**Project Methodology**

This project was executed using the following steps:

1. Data Preprocessing
2. Exploratory Data Analysis (EDA)
3. Model Training with different models like, LSTM, GRU, and BiLSTM
4. Model Evaluation (with error metrics like MSE, RMSE, MAE and MAPE) and Comparison results
5. Explainable AI with SHAP.
6. Feeder Allocation

**Initial Data shape imported:** 7728 rows and 13 columns: Datetime and all the 12 feeders comprise of the columns.

**Data Proprocessing**

The process carried out were both Data cleaning and Feature Engineering:

1. Data Cleaning: The following was done:

* Datetime conversion for the date column which was from Jan 1st, 2024 to Dec. 31st 2024.
* Column Renaming: renamed all the feeders columns to lowercase and converted the values from watt-hours (WH) to Megawatt-hours (MW).
* Handling missing values: linear interpolation technique was used to fill the 13 rows data missing from the data imported.

1. Feature Engineering: We extracted the following data as additional columns:

* time-based features such as **hour, day\_of\_week, month, day, and if weekend (as boolean)** was extracted from the datetime column
* Cyclical features such as **sine and cosine transformation of hour, month and day\_of\_week** was created to help ML model understand cyclical nature of time variables

This increased the shape of the datsaset from 7728 rows by 13 columns to 7728 by 25 columns.

**Exploratory Data Analysis (EDA)**

The following plots were generated using matplotlib and seaborn libraries to understand the data.

A graph of a line

AI-generated content may be incorrect.

Figure 1: Average Monthly Energy Consumption Pattern for all feeders

A graph of blue bars

AI-generated content may be incorrect.

Figure 2: Average Energy Consumption Pattern by each Feeders

A graph with a line

AI-generated content may be incorrect.Figure 3: Average Daily Energy Consumption Pattern

A graph showing a line

AI-generated content may be incorrect.

Figure 4: Average Hourly Energy Consumption Pattern

A graph of numbers and lines

AI-generated content may be incorrect.

Figure 5: Hourly Distribution of Energy Consumption

A screenshot of a graph

AI-generated content may be incorrect.

Figure 6: Correlation Heatmap by each feeders

A graph with blue lines

AI-generated content may be incorrect.

Figure 7: Total Energy Consumption Over time

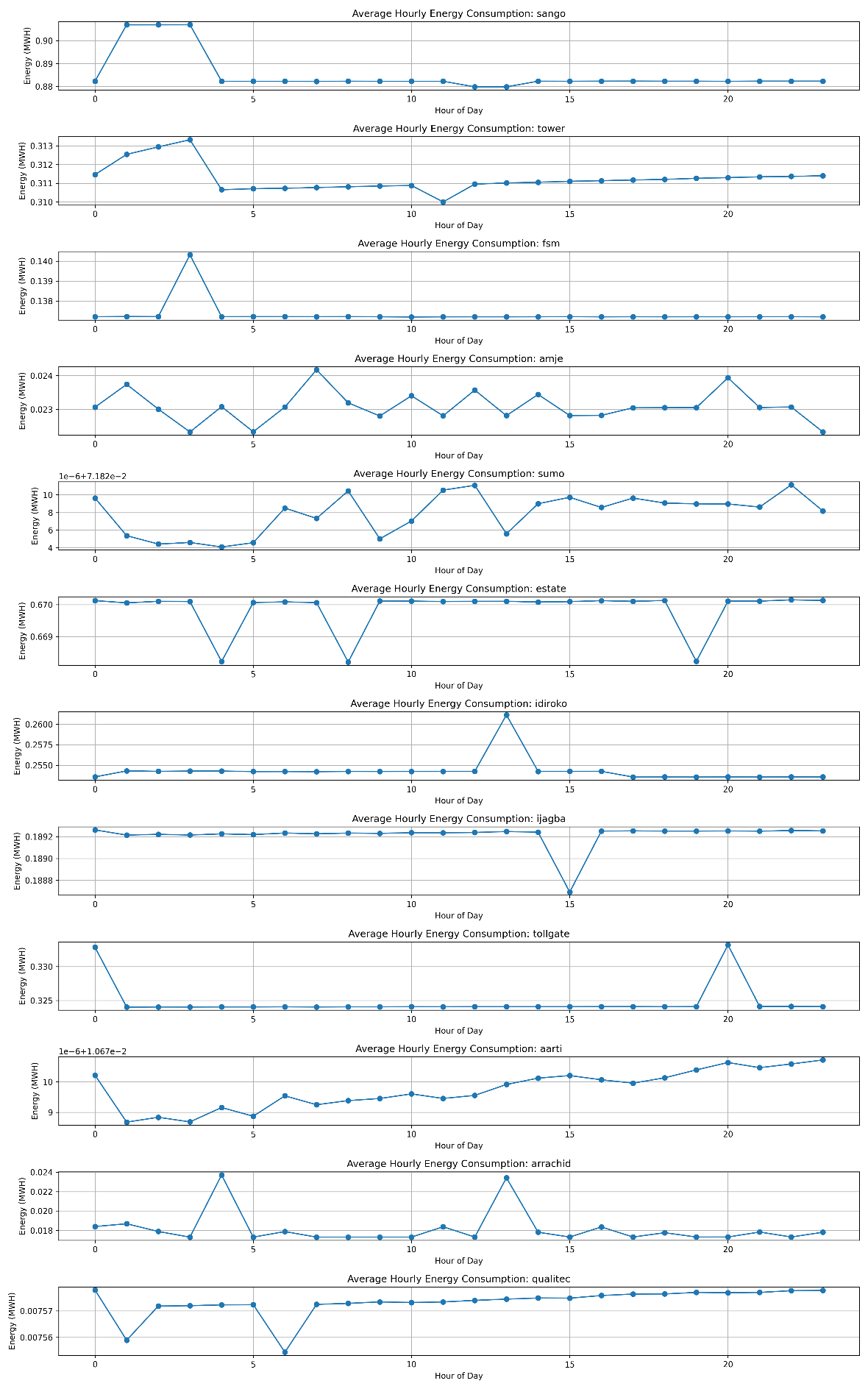


Figure 8: Average Hourly Energy Consumption of each Feeders

**N.B:** The Statistical Description of the Data can be viewed via this [link](https://docs.google.com/spreadsheets/d/1WgVAV048YdDtlkXb1wT2ZSVaYBjgv_byf2W2TD1NVUs/edit?usp=sharing), please go through it and copy to the report if needed.

**Model Training**

The Data was prepared for the LSTM and other RNN model to be trained upon, which followed the following process that aligns with the multivariate time-series training:

1. Removed the datetime column.
2. Feature Normalization: All numeric features was normalised to [0, 1] range using Min-max feature scaling as seen in the equation below:
3. Time-series sequencing: For temporal models like LSTM, GRU, and BiLSTM, the data was reshaped into sequences using a sliding window approach:

*FOR i = 0 TO N - time\_steps:*

*X[i] = X[i:i+time\_steps],*

*y[i] = y[i+time\_steps]*

Where N= Number of samples

Where, each sequence consisted of a defined number of previous time steps (time\_steps), and the corresponding target was the value at the next time step

N.B: Due to the hourly time-series focus, the time step used was 24, which denote it has use previous 24 hours of data to predict the next 24 hours.

1. Train-Test Splitting: Shuffle data randomly and allocate 70% for training, 10% validation and 20% for testing.
2. Train each model and then obtain their best hyperparameters using randomised search method.

**LSTM Model Training Result**

Model Architecture Summary

|  |  |  |
| --- | --- | --- |
| Layer (Type) | Output Shape | Parameters |
| LSTM (lstm\_4) | (None, 24, 64) | 19,712 |
| Dropout (dropout\_6) | (None, 24, 64) | 0 |
| LSTM (lstm\_5) | (None, 32) | 12,416 |
| Dropout (dropout\_7) | (None, 32) | 0 |
| Dense (dense\_6) | (None, 64) | 2,112 |
| Dense (dense\_7) | (None, 288) | 18,720 |
| Reshape (reshape\_3) | (None, 24, 12) | 0 |

**Total Parameters:** 52,960 (206.88 KB)

**Trainable Parameters:** 52,960 (206.88 KB)

**Non-trainable Parameters:** 0 (0.00 Byte)

Best LSTM Hyperparameters using Randomised Search method:

| **Parameter** | **Value** |
| --- | --- |
| units | 64 |
| activation | relu |
| dropout\_rate | 0.30 |
| learning\_rate | 0.00014031767388896486 |
| tuner/epochs | 10 |
| tuner/initial\_epoch | 4 |
| tuner/bracket | 1 |
| tuner/round | 1 |
| tuner/trial\_id | 0020 |

A graph of a loss

AI-generated content may be incorrect.

Figure 9: LSTM Model Training

A screenshot of a graph

AI-generated content may be incorrect.

Figure 10: Actual vs Predicted Energy Consumption of each feeders

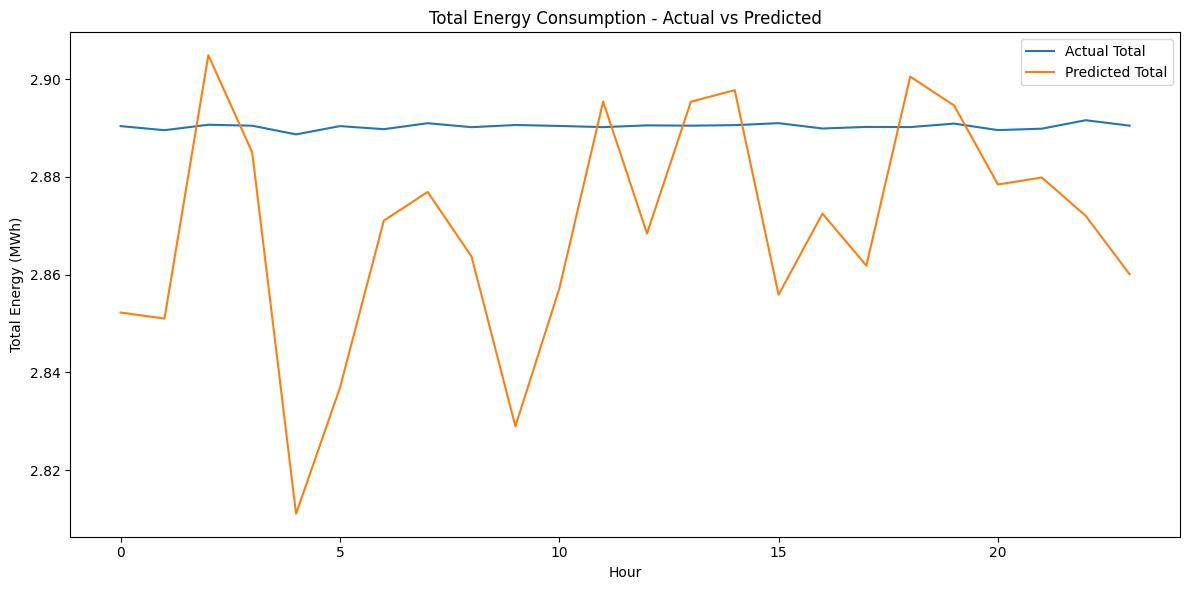


Figure 11: Actual vs Predicted Total Energy Consumption

GRU Model result