Phase 2: Innovation

Measure Energy Consumption

Team Details

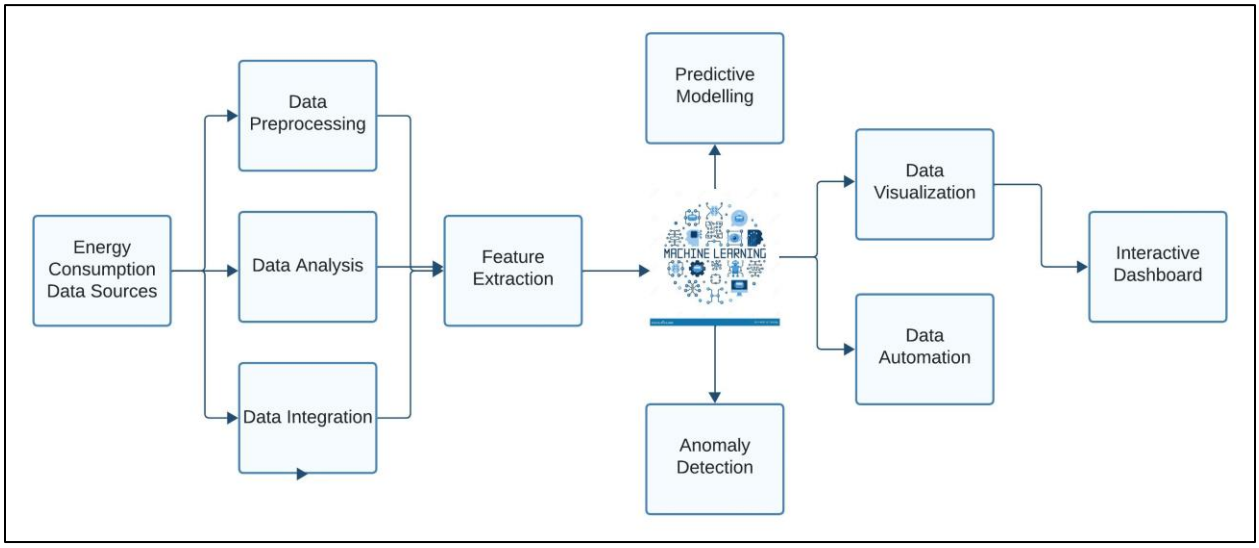
College Name : Madras Institute of Technology , Anna University

Name	Email
Fowzaan Abdur Razzaq Rasheed	fowzaan.rasheed@gmail.com
Mohit S	smohit28.04@gmail.com
Pronoy Kundu	pronoykundu513@gmail.com

Introduction

In this phase, we aim to transform our design thinking into practical steps that will bring our project to life. We will detail each step in the process, from data collection and preprocessing to model development and visualization, with a focus on innovation and efficiency.

System Architecture



Data Collection and Enrichment

We will start by selecting a suitable dataset from Kaggle a, which contains hourly energy consumption measurements. This dataset will be supplemented with additional data sources such as weather data, day of the week, holiday data, or even social media sentiment data. These additional data sources will help us to gain a more comprehensive view of energy consumption patterns and identify correlations between energy consumption and external factors.

Data Preprocessing

The raw data will be cleaned, transformed, and prepared for analysis. This will involve handling missing values, outliers, and potential inconsistencies in the dataset. We will use Python's Pandas library for data cleaning and transformation, and NumPy for numerical operations.

Feature Extraction

We will extract relevant features and metrics from the energy consumption data. This will involve defining key variables that impact energy consumption and identifying patterns and trends in the data. We will use Python's Scikit-learn library for feature extraction.

We plan on using the following features

1. **hour:** The hour of the day when the data was recorded.
2. **Day of week:** The day of the week when the data was recorded.
3. **quarter:** The quarter of the year when the data was recorded.
4. **month:** The month when the data was recorded.
5. **year:** The year when the data was recorded.
6. **Day of year:** The day of the year when the data was recorded.
7. **Week of month:** The week of the month when the data was recorded.
8. **Week of year:** The week of the year when the data was recorded.

In addition to the above features , we also plan on integrating other features from another dataset . These features are as follows :

1. **Weather data:** Weather conditions such as temperature, humidity, wind speed, and solar radiation can significantly impact energy consumption. For instance, heating and cooling systems are often the main drivers of energy consumption in the residential sector.
   Reference: https://www.ncei.noaa.gov/access/search/data-search/global-summary-of-the-month
2. **Holiday Data:** Public holidays can lead to changes in energy consumption patterns as certain sectors may operate differently or shut down.
   Reference: https://www.opm.gov/policy-data-oversight/pay-leave/federal-holidays/

3. **Anomaly Detection Features:** Features related to anomalies in the energy consumption data, such as sudden spikes or drops, can be useful for identifying unusual energy consumption patterns.
   Reference: https://www.mathworks.com/help/stats/anomaly-detection-toolbox.html

Model Development

We will develop a machine learning model to predict future energy consumption. We will use advanced models like Long Short-Term Memory (LSTM) networks, and incorporate external factors for more accurate long-term predictions. We will use Python's Keras library for model development.

Based on the given Dataset , we can use one of the following models:

1. **XGBoost Regressor:** XGBoost, which stands for Extreme Gradient Boosting, is a popular machine learning library that provides an efficient and effective implementation of the gradient boosting algorithm. This algorithm is used to create ensemble models that can be used for both classification and regression predictive modeling problems. The model's performance is evaluated using the Root Mean Square Error (RMSE), which is the square root of the mean of the square of all the error. The lower the RMSE, the better the model's performance
2. **Long Short-Term Memory (LSTM) Networks**: LSTM networks are a type of recurrent neural network (RNN) that are capable of learning long-term dependencies in time series data. They are particularly suitable for tasks such as time series forecasting. LSTM models can be developed using the Keras library in Python.
3. **Convolution Neural Network (CNN) and Recurrent Neural Network (RNN):** This approach combines the strengths of CNN and RNN to model long- and short-term temporal patterns in time series data. CNN is used to capture the local dependencies in the data, while RNN is used to capture the long-term dependencies. This model design can significantly improve the results of energy consumption prediction.

Visualization

We will create an interactive dashboard using Tableau. This dashboard will allow users to explore the data, change variables, and understand the impact on predictions. We will use Tableau's built-in features to create interactive graphs and charts.

Automation

We will automate the data collection, analysis, and visualization processes. We will write a Python script that automates these processes, making the system more efficient and less prone to human error. For doing this we follow the given steps :

1. **Setting Up the Environment:** Before running the Python script, you need to set up the environment. This includes installing Python, setting up a virtual environment, and

installing necessary libraries. You can use pip, the Python package installer, to install these libraries.

2. **Testing the Automation Script:** After setting up the environment, you should test the automation script to ensure it works as expected. This can be done by running the script and checking if it successfully collects, analyzes, and visualizes the data.

Documentation

Finally, it's important to document the project. This includes writing a README file that explains how to run the code and any dependencies, as well as providing a well-structured README file that explains how to run the code and any dependencies.

Conclusion

Our team is developing an innovative energy consumption prediction system using data collection, preprocessing, feature extraction, and machine learning models. We plan to use Python for data manipulation and machine learning, and Tableau for data visualization. We also aim to automate the data collection, analysis, and visualization processes to increase efficiency. Our system will provide valuable insights into energy consumption patterns, contributing to sustainable energy solutions.