Time Series Analysis and Forecasting of Electricity Usage

Group Name: YYYY
Group Member: Yueer Liu, Zimeng(Sandy) Li, Ziying Song, Yihan Yang

Contents

Part 1: Introduction

This project focuses on analyzing electricity consumption patterns using the Electricity Load Diagrams (2011-2014) dataset from the UCI Machine Learning Repository. The dataset contains electricity consumption data for Portuguese **370 clients** over a period of four years, recorded at **15-minute intervals**. This dataset provides a solid foundation for identifying consumption trends and developing robust forecasting models.

1.1 Background & Business Value

Accurate prediction of electricity usage is highly beneficial for customers, power providers, and policy holders. Due to factors such as seasonality, time of day, economic activity, and consumer behavior, electricity demand is constantly changing, making it difficult to optimize energy distribution effectively. A significant portion of Portugal's electricity grid is powered by wind, solar, and hydroelectric energy. Inaccurate predictions can lead to energy waste, grid disruptions, and increased operational costs, as energy suppliers may overproduce electricity (resulting in inefficiencies and financial losses) or underproduce, potentially causing blackouts. This could hinder Portugal's energy conservation and renewable energy integration policies. Furthermore, the increasing integration of renewable energy into the grid introduces additional unpredictability, necessitating more sophisticated forecasting methods to ensure stable energy supply.

From a business perspective, improved electricity demand forecasting holds significant value for multiple stakeholders. Utility companies can enhance grid efficiency by optimizing power generation, load shedding, and load balancing, reducing unnecessary energy production costs while ensuring reliable power supply. For commercial and industrial consumers, better forecasting enables smarter energy management, allowing businesses to lower electricity costs by shifting usage to off-peak periods. Individual consumers can also reduce unnecessary electricity consumption through better consumption forecasts. For policymakers, accurate electricity demand forecasting aids in long-term infrastructure planning, ensuring that investments in power plants, transmission lines, and renewable energy projects align with anticipated demand. Additionally, by leveraging machine learning-driven forecasts, utility companies can implement dynamic pricing models, offering consumers more competitive pricing structures based on real-time demand trends. Overall, the ability to predict electricity consumption more accurately can lead to cost savings, improved operational efficiency, and enhanced customer satisfaction across the energy sector.

1.2 Objective

One of our primary goals is to forecast the **daily electricity consumption for each client**, providing more personalized energy usage insights for residential and commercial consumers. By utilizing exploratory data analysis (EDA) to study electricity consumption behavior, we aim to understand daily patterns, peak usage times, and anomalies through forecasting. Understanding how different customers consume electricity enables precise time-series forecasting, improving their electricity management. Accurate forecasts help consumers optimize energy usage, reduce costs, and make informed energy efficiency decisions. With advanced machine learning models such as SARIMAX and LSTM, and multi-time series model HistGradientBoostingRegressor, this project aims to enhance forecasting accuracy, thereby assisting households and businesses in better energy planning.

From a broader perspective, we seek to predict macro-level electricity consumption behavior by analyzing usage differences among customers and segmenting them based on consumption patterns. By grouping customers with similar electricity usage behaviors, we can identify trends that help optimize energy distribution and management. Model evaluation will play a crucial role in determining the most effective forecasting method by comparing various models such as SARIMAX, LSTM, and HistGradientBoostingRegressor. Insights gained from this analysis can help utility providers, government agencies, and policymakers optimize grid operations, reduce peak load pressures, and improve overall energy efficiency. By integrating advanced time-series techniques and machine learning models, this project aims to enhance electricity demand forecasting, ensuring a more stable, cost-effective, and sustainable energy ecosystem.

1.3 Process and Change Management

The process of implementing advanced electricity forecasting models involves multiple stages, from data collection and preprocessing to model development, deployment, and continuous optimization. The proposed models will employ time-series forecasting techniques such as

SARIMAX, LSTM, and HistGradientBoostingRegressor to predict daily electricity demand. By incorporating customer segmentation, statistical features, and external variables (e.g., weather conditions), the models will provide actionable insights for energy suppliers and consumers. Once validated, the models will be integrated into utility providers' energy management systems, enabling real-time demand forecasting and improved load balancing. For commercial and residential consumers, forecasts will be accessible through dashboards and automated alerts to aid energy optimization.

The models need regular checks using performance metrics (like MAPE) and updates with new data to stay accurate. Adaptive learning will help the models adjust to changing consumption habits. Moving to AI-based forecasting will require a clear plan to help stakeholders, such as utility companies and policymakers, adapt smoothly. Since the models use advanced techniques like SARIMAX, LSTM, and HistGradientBoostingRegressor proper change management is needed to make the most of these predictions.

For utility companies, using AI to predict electricity demand will help automate how they manage power supply. In the past, utilities used manual methods or basic models to estimate demand, which often lead to inefficiencies. Now, with real-time AI predictions, they can better balance power generation and distribution, reducing waste and cutting costs. For example, a power company that used to produce too much electricity during busy times can now adjust supply based on AI forecasts. To make this work, grid operators will need training to understand the forecasts, use them in daily planning, and respond to sudden changes in demand.

For policymakers and regulators, AI-driven forecasts will provide valuable data for long-term energy planning and sustainability efforts. Governments can use these predictions to plan investments in power plants, transmission lines, and renewable energy projects. For instance, cities planning to build more electric vehicle (EV) charging stations can use AI forecasts to figure out when and where demand will be highest. To make this transition smooth, policymakers will need to work together, share data, and ensure AI forecasts support national and regional energy goals.

We aim to create accurate electricity consumption forecasts using time series models. By applying feature engineering, exploratory data analysis, and fine-tuning model parameters, we improve prediction accuracy. Our goal is to build a scalable forecasting system that helps manage energy better, keeps the grid stable, and supports smarter decision-making. In the long run, this will cut costs, improve efficiency, and make power distribution more reliable, for various stakeholders.

Part 2: Data Processing

2.1 Data Overview

This dataset is sourced from the UCI Machine Learning Repository and it records the electricity usage of over 370 customers from 2011 to 2014. The data is recorded at 15-minute intervals, which allows us to observe the temporal variations in energy usage closely. Each column in the dataset represents a different customer, while each row corresponds to a specific timestamp. Although there are no missing values in the data, some customers only started reporting their data after 2011, so their earlier records contain only zero values. Additionally, all timestamps are in local Portuguese time and include daylight saving time (DST) adjustments, which need to be taken into account during data analysis.

In addition, we include external data sources such as the Lisbon Weather Data to enhance our forecasting models. This dataset contains daily weather information from Lisbon, the capital of Portugal, including variables such as temperature, snowfall, sun hours, UV index, and humidity. For our analysis, we focus on the minimum and maximum daily temperatures, which we average to obtain the mean temperature for each day. Since electricity consumption is strongly affected by weather conditions—such as increased heating during winter or cooling in summer—this external weather data helps us capture seasonal patterns and temperature-driven fluctuations in energy demand.

2.2 Variable Analysis in the Datasets

The following is an analysis of the main variables in the datasets (Electricity Load Dataset and lisbon), as well as their importance in power consumption prediction.

2.2.1 Electricity Load Dataset: Timestamp (Index)

- Type: Datetime
- Description: This column records the date and time of each measurement point at 15-minute intervals.
- Importance: Timestamps are crucial for constructing time series data, as they help aggregate electricity consumption in desired units (in our cases, daily total consumption for each client). Additionally, they are used to adjust for Daylight Saving Time (DST) to ensure the temporal consistency of the data.

2.2.2 Electricity Load Dataset: Customer Electricity Consumption Columns (MT_001 to MT_370)

- Type: Numerical (kW)
- Description: Each column represents the electricity consumption of a specific customer at a particular time, with the unit being kilowatts (kW).
- Importance: These are the main target variables for prediction. Analyzing individual customers' electricity usage trends, outliers, and daily variations can help improve the accuracy of the prediction model. These variables need to undergo data transformation, such as converting kW to kWh, and handling potential outliers.

2.2.3 Lisbon: date_time

- Type: Datetime
- Description: This column records the date for each daily weather observation in Lisbon. Each entry corresponds to one calendar day.
- Importance: The date_time field allows us to align and merge the Lisbon weather data with the Electricity Load dataset based on date. This ensures that each day's weather conditions can be used as external features when modeling daily electricity usage.

2.2.4 Lisbon: Maximum and Minimum Temperature in Celsius

- Type: Numerical (Celsius)
- Description: maxtempC represents the highest recorded temperature in Celsius for a given day in Lisbon, and mintempC represents the lowest recorded temperature for the same day. These values reflect the daily temperature range.

- Importance: We use the maximum and minimum temperatures to calculate the mean daily temperature, which is added as a feature to our forecasting model. Since electricity usage is often influenced by weather (e.g., increased cooling during hot days), including temperature helps the model capture weather-driven changes in electricity demand.

2.2.5 Daily Electricity Consumption per Client (Target Variable)

- Type: Numerical (individual daily electricity usage per customer)
- Description: Represents the actual daily power consumption of each customer over time. This is the primary variable being forecasted in our modeling process.
- Importance: This variable is the target of our predictive model, which aims to capture personalized usage behavior and temporal dynamics for each customer. Predicting this value enables granular demand analysis, anomaly detection, and potentially client-level energy optimization strategies.

2.2.6 Rolling statistical feature (derived variable)

- Type: Numerical (Aggregated values based on time windows)
- Description: These are derived features based on rolling windows (7-day and 14-day) applied to each client's historical electricity usage. The statistics include mean, standard deviation, median, minimum, maximum, ratio of min to max, and coefficient of variation.
- Importance: These features help smooth out short-term noise and capture recent trends or variability in usage. This improves model stability and accuracy by allowing it to learn from both individual client behavior and broader time-based patterns. They are computed automatically using the RollingFeatures module during model training.

2.3 Pre-processing

2.3.1 Standardized numerical format

In the original dataset, the decimal point is represented by a comma (,) instead of a period (.), which may cause issues when processed in Python. Therefore, we will replace these commas with periods to ensure that the values are correctly recognized as floats

2.3.2 Unit Conversion (kW → kWh)

The power consumption unit in the data is kilowatts (kW), but each reading corresponds to the consumption over a 15-minute period. To convert it to the standard kilowatt-hours (kWh), we divide all the values by 4 to ensure consistent units, facilitating data comparison and trend analysis.

2.3.3 Delete the data from 2011.

As many customers had not yet started reporting data in 2011, a large number of records for that year were zero values. To reduce unnecessary noise, we have decided to start the analysis from 2012 to ensure the reliability of the data set.

2.3.4 Handling the Impact of Daylight Saving Time (DST)

In March, every year, Portugal moves its clocks forward by one hour, resulting in a data gap from 1:00 to 2:00 in the morning. To address this issue, we use interpolation, taking the average of the values at 00:45 and 02:00 to fill in the missing part, ensuring a smooth transition of energy consumption data.

In October, every year, Portugal moves its clocks back by one hour, causing the data from 1:00 to 2:00 in the morning to be recorded twice. If this data is used directly, it will artificially increase the energy consumption for this period. To ensure the accuracy of the data, we take the average of the energy consumption during the repeated time period (i.e., divide by 2) to prevent deviation.

Part 3: Pre-Modeling

3.1 Data Aggregation: Grouped by day

To analyze the electricity usage trends at different time scales, we resampled the data:

Daily aggregation: Sum up the readings taken at 15-minute intervals to calculate the total electricity consumption for each day, which help us observe the various patterns among different clients.

3.2 Train - Test Split

We split the data chronologically, using electricity consumption data from 2013-2013 for training, and the last 365 days (1 year of 2014) as a testing data set. This setup preserves the natural time order and ensures that at least two full years of seasonal patterns are captured in the training set, which is important for reliable time series forecasting.

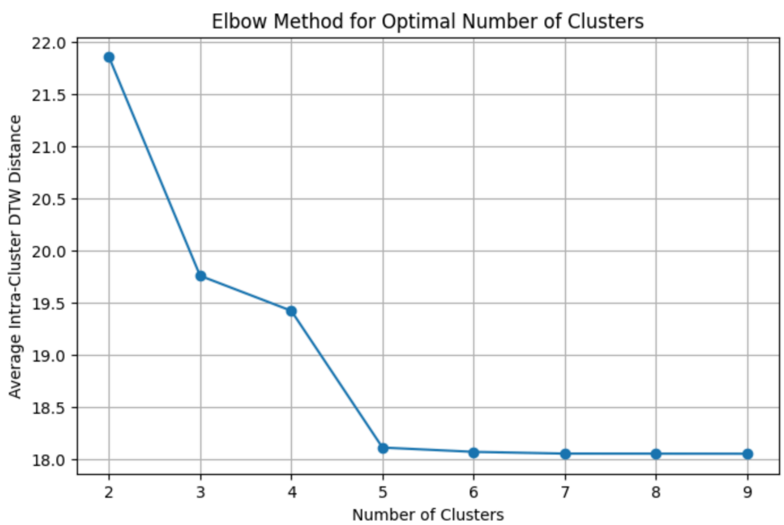
3.3 Clustering Through Dynamic Time Warping

Dynamic Time Warping (DTW) is a method to measure the similarity between two time series, even if they are not perfectly aligned in time. It works by "warping" the time axis to match similar shapes, even when peaks and valleys happen at different times. This makes it more flexible than traditional distance measures like Euclidean distance, which require point-to-point matching.

In the previous report, we use K-means clustering based on Euclidean distance and assume all data points are aligned, which doesn’t work well for time series like electricity usage. Different clients may have similar usage patterns that are just shifted in time. DTW can capture those similarities, while K-means would miss them. That’s why we now shift to DTW, which is more suitable for clustering time series data.

Our goal is to predict the daily electricity consumption for each client. Clustering clients with similar historical usage patterns helps us understand different types of users. For example, some clients use electricity steadily, while others have sharp peaks or seasonal trends.By using DTW-based clustering, we can group clients with similar behavior. This can improve our forecasting models because we can handle similar clients together, add cluster labels as features for prediction, and even understand which types of clients are more predictable. So DTW clustering is not the final goal but helps make our prediction models more accurate and more explainable.

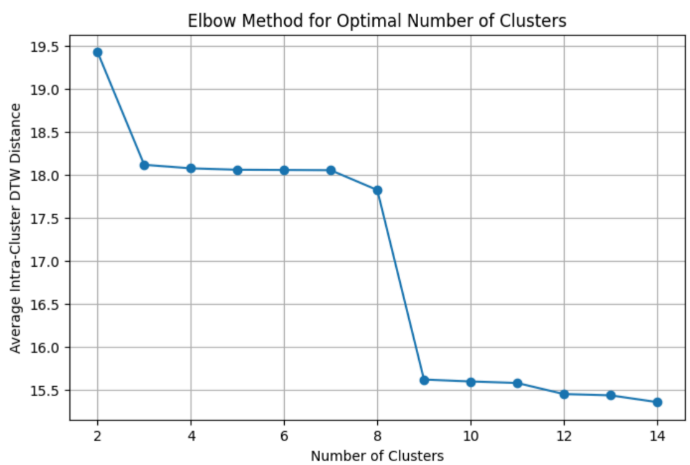
To decide how many clusters to use, we applied the Elbow Method. We tested different numbers of clusters from 2 to 9. For each value, we calculated the average DTW distance between clients within the same cluster. The idea is that smaller distances mean the clients in a cluster are more similar. We then plotted the number of clusters against the average distance. At k = 5, we found that adding more clusters didn't improve the distance much, and such a point is called the 'elbow'. According to the graph, we choose to make 5 clusters.



After running the clustering with 5 groups, we got the following result:

```
[370 rows x 2 columns]
Cluster
0    280
2     53
4     28
1      7
3      2
```

Most clients were assigned to Cluster 0, while Clusters 1 and 3 had very few clients, which appeared to be outliers. To address this imbalance, we removed the clients in Clusters 1 and 3 and re-ran the clustering process on the remaining clients. Our goal was to achieve a more balanced distribution of clients across clusters. However, after applying the elbow method, the new optimal number of clusters appeared to be either 3 or 9. The resulting cluster distribution was still uneven and did not show a clear improvement over the original clustering. Because of this, we decided to keep the original clustering result with 5 clusters.



Cluster	
0	26
1	201
2	50
3	69
4	3
5	1
6	1
7	1
8	9

We then proceed to analyze the pattern for each cluster. We discover that

Cluster 0: Steady Usage with Minor Variability (280 clients)

Clients in this cluster use electricity in a stable and consistent way. Their usage doesn't change much over time — there are no big spikes or drops. The small ups and downs are likely due to small changes in operations. This kind of pattern is common in places like factories, hospitals, or data centers that run all day, every day. These clients usually need a constant level of electricity to support regular operations.

Cluster 1: Highly Fluctuating Usage (7 clients)

The clients in Cluster 1 have highly unpredictable electricity use. Their usage goes up and down a lot, with no clear pattern or trend. This kind of behavior is often seen in businesses that don't follow a fixed schedule — like certain factories or commercial operations that change production levels or hours often. Their electricity use changes depending on things like demand, supply, or work schedules.

Cluster 2: Gradually Increasing Trend (53 clients)

Clients in this group show a slow and steady increase in electricity use over time. There are no big swings, but the overall trend goes upward. This could be due to business expanding or using more energy as time goes on. These users are growing, but in a gradual and controlled way.

Cluster 3: Initial Spike Followed by a Long Decline (2 clients)

This cluster includes clients who had high electricity use at the beginning but then dropped off sharply and stayed low. Such a pattern is often associated with businesses that initially operated at full capacity but then experienced a shutdown or a significant scale-down in activity. Another possibility is that these are seasonal businesses that only operate during certain times of the year. After their peak period, they stay mostly inactive with low energy use. So in this cluster, business ceased or significantly reduced operations after an initial high-energy phase.

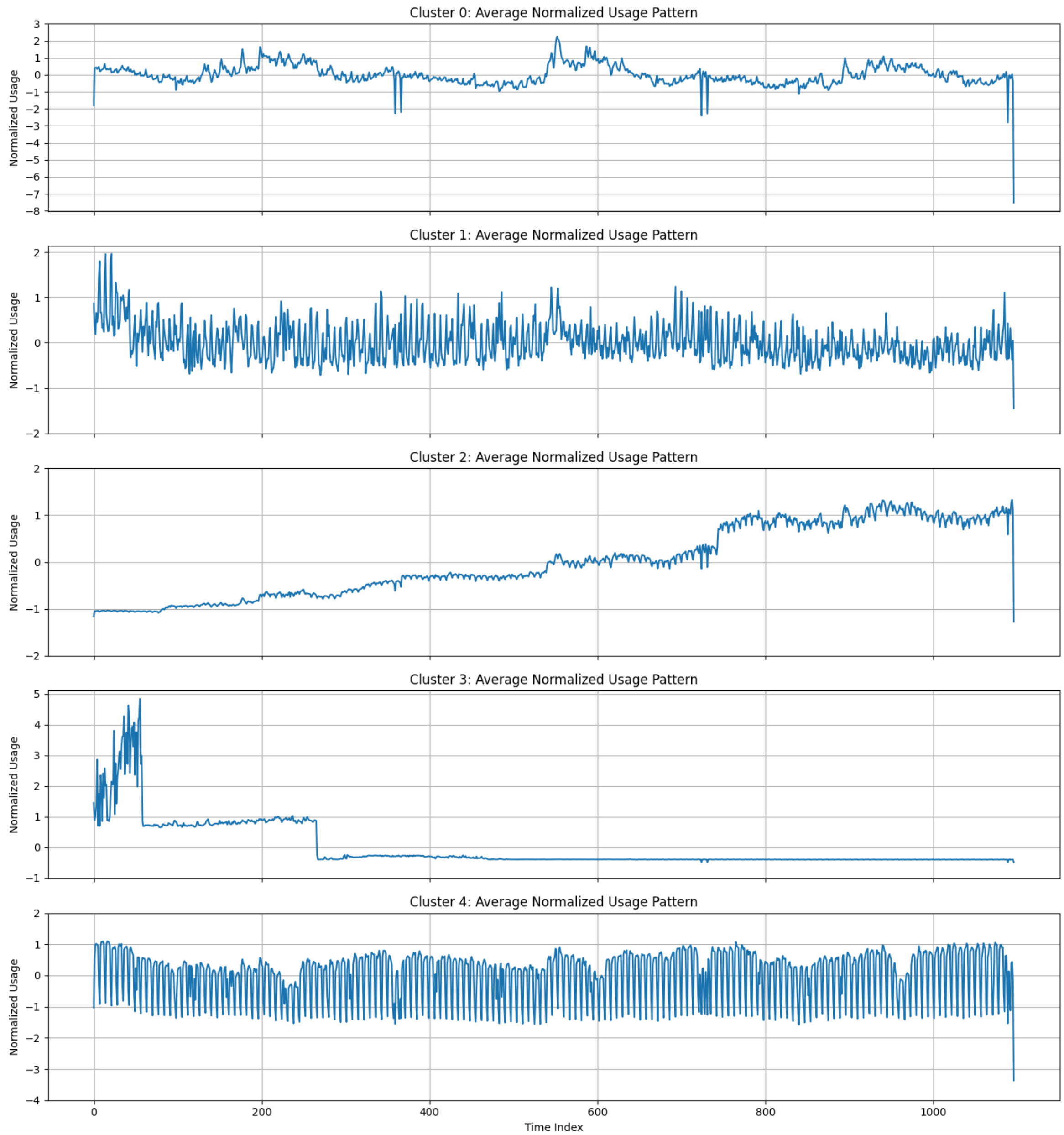
Cluster 4: Periodic Cyclical Pattern (28 clients)

Clients in Cluster 4 have a clear pattern that repeats over time. Their electricity use goes up and down in a regular way — like higher during weekdays and lower on weekends or nights. This suggests they follow a fixed schedule. These are likely schools, offices, or other businesses that operate on a weekly cycle. Their energy use is very predictable and consistent over time.

Final Interpretation

These clusters show that different clients have different electricity usage patterns. Some use electricity in a steady and reliable way (like Cluster 0), while others are growing (Cluster 2), unstable (Cluster 1), or shutting down (Cluster 3). Cluster 4 follows a clear and regular schedule.

Understanding these patterns helps us build better forecasting models. It also helps utility companies or energy planners make better decisions — like finding which clients are more predictable, which ones may need support, or where energy-saving plans might work best. Later on, we can also look at things like weather or business type to understand why clients behave the way they do.



In the modeling process, we use the cluster labels as one of the features in our forecasting models. Incorporating clustering helps us build more accurate predictions by capturing differences in electricity usage patterns across client groups, and allows us to better understand how demand varies among different types of users.

3.4 Dictionary:

In this project, we aim to build a multi-time series forecasting model that predicts the daily electricity consumption for each client. Instead of training a separate model for each of the 370 clients, we use a single model that learns across all clients, while still allowing each client to have their own time series. To organize and manage this structure, we use Python dictionaries, where each client is treated as an individual time series, but processed together under a shared model.

3.4.1 Client Time Series Dictionary (**series_dict**)

The first dictionary we created, called series_dict, is used to store the historical electricity consumption data for each client. Each key in the dictionary is a client ID (e.g., "MT_001"). Each value is a pandas Series with a datetime index representing that client's daily electricity usage. We also cleaned each time series by removing leading zero values at the beginning, which usually indicate periods before the client was active. This ensures the model only learns from valid consumption periods.

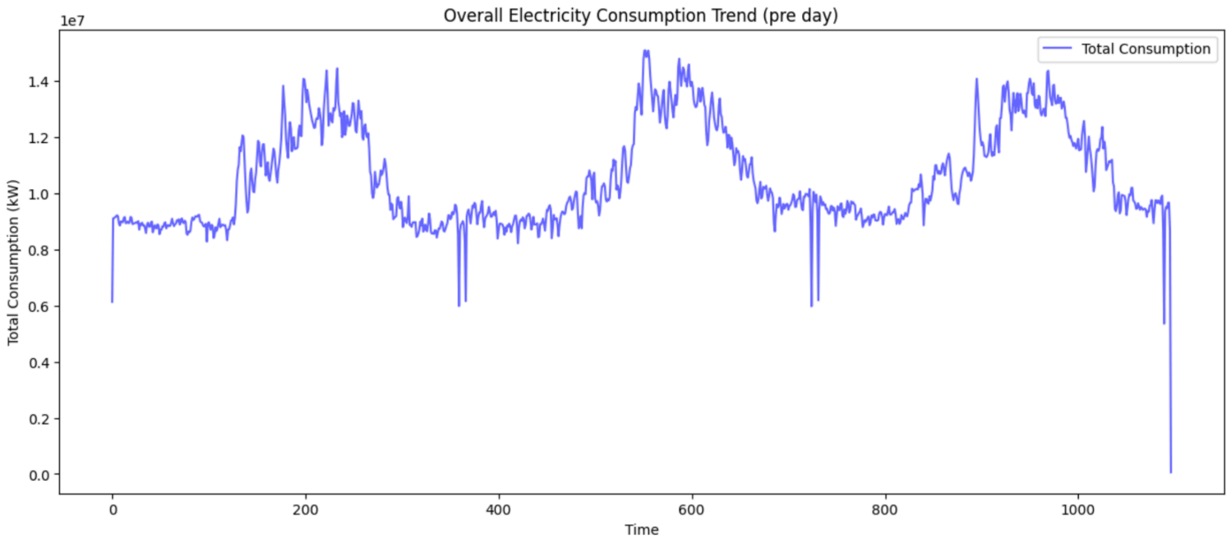
3.4.2 Exogenous Features Dictionary (**exog_dict**)

To help our forecasting model capture seasonal and group-based patterns, we created a second dictionary called exog_dict, which stores exogenous (external) features for each client.

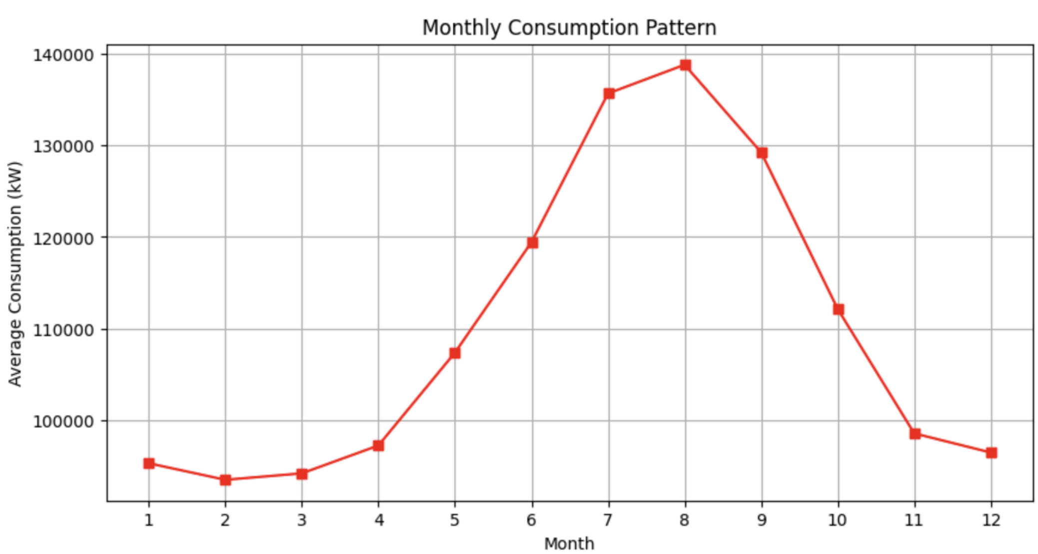
Each entry in this dictionary corresponds to one client and contains a DataFrame with time-aligned features. These features include:

- Cyclical time features: sin_dow, and cos_dow, which represent the sine and cosine transformations of the day of the week and help us capture weekly cycles. sin_doy, cos_doy, which are sine and cosine transformations of the day of the year to capture yearly seasonality.
- Average daily temperature (avg_temp): the average of the daily minimum and maximum temperatures in Lisbon. Electricity consumption is often influenced by weather — especially for heating and cooling — so including temperature helps the model learn these relationships.
- Cluster label (cluster): Based on DTW-based time series clustering, each client was assigned a cluster representing its typical consumption pattern. This cluster label is added as a categorical feature to help the model differentiate between types of clients (e.g., stable users vs. highly variable users).

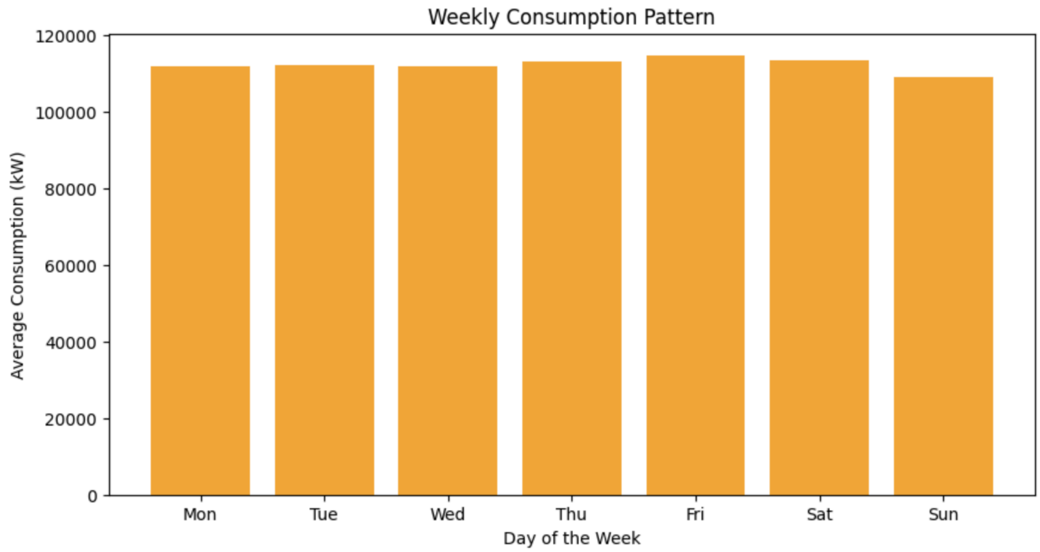
3.5 Exploratory Data Analysis



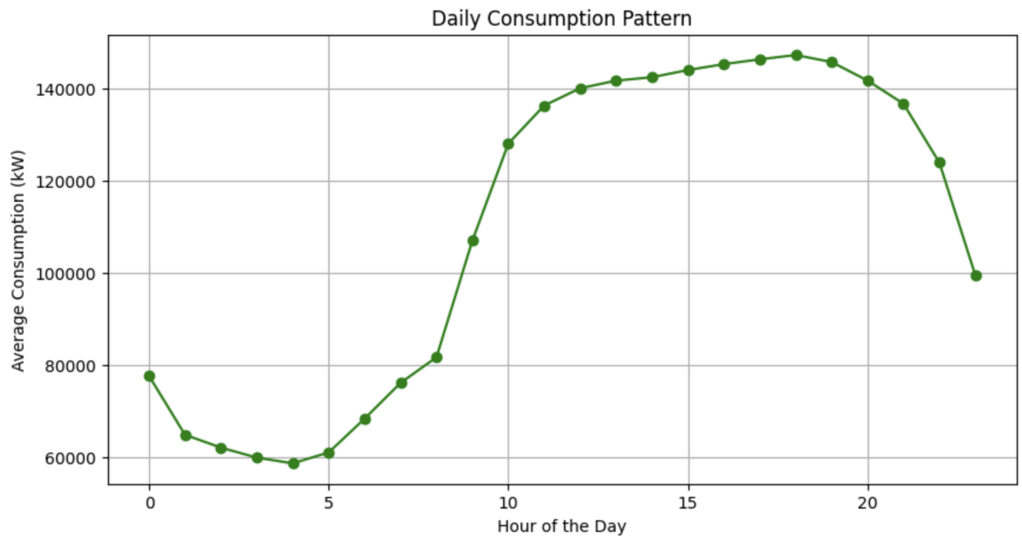
The daily trend of electricity consumption shows regular fluctuations, with obvious peaks and troughs that may be related to seasonal changes, such as increased heating in winter or air conditioning usage in summer. Sudden drops may be caused by holidays or unexpected factors. Overall, the periodicity of electricity demand is quite strong, which is crucial for power supply management and forecasting. Additionally, certain prominent changes merit further examination to ensure data quality or identify abnormal events. Mastering these patterns helps optimize energy allocation and enhance the accuracy of prediction models.



As can be seen from the graph, the monthly electricity consumption shows a distinct seasonal variation. The electricity consumption is relatively low during winter (from January to March) and late autumn (from November to December), while it gradually increases as spring arrives and reaches its peak in July and August. This might be related to the increased demand for air conditioning cooling. Subsequently, from September onwards, the electricity consumption gradually decreases and enters the trough period of the autumn and winter seasons. This trend reflects the direct impact of seasonal temperature changes on electricity demand and holds significant reference value for power dispatch and supply planning.



As can be seen from the graph, the electricity consumption is relatively stable within a week, but the electricity usage on weekends (especially on Sunday) shows a slight decrease. The electricity consumption from Monday to Friday is close to the peak level, which might be attributed to the high electricity demand from industries, businesses and office buildings during working days. While the electricity usage on Saturday and Sunday is slightly lower, possibly due to the reduced operation of some factories and companies and the different electricity usage patterns of residents. This trend indicates that the electricity demand is higher during working days and slightly decreases on weekends, which is of great reference value for power grid dispatching and energy management.



As can be seen from the graph, the electricity consumption shows a distinct daily variation pattern. The electricity consumption is the lowest during the early morning period (0 - 5 a.m.), which might be because most households and enterprises are in a rest state. After 5 a.m., the electricity consumption begins to rise and increases rapidly around 10 a.m., indicating an increase in electricity usage for industries, businesses, and residents. During the daytime (10 - 20 a.m.), the electricity consumption remains at a high level, suggesting that this period is the peak electricity usage time, possibly related to office, production, and air conditioning usage. After 8 p.m., the electricity consumption begins to decline and returns to a lower level in the early hours of the night. This trend reflects the influence of daily and working patterns on electricity demand and helps optimize grid dispatch and load management.

Part 4: Modeling

4.1 SARIMAX

SARIMAX is a time series forecasting method that extends ARIMA by incorporating both seasonal patterns and external variables. This makes it well-suited for modeling electricity demand, which often follows weekly cycles and is influenced by external factors like weather.

In our approach, we applied SARIMAX at the cluster level: for each client cluster obtained through DTW-based clustering, we selected up to five representative clients and concatenated their daily electricity consumption data to train a single SARIMAX model. To account for external influences, we used **average daily temperature as an exogenous variabl**e. Including temperature helps the model better capture fluctuations in energy demand due to heating or cooling needs.

Each SARIMAX model was configured with an order of (1,1,1) and a seasonal order of (1,1,1,7) to reflect weekly seasonality. We trained the model on historical data and tested its forecasting accuracy over the most recent year (365 days) of observations. This approach allowed us to explore how seasonal and weather-related patterns affect consumption across different types of clients.

4.2 LSTM - Long Short-Term Memory

We selected an **LSTM-based Recurrent Neural Network** using the RecurrentNetwork module from the pytorch_forecasting library to model electricity consumption patterns. LSTM networks are selected for this task because they are designed to capture long-term dependencies in sequential data. This is especially important for electricity usage forecasting, where user behavior is often influenced by historical trends, weekly cycles, and gradual shifts. Compared to traditional models, LSTMs can better model nonlinear temporal dynamics and maintain memory across time steps.

In our implementation, we used the TimeSeriesDataSet class to preprocess the data. Each sample consisted of 30 days of historical usage (encoder length) to predict the next 24 hours (prediction length). We trained the model globally across all users, using group_id to distinguish individual time series. Our model configuration included a hidden size of 16, one LSTM layer, a learning rate of 0.03, a dropout of 0.1 to prevent overfitting, and SMAPE (Symmetric Mean Absolute Percentage Error) as the loss function. We trained the model using the Trainer class for 10 epochs, with gradient clipping to stabilize training. This setup allowed the model to learn general patterns across users while also adapting to individual consumption behaviors, making it a strong fit for our multi-series forecasting task.

4.3 Histogram-based Gradient Boosting Model

4.3.1 Introduction to Multi-series Recursive Forecaster

For this forecasting task, we also selected **HistGradientBoostingRegressor** as the prediction model. Our dataset contains over 1,000 days of electricity usage for each of the 370 users, along with additional features such as rolling statistics (e.g., moving averages), average daily temperature, calendar-based variables (like day of week and day of year), and cluster labels that group users by behavioral similarity.

We chose this model for several reasons. First, it handles **complex nonlinear relationships** between features without requiring us to manually define those interactions. It also performs well on **high-dimensional datasets**, which is important given the number of engineered features we included. Additionally, it naturally supports **missing values** and **categorical variables**, reducing the need for extensive preprocessing.

Unlike time series models that require the data to be stationary, HistGradientBoostingRegressor makes **fewer assumptions about data distribution**, which is especially useful for modeling real-world electricity usage that often fluctuates and follows irregular patterns. Its histogram-based algorithm is also **efficient and scalable**, making it well-suited for large datasets like ours. Overall, this model strikes a good balance between flexibility, speed, and accuracy, and pairs well with our use of the ForecasterRecursiveMultiSeries framework to deliver reliable multi-step forecasts across many users.

4.3.2 Model Training and Prediction

We first aggregated the electricity usage data into daily consumption per user to train our model. Then, we applied a **sliding window approach** to generate training samples. Specifically, we used the past 30 days of each user’s consumption and features to predict the electricity usage for the next 7 days. This rolling window technique ensures that the model captures recent patterns and adapts to temporal dynamics in the data.

We trained a **global model** using the ForecasterRecursiveMultiSeries wrapper, which allowed the HistGradientBoostingRegressor to learn shared consumption patterns across all 370 users. This method is particularly effective in leveraging similarities among users, which improves generalization and performance—especially for users with limited data.

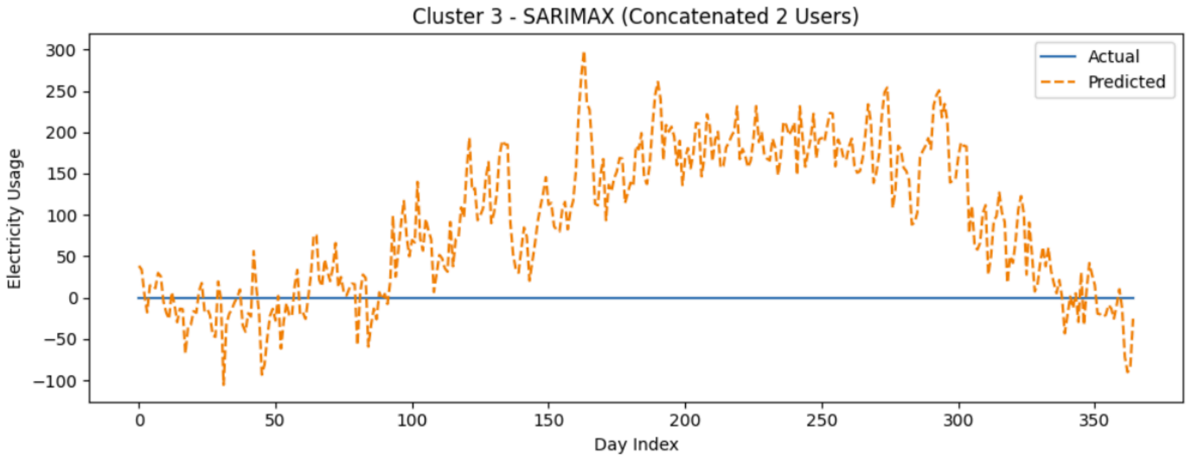
To perform **multi-step forecasting**, we adopted a recursive prediction strategy: the model predicts one day, then uses that prediction as input to forecast the next day, and continues this process iteratively for the entire horizon (up to 7 days). This simulates real-world usage, where future values are not known and must be estimated sequentially.

Part 5: Result

5.1 SARIMAX Model Evaluation

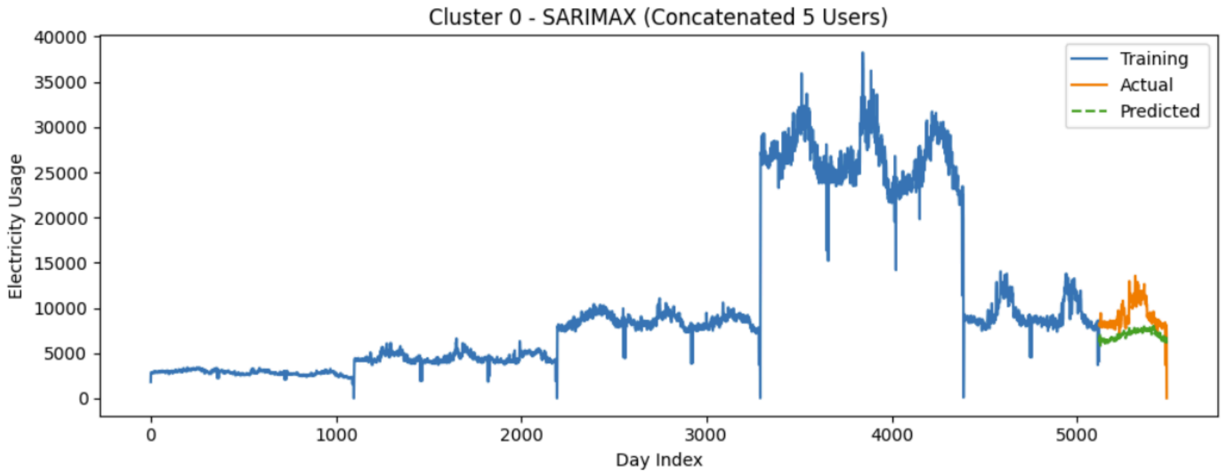
After training SARIMAX models on each cluster, we decided to remove Cluster 3 from evaluation. This is because Cluster 3 only had 2 clients and showed very strange usage patterns, including flat or near-zero values with no useful variation. This made the SARIMAX model unable to learn meaningful trends, and the predictions became unstable. Therefore, we treated it as an outlier cluster and excluded it from the final comparison.

```
Training SARIMAX for Cluster 3 with 2 users...
```

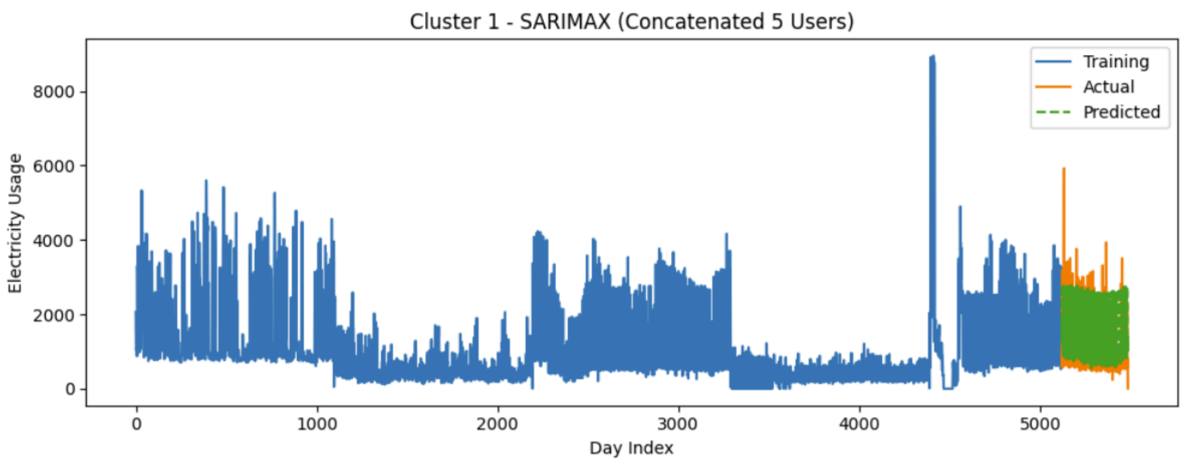


For the remaining clusters (0, 1, 2, and 4), we evaluated SARIMAX performance using RMSE, MAE, and MAPE. The MAPE scores were generally high, with Cluster 0 at 72.40%, Cluster 1 at 75.18%, Cluster 2 at a very high 426.93%, and Cluster 4 at 67.70%. The average MAPE across all clusters was 160.55%, which is too high for practical forecasting. This shows that SARIMAX struggled to provide accurate predictions, especially for clusters with large usage variation or unstable patterns.

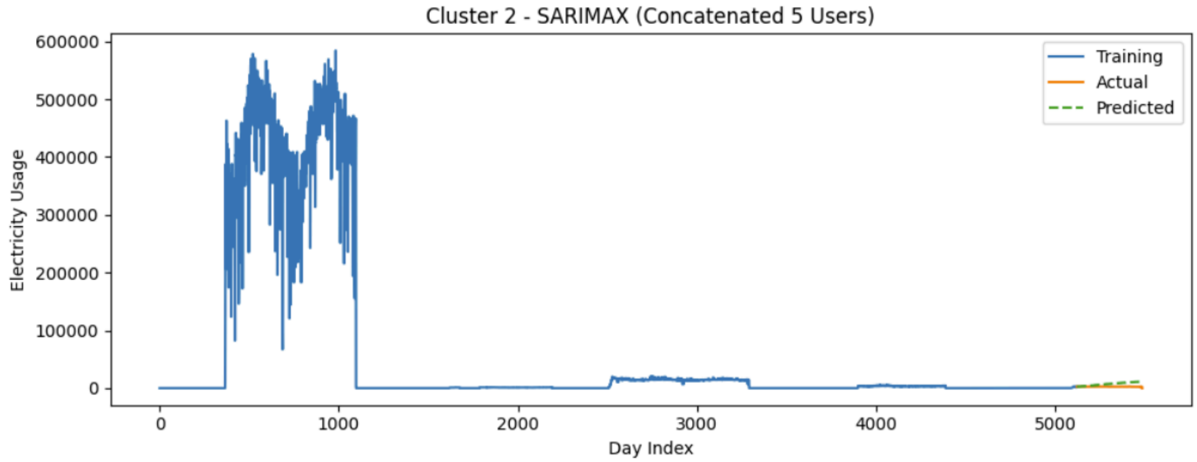
```
Cluster-Level Evaluation
Cluster 0 (users: 5): RMSE = 2468.21, MAE = 2207.69, MAPE = 72.40%
Cluster 1 (users: 5): RMSE = 627.67, MAE = 398.87, MAPE = 75.18%
Cluster 2 (users: 5): RMSE = 5383.02, MAE = 4645.42, MAPE = 426.93%
Cluster 4 (users: 5): RMSE = 9422.94, MAE = 7043.89, MAPE = 67.70%

Average MAPE across clusters: 160.55%
```



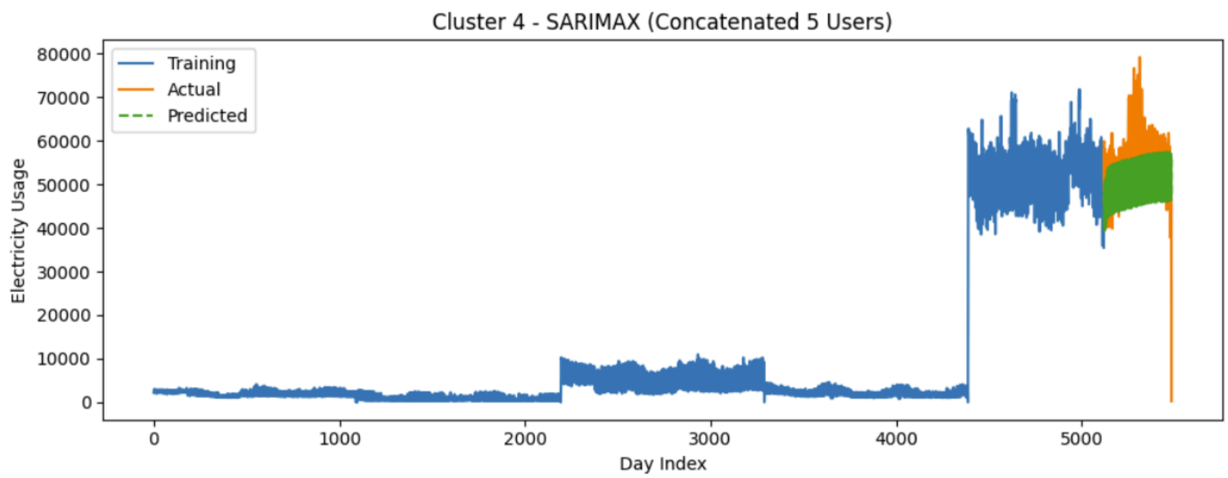

```
Training SARIMAX for Cluster 1 with 5 users:
```



Training SARIMAX for Cluster 2 with 5 users:



Training SARIMAX for Cluster 4 with 5 users:

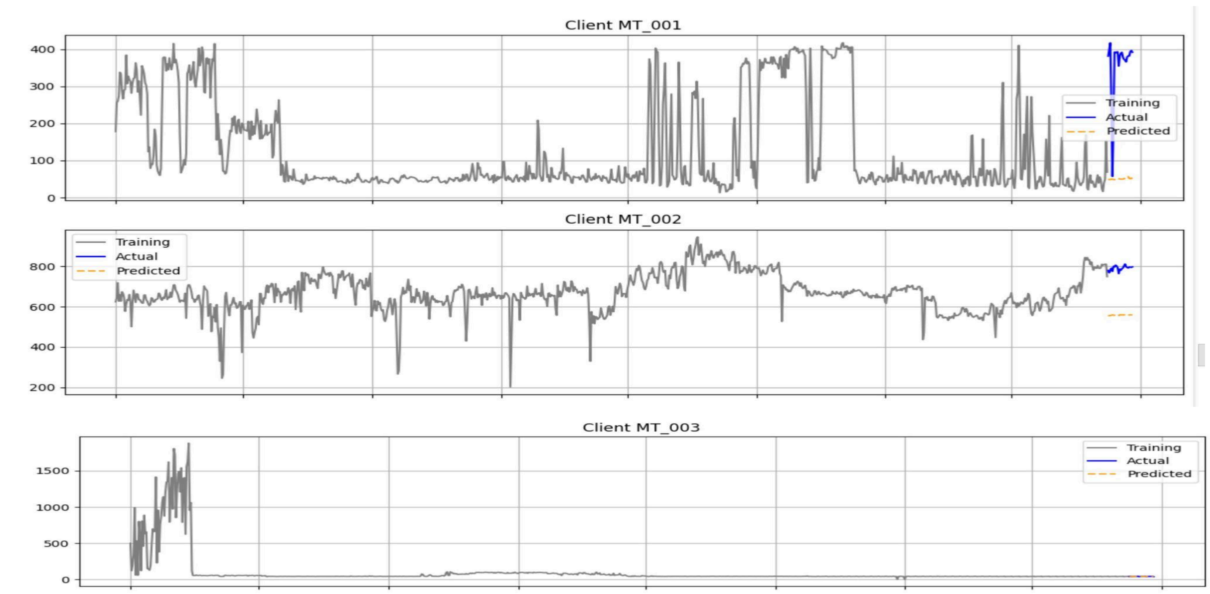


Because of these results, we did not choose SARIMAX as our final model. First, SARIMAX is for a single time series and can only be used at the cluster level, not for individual clients. It assumes all time series in the training data follow the same pattern, which doesn't work well when each client has unique behavior. Second, since we had to concatenate the time series of several clients for training, we could only select a few users per cluster. If we included too many, the combined series became too large and slow to run. These limitations make SARIMAX less scalable and less flexible for our goal of forecasting electricity usage at the individual client level. Instead, we used a multi-series machine learning model (e.g. HistGradientBoostingRegressor) that could handle all clients at once, with more efficient and accurate results.

5.2 LSTM Model Evaluation

We applied our trained LSTM model to forecast the next 24 days of electricity consumption across 369 users. The plots below illustrate actual vs. predicted values for a few representative clients, where we also include the historical training portion for context. For many users, especially those with consistent patterns, the model can capture the overall trend. However, for clients with highly irregular or flat historical data, the forecasts appear less accurate and often underperform. This is reflected in the overall MAPE score of **135.94%**, which is notably high.

The elevated MAPE suggests that the LSTM struggled to generalize across all user patterns, likely due to data sparsity, noise, or abrupt changes in user behavior that are difficult to capture from a limited 30-day context window. In some cases, like MT_003, the actual values are close to zero, making percentage-based errors extremely sensitive and inflating the MAPE—even when absolute error is relatively small. Despite these challenges, the model does learn general consumption dynamics for many users, which indicates room for further tuning, such as longer history windows, better feature engineering, or hybrid models.

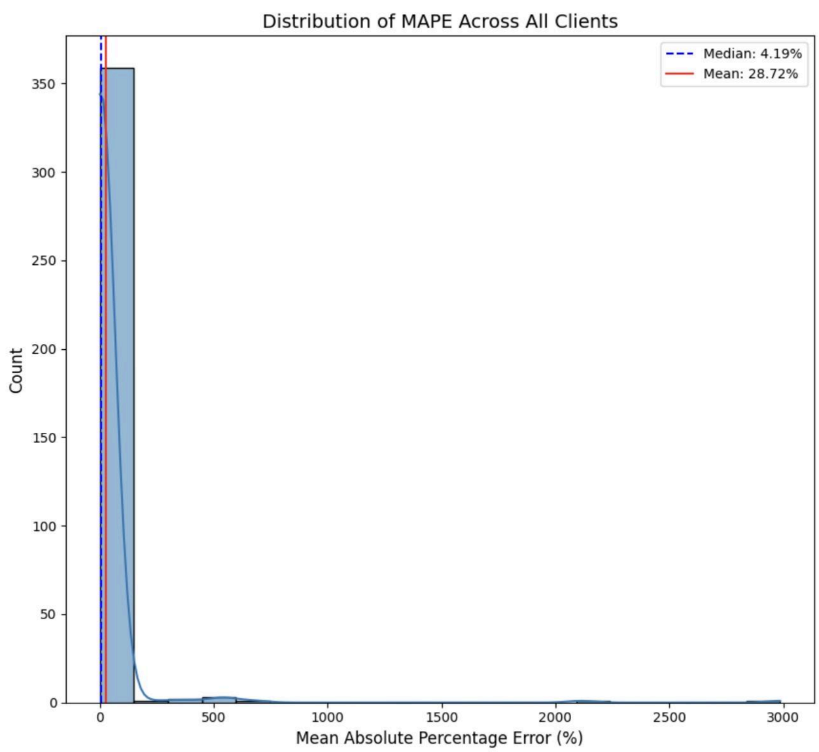


To further improve forecasting performance, we also explored the use of **HistGradientBoostingRegressor**. While LSTM is effective for capturing temporal dependencies, tree-based models like HistGradientBoostingRegressor are often more robust to noise and irregular patterns, especially in datasets with high variability across users. This model proved to be significantly more accurate and stable in our experiments, delivering lower error rates and better generalization. In the following section, we detail the setup and results of this approach.

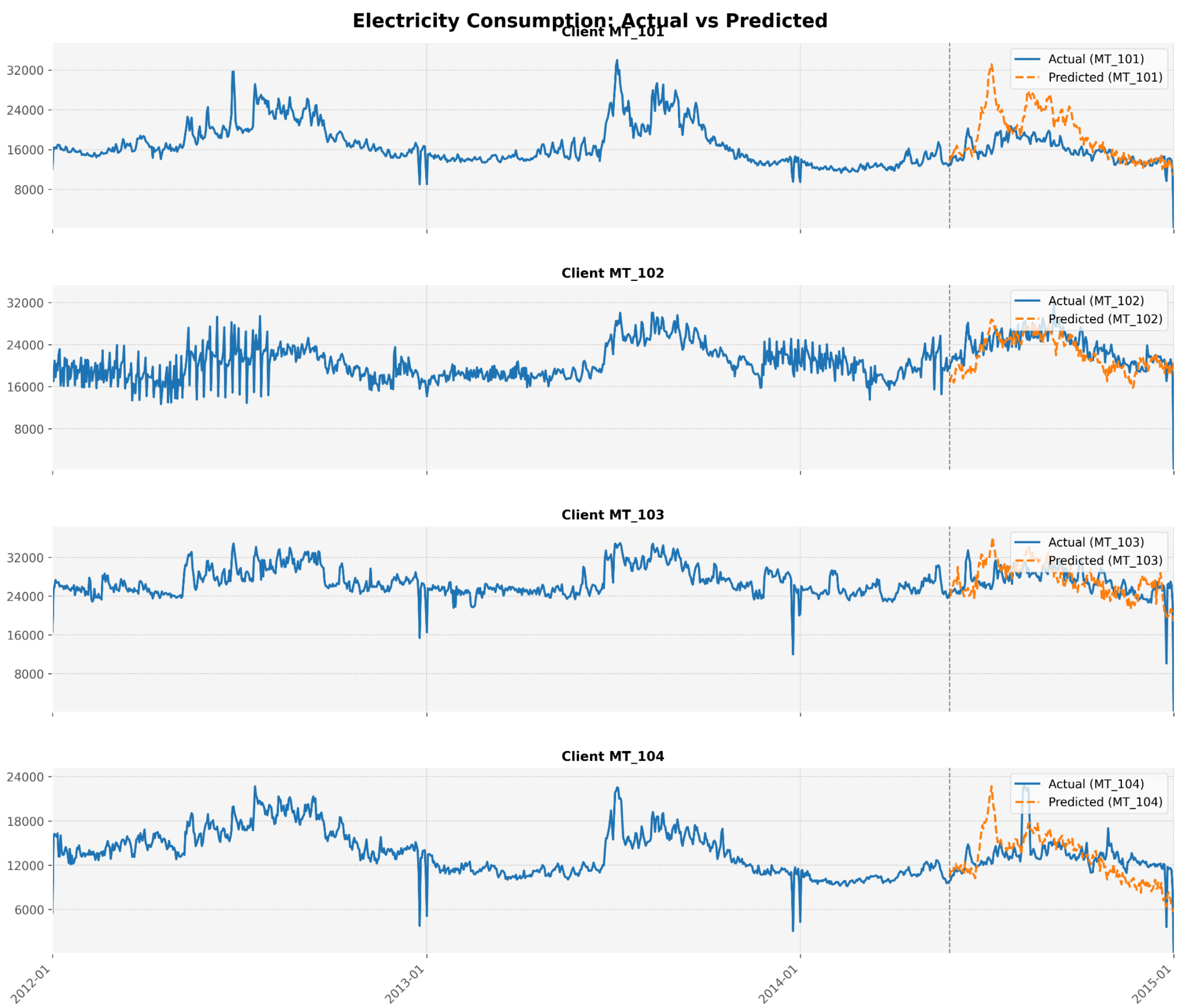
5.3 Histogram-based Gradient Boosting Model Evaluation

We evaluated the model using standard metrics, including **MAE**, **RMSE**, and **MAPE**, across all users and forecast horizons. Overall, the model achieved the following results:

- **MAE (mean absolute error):** 655.66
- **RMSE (root mean square error):** 779.34
- **MAPE (mean absolute percentage error):** 30.20%

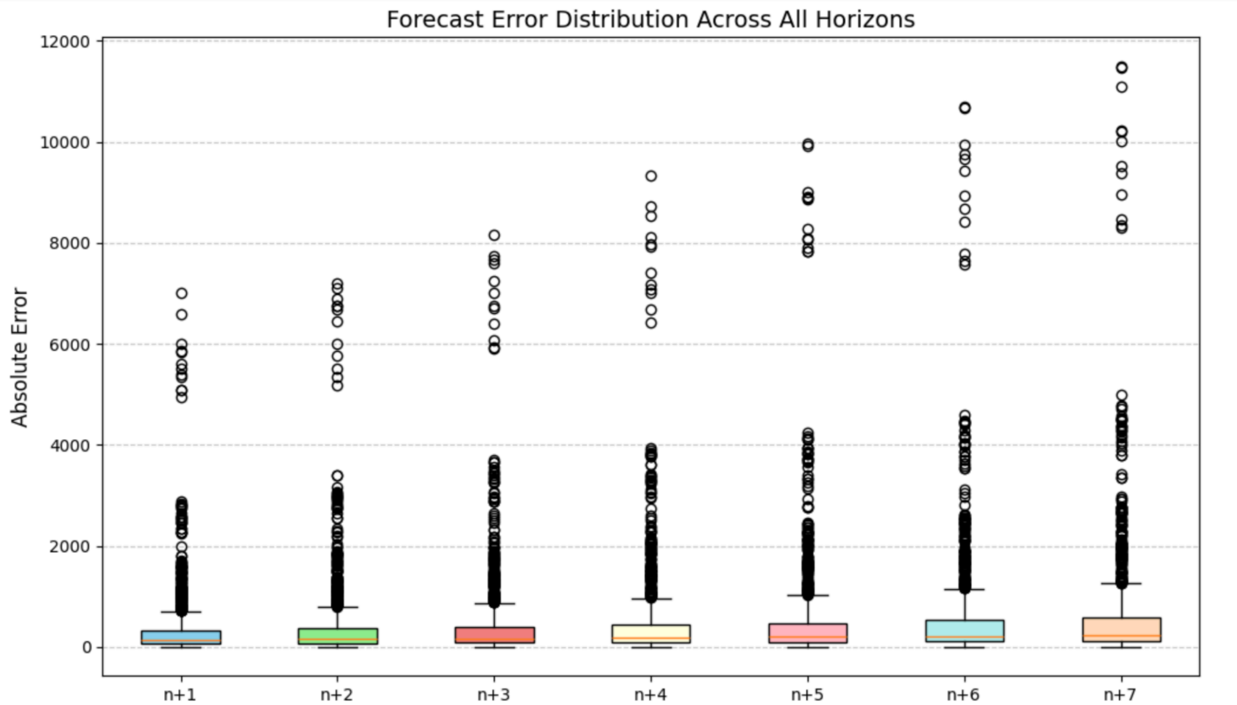


The chart shows the distribution of MAPE across all clients in our dataset. While the mean MAPE is 28.72%, the **median MAPE is only 4.19%**, which highlights that the majority of clients had relatively low prediction errors. The large gap between the mean and median is caused by a small number of extreme outliers with very high MAPE values (in some cases above 1000%), which significantly raise the average. This suggests the model is generally reliable and performs well for most clients, but there are a few clients with highly irregular or unpredictable usage patterns that affect the overall error statistics.



These metrics show that the model provides reliable predictions across a large and diverse user base. Additionally, we analyzed the forecast error by the horizon and observed that prediction errors increase slightly as the forecast window extends further, which is expected in recursive forecasting.

To better understand the performance by day, we visualized the absolute error distributions for each forecast horizon (rolling window validation). The results show tighter error bounds in short-term predictions (e.g., Day 1) and gradually wider distributions as the forecast steps increase. However, the majority of predictions remain within a reasonable range, indicating stable performance over time.



This Figure shows the absolute error distribution for each forecast horizon (Day 1 to Day 7). As expected in recursive forecasting, the errors tend to grow as the prediction window extends further into the future. The boxplots show that while short-term forecasts (e.g., n+1) maintain tighter error ranges, longer horizons (e.g., n+7) experience a wider spread, particularly in the upper quartiles and outliers. Still, the overall distribution remains stable, and most predictions stay within a reasonable error range, demonstrating consistent performance over multiple steps.

Part 6: Conclusion

Our study of electricity consumption forecasting using the Portuguese electricity data from 2011-2014 has yielded several important results. We initially explored multiple forecasting approaches including tree-based models, LSTM networks, and SARIMAX models. However, these initial approaches predicted only total consumption, which did not satisfy our goal of generating individual forecasts for each client.

To support the forecasting model, we incorporated exogenous features such as average temperature, user cluster labels, and cyclical time indicators (sine and cosine transformations of day-of-week and day-of-year). These features were stored in exog_dict and provided important contextual signals to help the model capture seasonality and user behavior dynamics.

To better achieve this objective, we conducted clustering using Dynamic Time Warping (DTW), which more effectively captures similarity in time series data compared to traditional K-means clustering. After grouping clients into clusters based on usage patterns, we initially assumed that clients within the same cluster would exhibit similar consumption behaviors. This assumption led us to train one model per cluster by concatenating the time series of clients within each cluster into a single sequence.

Both LSTM and SARIMAX models performed poorly using this training approach, even after extensive hyperparameter tuning. This underperformance was primarily because these models are fundamentally univariate time series models that rely heavily on cluster homogeneity. Since our clients were not evenly distributed across the five clusters (with most clients falling into a single cluster), only a few representative clients from each cluster could be selected for training due to computational constraints. This approach effectively created one model per cluster rather than truly predicting individual client behavior, ignoring important client-specific details.

To address these limitations, we implemented a multivariate time series approach using the HistGradientBoostingRegressor within a multi-series recursive framework. This model could simultaneously handle all clients while incorporating important features such as rolling statistics, temperature data, and cluster labels. The model achieved an MAE of 655.66, RMSE of 779.34, and a MAPE of 30.20% across all clients, significantly outperforming the SARIMAX approach (which had an average MAPE of 160.55% across clusters).

We utilized rolling window validation to rigorously evaluate our model's performance across different forecast horizons. This approach simulates real-world forecasting scenarios by training the model on a fixed window of historical data and then testing it on subsequent periods. Our results show that prediction accuracy gradually decreases as the forecast horizon extends, which is expected in recursive forecasting. The mean absolute error increases from 325.72 on day 1 to 573.19 on day 7, with the standard deviation nearly doubling from 617.07 to 1132.65. Despite this increase in error, the model maintains reasonable accuracy even at longer horizons.

The distribution of MAPE across clients reveals that most clients have relatively low error rates (median MAPE of 4.19%), while a smaller number of clients with more volatile consumption patterns contribute to the higher mean MAPE of 28.72%. This skewed distribution indicates that our model performs exceptionally well for the majority of clients but struggles with a subset of clients who have more unpredictable usage patterns.

The HistGradientBoostingRegressor's success can be attributed to several key advantages: its ability to handle nonlinear relationships between features, its effective management of high-dimensional data including both numerical and categorical variables, and its built-in handling of missing values. Most importantly, by leveraging a global model that learns shared patterns while still maintaining client-specific features, we achieved a scalable and accurate forecasting system applicable to all 370 clients in our dataset.

Overall, our forecasting results offer useful insights for both energy providers and customers. For utility companies, better predictions can help manage the electricity grid more efficiently and reduce unnecessary energy production. The model works across hundreds of clients, making it practical for large-scale use. For individual users, having more accurate forecasts can support better energy planning and cost savings. By learning both common trends and individual usage patterns, our approach can lead to smarter electricity use and more reliable energy systems.

Part 7: Challenge and Future Steps

7.1 Improving Cluster Balance

One of the main challenges we encountered was the imbalance in the clustering results. Out of 370 clients, over 280 were grouped into a single cluster, while the rest were spread across the remaining clusters. This suggests that our current clustering method (using DTW and Agglomerative Clustering) may not fully capture the diversity of client behaviors. In future work, we plan to explore more advanced clustering methods to improve the separation between client groups.

7.2. Limited Feature Availability

The original dataset only included client IDs and electricity usage over time, with no additional context. However, time series forecasting relies heavily on meaningful features. Without them, the model struggles to learn patterns beyond the raw values. To improve prediction quality, we had to manually engineer features like day-of-week, seasonal cycles, temperature, and clustering labels. In future steps, we aim to enrich the dataset with more external information, such as Holiday indicators, Client type (residential, commercial, industrial), Economic activity, or weather anomalies, which might improve model performance.

Part 8: Reference

Nhat-Duc, H., & Van-Duc, T. (2023). Comparison of histogram-based gradient boosting classification machine, random forest, and deep convolutional neural network for Pavement Raveling Severity Classification. *Automation in Construction*, *148*, 104767. https://doi.org/10.1016/j.autcon.2023.104767

Rodrigo, J. A., & Ortiz, J. E. (2022, October). *Global forecasting models: Modeling multiple time series with machine learning*. Global forecasting models: multiple time series forecasting with skforecast. https://cienciadedatos.net/documentos/py44-multi-series-forecasting-skforecast

Vivas, L. (2020, April 23). *Spain Portugal weather*. Kaggle. https://www.kaggle.com/datasets/luisvivas/spain-portugal-weather?resource=download&select=lisbon.csv

Part 9: Team Information

Group Name: YYYY
Group Member: Yueer Liu, Zimeng(Sandy) Li, Ziying Song, Yihan Yang