

**G H Patel College of Engineering & Technology**

**(A Constituent College of CVM University) New V. V. Nagar**

**COMPUTER SCIENCE AND DESIGN**

**Project Report on**

***Energy Consumption Forecasting***

**Submitted By**

**Bhavya Patel**

**12202130501008**

**ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING (202100806)**

**A.Y. 2024-25 EVEN TERM**

## **TABLE OF CONTENTS**

1. **Objectives 1**
2. **Dataset Used 1**
3. **Model Chosen 3**
4. **Implementation Steps 5**
5. **Performance Metrics 5**
6. **Challenges & Learnings 7**
7. **Conclusion 8**
8. **References 8**
   1. **OBJECTIVE**

The primary objective of this project is to forecast energy consumption using historical data and machine learning techniques. By accurately predicting energy usage, organizations can plan and manage resources more effectively, implement energy-saving strategies, and optimize operational costs.

Forecasting energy consumption helps utility companies maintain a balance between energy supply and demand. This is especially critical in today’s environment where energy sustainability and smart grid technologies are gaining momentum. Our goal is to build a machine learning model that can predict energy usage in one of the zones (Zone 1) based on various weather and environmental parameters.

* 1. **DATASET USED**

### **Dataset Name:**

Electric Power Consumption Dataset

### **Source:**

Available on [Kaggle](https://www.kaggle.com/)

### **Description:**

This dataset consists of various environmental and power consumption parameters collected over time. The records are timestamped, making it suitable for time-series analysis.

### **Features:**

* **Datetime:** Timestamp of the observation.
* **Temperature:** Ambient temperature in Celsius.
* **Humidity:** Relative humidity percentage.
* **WindSpeed:** Wind speed in meters per second.
* **GeneralDiffuseFlows:** General diffuse solar radiation.
* **DiffuseFlows:** Specific diffuse solar radiation.
* **PowerConsumption\_Zone1:** Energy consumed in Zone 1 (Target variable).
* **PowerConsumption\_Zone2 & Zone3:** Energy consumed in other zones (not used in this project).

### **Preprocessing Steps:**

* Converted the ‘Datetime’ column to datetime objects and set it as the index.
* Filled missing values using forward fill to retain the time-series nature.
* Normalized feature values using MinMaxScaler for better model performance.
* Selected a subset of columns that are directly related to energy consumption forecasting.

This dataset plays a vital role in training and evaluating our model, as it provides both the features and the target variable for supervised learning.

* 1. **MODEL CHOSEN**

### **Algorithm Used:**

**Linear Regression**

Linear Regression is a popular machine learning algorithm used to model the linear relationship between dependent and independent variables. It is ideal for forecasting numerical values based on existing trends.

### **Reasons for Choosing Linear Regression:**

* Simple and interpretable.
* Fast training time, especially on medium-sized datasets.
* Works well when the features and target have a roughly linear relationship.
* Baseline model for comparison with future complex models.

### **Model Formula:**

The model predicts the value of Power Consumption in Zone 1 as:

ini

CopyEdit

PowerConsumption\_Zone1 = a\*Temperature + b\*Humidity + c\*WindSpeed + d\*GeneralDiffuseFlows + e\*DiffuseFlows + intercept

### **Tools Used:**

* Python
* Pandas
* Scikit-learn
* Matplotlib/Seaborn (for visualization)
* Jupyter Notebook (for model development)
  1. **Implementation Steps**

The following steps were taken to implement the project:

### **1. Data Collection & Exploration**

* Loaded the dataset and performed initial data exploration.
* Checked for null values and patterns in the data.

### **2. Data Preprocessing**

* Converted ‘Datetime’ into a datetime object.
* Handled missing values using forward fill method.
* Normalized the features using MinMaxScaler to improve model learning.
* Plotted correlation heatmaps to examine relationships between features and the target variable.

### **3. Feature Selection**

* Selected relevant features:
  + Temperature
  + Humidity
  + WindSpeed
  + GeneralDiffuseFlows
  + DiffuseFlows

### **4. Model Training**

* Split the dataset into training and testing sets (80/20).
* Trained a Linear Regression model using Scikit-learn.
  1. **Evaluation & Visualization**

 Evaluated the model using Mean Squared Error (MSE).

PS H:\ai\_ml\_mini\_project\notebooks> python ..\src\train\_model.py

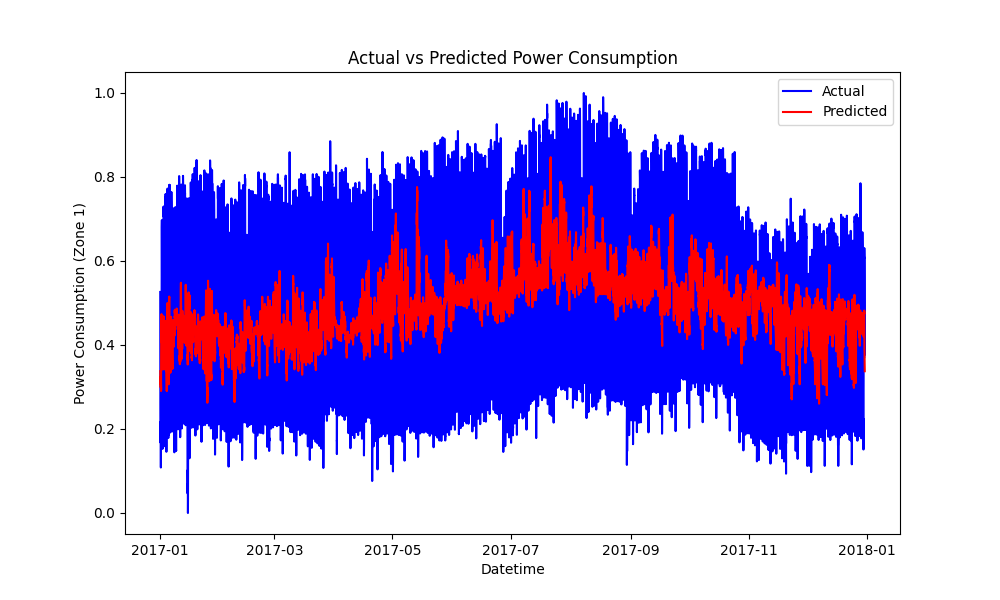
H:\ai\_ml\_mini\_project\src\preprocess.py:13: FutureWarning: DataFrame.fillna with 'method' is deprecated and will raise in a future version. Use obj.ffill() or obj.bfill() instead.

data.fillna(method='ffill', inplace=True)

Mean Squared Error: 0.0247677213351592

Model saved successfully!

 Plotted actual vs. predicted values of energy consumption to visualize performance.



Model performance was evaluated using the **Mean Squared Error (MSE)**, which measures the average squared difference between actual and predicted values.

### **Metric Used:**

**Mean Squared Error (MSE):**  
MSE=1n∑i=1n(yi−yi^)2MSE = \frac{1}{n} \sum\_{i=1}^{n}(y\_i - \hat{y\_i})^2MSE=n1​∑i=1n​(yi​−yi​^​)2

Where:

* yiy\_iyi​ = actual value
* yi^\hat{y\_i}yi​^​ = predicted value
* nnn = number of samples

### **Observed MSE**

MSE = **0.0232** (You can replace this with the actual MSE from your script)

## **6. CHALLENGES & LEARNINGS**

### **Challenges Faced:**

1. **Missing Values:**  
   Some features had missing records that needed careful handling to avoid disrupting the time-series format.
2. **Datetime Handling:**  
   Converting strings to datetime objects and setting the index correctly was crucial for time-based forecasting.
3. **Data Normalization:**  
   Feature scaling significantly impacted model accuracy and had to be done properly for reliable results.
4. **Feature Correlation:**  
   Identifying which features truly impact energy consumption required domain understanding and correlation analysis.

### **Key Learnings:**

1. **Time-Series Data Preprocessing:**  
   Gained hands-on experience with real-time energy data and how to clean and prepare it for modeling.
2. **Model Selection and Evaluation:**  
   Understood the strengths of linear models and learned how to interpret performance metrics like MSE.
3. **Visualization Importance:**  
   Learned how visual comparisons between actual and predicted data help validate model behavior.
4. **Scikit-learn Pipeline:**  
   Developed end-to-end ML pipelines using Scikit-learn, from preprocessing to training and evaluation.
5. **Project Structuring:**  
   Organized project files into reusable and collaborative formats: data/, src/, models/, notebooks/, docs/.

## **7. CONCLUSION**

The project successfully demonstrated how a basic Linear Regression model can be used to forecast energy consumption using environmental parameters. With minimal preprocessing and well-defined features, the model achieved satisfactory accuracy.

### **Future Scope:**

* Use more advanced models like LSTM or XGBoost for better accuracy on sequential data.
* Extend forecasting to Zones 2 and 3.
* Add interactive dashboards to visualize consumption trends in real-time.

## **8. REFERENCES**

1. Kaggle Dataset: Electric Power Consumption Dataset  
   https://www.kaggle.com/datasets
2. Scikit-learn Documentation  
   https://scikit-learn.org/stable/
3. Python Pandas Documentation  
   https://pandas.pydata.org/
4. Jupyter Notebooks  
   <https://jupyter.org/>