse gas emissions and mitigating climate change.

8. Blockchain-Based Energy Tracking:

- Emerging research on the use of blockchain technology for transparent and secure energy consumption tracking.

9. Future Trends and Challenges:

- An examination of the latest trends in energy measurement, including the development of smart appliances and the potential of 5G technology in energy monitoring.

Notable works, authors, and publications can be found within these subfields, helping researchers and practitioners

gain a deeper understanding of the state of the art in energy consumption measurement and management.

5)Problem Defining,Design

Thinking,Innovation&Problem Solving:

Statement:

Energy infrasture is inadequate to meet the growing demand for electricity.The environmental problems related to energy consumption include air pollution,water pollutionand solid waste disposal.Electrical shortages hurt industrial output.the one of the biggest problem is the green house gas emissions.

Problem Statement	I am	I am trying to	But	Because	Which makes me fell
PS-1	Electricity usage	use less amount of energy	The usage of energy is high	The use electical appliances	helpless
Ps-2	Vehicle usage	Less Use of energy	Vehicle usage	The vehicle is needed now a days	helpless

Table 1

6)Importing Data Set&Performing Data cleaning&Analysis:

Importing a dataset is a fundamental concept in data analysis and data science. It refers to the process of bringing external data into a data analysis environment or software tool so that you can work with and analyze the data. Here's an explanation of the theory behind importing a dataset:

Data Source:

Datasets can come from various sources, such as spreadsheets (e.g., Excel files), database (e.g., SQL databases), text files (e.g., CSV files), APIs (Application Programming Interfaces), web scraping, or other data storage systems. The first step is to identify and access the data source.

Data Format:

Datasets can be stored in various formats, and you need to understand the format of the data you are working with. Common data formats include CSV (Comma-Separated Values), JSON (JavaScript Object Notation), XML (eXtensible Markup Language), Excel, and many others. Understanding the format is crucial because it determines how you will import the data.

Data Analysis Environment:

To work with a dataset, you need to use a data analysis environment or software tool. Popular choices include Python with libraries like Pandas, R, SQL databases, Excel, and more. Different environments have their methods for importing data.

Import Methods:

Each data analysis environment provides methods or functions to import data. For example, in Python's Pandas library, you can use the `pd.read_csv()` function to import data from a CSV file.

Parameters and Options:

When importing data, you often need to specify parameters and options. For instance, you might need to specify the file path or URL, delimiter for CSV files, encoding type for text files, and more. These parameters ensure that the data is imported correctly.

Data Validation:

After importing, it's crucial to validate the data. Check for missing values, data types, and anomalies. Data cleansing and preprocessing might be required to make the data suitable for analysis.

Data Exploration:

Once the data is successfully imported and cleaned, you can explore it. This includes summary statistics, data visualization, and more in-depth analysis. Understanding the dataset's characteristics is essential for drawing meaningful insights.

Data Analysis and Visualization:

After importing and exploring the data, you can perform various data analysis tasks, such as regression, classification, clustering, or time series analysis. You can also create data visualizations to help convey insights to others.

Report or Decision Making:

Finally, the results of your analysis can be used to make informed decisions or create reports for stakeholders.

Data cleaning and analysis are integral steps in the data analysis process. Let's explore the theory behind each of these concepts:

Data Cleaning:

Data cleaning, also known as data preprocessing, is the process of identifying and correcting errors, inconsistencies, and inaccuracies in a dataset to ensure that it is suitable for analysis.

Data Quality:

The quality of data is crucial for accurate analysis. Data can contain various issues, including missing values, outliers, duplicate records, inconsistent formatting, and more. Data cleaning aims to rectify these issues.

Data Profiling:

Before cleaning the data, it's essential to understand its characteristics. Data profiling involves examining the dataset to identify the types of issues that need to be addressed. This step helps in creating a data cleaning plan.

Missing Data Handling:

One common issue in datasets is missing data. Missing data can be handled through methods like imputation (replacing missing values with estimates), removal of rows or columns with missing data, or special treatment based on domain knowledge.

Outlier Detection:

Outliers are data points that deviate significantly from the rest of the data. They can impact the analysis, and data cleaning involves identifying and addressing outliers through techniques like z-scores or IQR (Interquartile Range).

Data Standardization:

Data may have inconsistent formatting, such as date formats, units of measurement, or categorical values. Standardization involves transforming data into a consistent format for easier analysis.

Duplicate Record Removal:

Duplicates can distort analysis results. Identifying and removing duplicate records is essential to ensure the dataset's integrity.

Data Transformation:

Sometimes, data may need to be transformed to meet analysis requirements. This includes converting data types, encoding categorical variables, or scaling features.

Data Integration:

In cases where data comes from multiple sources, integration is required to consolidate the data into a single dataset. It may involve resolving inconsistencies between different sources.

Data Validation:

After cleaning, the data should be validated to ensure that all issues have been addressed. Data validation involves running checks to verify that the data is now clean and ready for analysis.

Data Analysis:

Data analysis is the process of exploring, examining, and drawing insights from the cleaned dataset.

7)Data Visualization:

Data visualization is a powerful technique for representing data graphically to help people understand and make sense of complex information. It involves the use of visual elements such as charts, graphs, and maps to convey data patterns, trends, and insights. Here's an explanation of the theory behind data visualization:

Visual Encoding:

Visual encoding is the process of mapping data attributes (e.g., values, categories) to visual properties (e.g., position, size, color, shape). For example, in a bar chart, the height of the bars represents data values, and the x-axis or y-axis position represents categories or time.

Data Types and Visualizations:

Nominal Data: Nominal data represents categories or labels with no inherent order, such as colors, names, or types. Nominal data can be visualized using bar charts, pie charts, or color-coded maps.

Ordinal Data:

Ordinal data has categories with a specific order or ranking, like customer satisfaction ratings (e.g., "poor," "average," "excellent"). Ordinal data can be represented using bar charts or dot plots.

Interval Data:

Interval data has a consistent measurement scale with equal intervals between values but no true zero point (e.g., temperature in Celsius). Line charts and histograms are often used for interval data.

Ratio Data:

Ratio data has a meaningful zero point (e.g., age, income, height). Scatter plots, histograms, and bar charts are common choices for visualizing ratio data.

Types of Visualizations:

Bar Charts:

Bar charts represent data using horizontal or vertical bars, with the length or height of each bar corresponding to the data value. They are suitable for comparing values across categories.

Line Charts:

Line charts connect data points with lines, making them ideal for showing trends and changes over time.

Pie Charts:

Pie charts display data as a circle divided into slices, where each slice represents a portion of the whole. They are useful for showing parts of a whole.

Scatter Plots:

Scatter plots display individual data points as dots on a two-dimensional plane. They are used to visualize the relationships between two variables.

Histograms:

Histograms group data into intervals (bins) and display the frequency or density of data points within each bin. They are used to visualize data distribution.

Box Plots:

Box plots show the distribution of data using quartiles, providing insights into the spread and skewness of the data.

8)Model Development & Evaluation:

Data Acquisition Module:

Implement code to interface with energy meters or sensors to collect real-time energy consumption data. Utilize libraries like pyserial or GPIO (for Raspberry Pi) to establish communication with hardware devices.

Data Processing Module:

Develop algorithms to process raw energy data, including parsing, cleaning, and converting data into usable formats (kWh, Joules, etc.). Apply data filtering techniques to remove noise and anomalies from the acquired data.

Data Storage Module:

Choose a suitable database system (e.g., SQLite, MySQL) to store processed energy consumption data. Implement code to establish a connection with the database and store data securely.

Visualization Module:

Use data visualization libraries such as Matplotlib or Plotly to create interactive charts and graphs representing energy consumption patterns. Display real-time data on a graphical user interface (GUI) for easy interpretation.

User Interface Module:

Design a user-friendly interface using GUI libraries like Tkinter or PyQt. Allow users to view historical data, set energy usage thresholds, and receive notifications when consumption exceeds specified limits.

Energy Analytics Module:

Develop algorithms to analyze energy usage patterns over time. Implement machine learning models for predicting future energy consumption based on historical data (optional).

EVALUATION OF THE PROJECT:

Accuracy:

Evaluate the accuracy of energy consumption measurements by comparing the data collected by your system with a known standard or reference data.

Performance:

Measure the performance of your system in terms of data processing speed, response time for user queries, and real-time data visualization capabilities.

User Experience:

Gather feedback from users regarding the usability and intuitiveness of the interface. Evaluate whether users find it easy to set thresholds and receive notifications.

Reliability:

Test the reliability of the notification system by simulating various scenarios, including network issues and unexpected system failures.

Scalability:

Evaluate how well the system handles a large volume of data and users. Assess whether the system performance degrades under heavy loads.

Security:

Ensure that the data storage and communication channels are secure. Evaluate the system against common security threats and vulnerabilities.

Documentation:

Evaluate the completeness and clarity of project documentation, including user manuals, code comments, and technical guides.

Algorithm Name: Linear Regression for Energy Consumption Prediction

Explanation:

Loading Data:

In this example, the `pjm_hourly_est.csv` file dataset is loaded. You can replace it with your specific dataset by changing the features and target variables according to your dataset's column names.

Data Preprocessing:

The dataset is split into features (X) and the target variable (y). Then, it's further split into training and testing sets (80% training, 20% testing) using `train_test_split`.

Model Training:

A Linear Regression model is initialized and trained using the training data (X_train and y_train).

Prediction:

The trained model predicts energy consumption based on the test features (X_test).

Model Evaluation:

Mean Squared Error (MSE) and R-squared value are calculated to evaluate the model's performance. MSE measures the average squared difference between predicted and actual values. R-squared indicates the proportion of the variance in the target variable that is predictable from the features.

9)Code Sample:

```
#training data

import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error, r2_score
df=pd.read_csv("pjm_hourly_est.csv")

X = df[['PJME_MW','PJM_MW']] # Features
y = df['PJME_MW'] # Target variable
X_train, X_test, y_train, y_test = train_test_split(X, y,
test_size=0.2, random_state=42)
model = LinearRegression()
model.fit(X_train, y_train)
```

```

y_pred = model.predict(X_test)
mse = mean_squared_error(y_test, y_pred)
r2 = r2_score(y_test, y_pred)

print(f"Mean Squared Error: {mse}")
print(f"R-squared: {r2}")

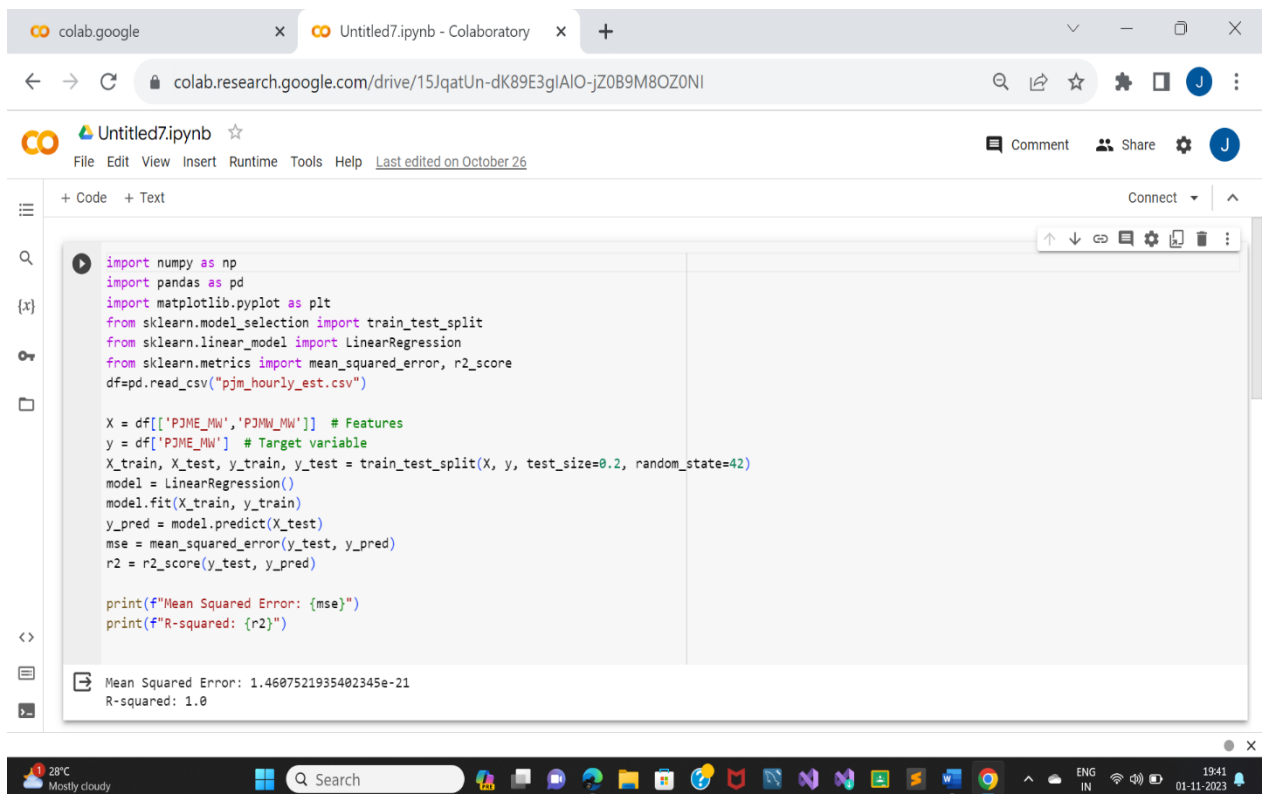
#testing data

import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error, r2_score
# Load your dataset
df = pd.read_csv('pjm_hourly_est.csv')
X = df[['PJME_MW', 'PJM_MW']] # Features
y = df['PJME_MW'] # Target variable
X_train, X_test, y_train, y_test = train_test_split(X, y,
test_size=0.2, random_state=42)
model = LinearRegression()
model.fit(X_train, y_train)
y_pred = model.predict(X_test)
mse = mean_squared_error(y_test, y_pred)
r2 = r2_score(y_test, y_pred)

print(f"Mean Squared Error: {mse}")
print(f"R-squared: {r2}")

```

10)Output Screenshot:



The screenshot shows a Google Colab notebook titled 'Untitled7.ipynb'. The code in the cell imports necessary libraries (numpy, pandas, matplotlib, sklearn) and loads a dataset 'pjm_hourly_est.csv'. It then splits the data into training and testing sets, fits a LinearRegression model, and prints the Mean Squared Error (MSE) and R-squared value. The output at the bottom of the cell shows 'Mean Squared Error: 1.4607521935402345e-21' and 'R-squared: 1.0'.

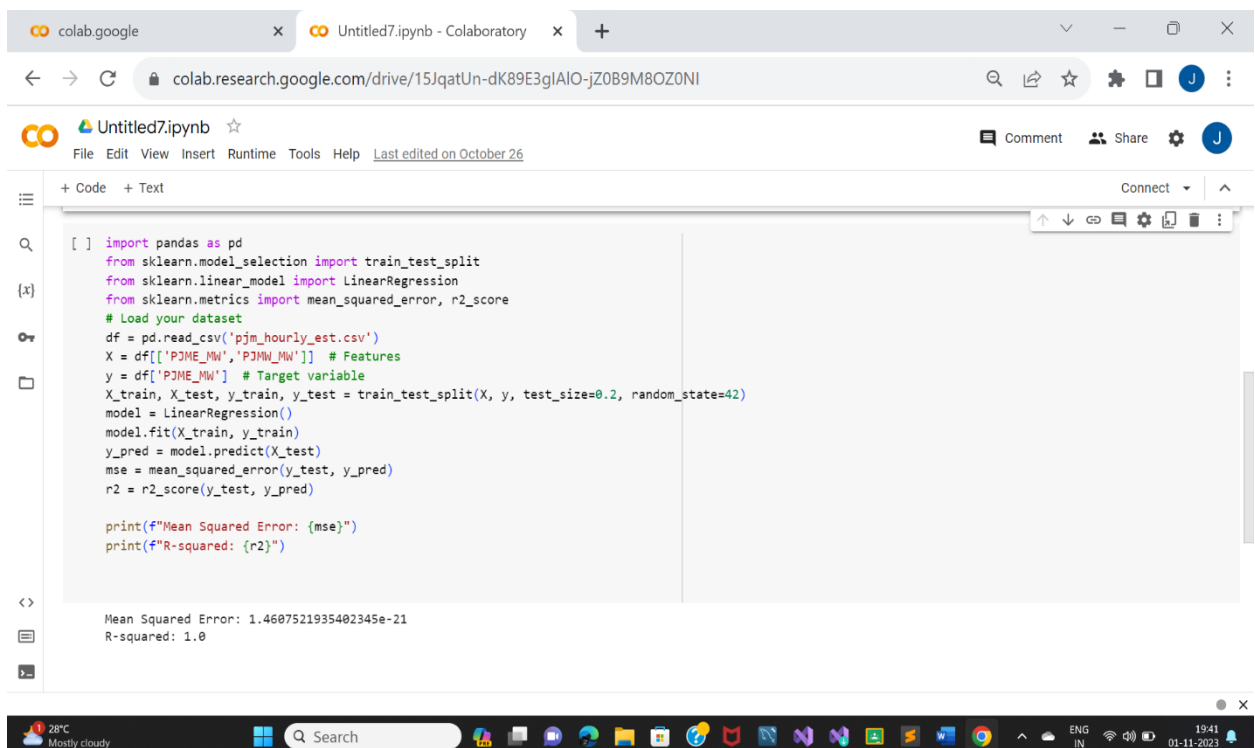
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error, r2_score
df=pd.read_csv("pjm_hourly_est.csv")

X = df[['PJME_MW', 'PJMW_MW']] # Features
y = df['PJME_MW'] # Target variable
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
model = LinearRegression()
model.fit(X_train, y_train)
y_pred = model.predict(X_test)
mse = mean_squared_error(y_test, y_pred)
r2 = r2_score(y_test, y_pred)

print(f"Mean Squared Error: {mse}")
print(f"R-squared: {r2}")
```

Mean Squared Error: 1.4607521935402345e-21
R-squared: 1.0

Fig-1



The screenshot shows a Google Colab notebook titled 'Untitled7.ipynb'. The code in the cell imports necessary libraries (pandas, sklearn) and loads a dataset 'pjm_hourly_est.csv'. It then splits the data into training and testing sets, fits a LinearRegression model, and prints the Mean Squared Error (MSE) and R-squared value. The output at the bottom of the cell shows 'Mean Squared Error: 1.4607521935402345e-21' and 'R-squared: 1.0'.

```
[ ] import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error, r2_score
# Load your dataset
df = pd.read_csv('pjm_hourly_est.csv')
X = df[['PJME_MW', 'PJMW_MW']] # Features
y = df['PJME_MW'] # Target variable
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
model = LinearRegression()
model.fit(X_train, y_train)
y_pred = model.predict(X_test)
mse = mean_squared_error(y_test, y_pred)
r2 = r2_score(y_test, y_pred)

print(f"Mean Squared Error: {mse}")
print(f"R-squared: {r2}")
```

Mean Squared Error: 1.4607521935402345e-21
R-squared: 1.0

Fig-2

11)Conclusion &Future Enhancement

References:

In conclusion, AI has forged a path toward a future where energy is not merely consumed but optimized, conserved, and synchronized with environmental goals.

The convergence of data, AI, and energy systems offers a beacon of hope for a more sustainable planet, where every watt is purposefully utilized, and where conservation and efficiency are not mere aspirations but daily realities. This survey demonstrates that AI is poised to play a pivotal role in this transformative journey toward responsible and informed energy consumption. As we embrace the possibilities that AI brings, we move closer to a world where energy is both smarter and greener.

