ID2223
Scalable Machine Learning and Deep Learning

Energy Load Prediction using weather data in Sweden

Project report
8 January 2025

Pierre Jousselin (pgmjo@kth.se)
Eron Ariodito Hermanto (ariodito@kth.se)

Introduction

Predicting energy load has become an increasingly significant challenge due to the rising complexity of energy systems and the global push toward sustainable practices. Accurate energy load forecasting is critical for efficient grid management, resource optimization, and ensuring stability in energy supply. It helps to anticipate demand surges, plan energy production, and manage costs effectively, especially in regions with high variability in weather conditions and energy consumption.

In this project, we focused on predicting the energy load in Sweden, which is divided into four distinct energy regions (SE1, SE2, SE3, and SE4). To build an effective predictive model, we combined historical and real-time energy consumption data with weather forecasts, recognizing the strong correlation between weather and energy usage. This report summarizes our methodology, the data sources we used, and the predictive modeling approach that enables us to estimate energy loads for the next 10 days. A summary diagram of the architecture of the system built is illustrated in figure 1.

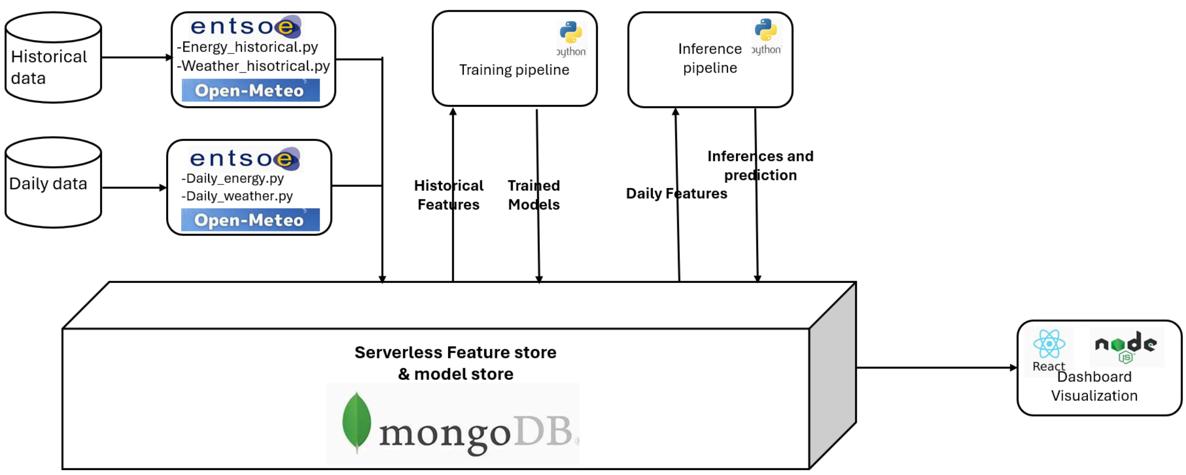

Figure 1. Summary Diagram

Data Collection and Preparation

Energy Load Data from ENTSO-E

To retrieve historical and daily energy load consumption, we utilized the ENTSO-E Transparency Platform, a trusted source of energy-related data in Europe. ENTSO-E provided hourly energy load data, which we averaged into daily values for consistency with our weather data. Access to the API required obtaining an API key, which was kindly provided by ENTSO-E after registration.

As Sweden has four different counties, we decided to keep them separated and have four different models so that each county has its own model.

Weather Data from Open-Meteo

Weather has a significant impact on energy consumption, as it influence the heating demand which requires electricity or in the summer cooling device. To incorporate this into our analysis, we used the Open-Meteo API to retrieve both historical weather data and daily 10-day forecasts. For each Swedish region, we selected four cities as weather sensors to represent the regional climate. The weather data included key variables such as daily temperature, precipitation, wind speed, and wind direction.

Data Storage and Management

Given the instability of some platforms like Hopsworks, we opted to use **MongoDB** for data storage due to its flexibility and reliability. We created three separate databases:

1. **Weather Database**: This contains weather data for each sensor, with fields for temperature, wind speed, precipitation, and corresponding county codes.
2. **Energy Database**: This stores daily energy load values for each county code.
3. **Feature View Database**: This acts as a preprocessed dataset for model training, containing daily concatenated weather data (aggregated across cities in each county) and corresponding energy load values.
4. **Inference Database** : which contains all of our model predictions in order to keep track of its performance.

Methodology

Daily Pipelines for Data Updates

To keep our predictions current, we implemented daily pipelines for both energy and weather data:

1. **Energy Pipeline**: This pipeline retrieves daily energy load data from ENTSO-E and updates the energy database.
2. **Weather Pipeline**: This pipeline fetches the next 10-day weather forecasts from Open-Meteo for 4 cities in each county and updates the weather database.

Model Training

To ensure robust predictions, we trained a separate model for each energy region (SE1, SE2, SE3, and SE4). This was necessary due to the significant differences in energy consumption ranges between regions—for instance, SE3 averages around 12,000 MW, while SE1 averages closer to 1,800 MW.

For each region:

- We trained Random Forest models using the feature view data.
- We stored the trained models in the database alongside performance metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), and R^2 score.

This modular approach ensures that the models capture region-specific characteristics effectively. In order to search the best hyperparameters we used the GRID SEARCH methods on those parameters: ***n_estimators ; max_depth ; min_samples_split; min_samples_leaf***

Inferences and predictions:

Once the daily data are retrieved and the latest weather forecast has been made, we use our trained model to make the predictions on the next days. Each of these predictions are then saved in our inference database that is used to monitor the models performance on our dashboard.

Dashboard Visualization

The dashboard takes the data from mongoDB and display it into prediction chart and hindchart. The purpose of the prediction chart is to visualize all prediction that is done using the system while the hindchart is used to compare the prediction of the system to actual load data as it updates trough the days. A map view is also available to visually see the activity of energy load within the 4 regions from its latest data, hovering over the map also shows the latest actual load data and the prediction of the next day.

The dashboard can be accessed through https://id-2223-project-dashboard-view.vercel.app/