ENERVISION

An AI-Powered Energy Consumption Prediction and Carbon Footprint Estimator

"Predict hourly, daily, and monthly energy usage with cost and CO_2 footprint insights."

AUTHOR: AURIC DUTT

PROBLEM STATEMENT:

With the growing concerns about energy conservation and carbon emissions, households and businesses need a simple yet accurate way to **predict their electricity consumption** and **estimate the associated costs and environmental impact**.
Traditional energy bills only provide retrospective information, leaving consumers unaware of how their daily habits and external factors (such as temperature or time of day) affect their energy usage in real time.

The lack of **predictive tools** makes it difficult for individuals and organizations to:

- Plan their electricity usage efficiently.
- Reduce unexpected high energy bills.
- Track and lower their carbon footprint.

There is a need for a **data-driven solution** that can forecast hourly, daily, and monthly energy consumption while also calculating **estimated costs** and **CO_2 emissions**, empowering users to make **proactive and eco-friendly decisions**.

DATASET CREATION AND FEATURE ENGINEERING

To capture energy-usage patterns across different time scales, I built **three separate datasets** derived from the original **UCI "Individual Household Electric Power Consumption" (France)** dataset (link: https://archive.ics.uci.edu/dataset/235/individual+household+electric+power+consumption) and enriched them with synthetic yet realistic features.

1. HOURLY DATASET

- **Source:** UCI dataset resampled to hourly intervals.
- **Synthetic Enhancements:**
  - **Weather Data:** Integrated real historical temperature from the **Meteostat API** to reflect outdoor conditions.
  - **Holiday/Weekend Flags:** Queried a **public holiday API** to mark French national holidays and weekends.
  - **Electrical Features:** Added user-focused metrics such as **power factor**.
- **Target:** Hourly energy consumption in kWh.

Dataset Link: **https://www.kaggle.com/datasets/auricdutt/hourly-energy-consumption-dataset-france**

2. DAILY DATASET

- **Source:** Aggregated hourly data into daily totals.
- **Feature Engineering:**
  - Cyclical **month encoding** (`sin` and `cos` of month number) to preserve seasonal effects.
  - Daily average temperature from Meteostat.
  - Binary flags for **weekend** and **holiday** status.
- **Target:** Total daily kWh usage.

DATASET LINK: **https://www.kaggle.com/datasets/auricdutt/daily-energy-consumption-dataset-france**

3. MONTHLY DATASET

Initially, the monthly dataset contained **48 real observations** (covering 2006–2010).
To extend the dataset and support long-term prediction (up to 2025), **~200 additional synthetic records** were generated. Synthetic samples were created by taking the **mean of historical values** for each feature and applying **random noise multipliers**, ensuring realistic variability while preserving the overall data distribution.

- **Source:** Summed daily data to monthly totals.
- **Features:**
  - Monthly mean temperature.
  - Sub-meter readings representing kitchen, laundry, and bedroom loads.
  - Average power factor per month.
- **Target:** Total monthly kWh consumption

DATASET LINK: **https://www.kaggle.com/datasets/auricdutt/monthly-energy-consumption-dataset-france**

DATASET DESCRIPTION

The project uses three engineered datasets—**Hourly**, **Daily**, and **Monthly**—all derived from the UCI "Individual Household Electric Power Consumption (France)" dataset and enriched with additional features.

COMMON BASE FEATURES (FROM UCI DATASET)

Each dataset retains the core electrical measurements, originally recorded at one-minute intervals:

1. **date** – Date of measurement in `dd/mm/yyyy` format.
2. **time** – Time of measurement in `hh:mm:ss` format (aggregated appropriately for daily and monthly datasets).
3. **global_active_power** – Household global minute-averaged active power (kW).
4. **global_reactive_power** – Household global minute-averaged reactive power (kW).
5. **Voltage** – Minute-averaged voltage (V).
6. **global_intensity** – Household global minute-averaged current intensity (A).
7. **sub_metering_1** – Energy sub-metering No. 1 (Wh of active energy). Represents the **kitchen** (dishwasher, oven, microwave; hot plates are gas).
8. **sub_metering_2** – Energy sub-metering No. 2 (Wh). Represents the **laundry room** (washing machine, tumble dryer, refrigerator, lighting).
9. **sub_metering_3** – Energy sub-metering No. 3 (Wh). Represents **electric water heater** and **air-conditioner**.

HOURLY DATASET (ENGINEERED)

- **datetime -** Combined date and time stamp in `YYYY-MM-DD HH:MM:SS` format, representing the start of each hourly interval.
- **hourly_energy_kWh – (Target variable)** Total household energy consumption for that hour (kilowatt-hours).
- **Global_active_power_kW -** Hourly mean of global active power drawn by the household (kilowatts).
- **Global_reactive_power -** Hourly mean of global reactive power (kilowatts).
- **Voltage -** Hourly averaged supply voltage (volts).
- **Sub_metering_1 -** Hourly energy (watt-hours) for **kitchen appliances** (dishwasher, oven, microwave).
- **Sub_metering_2 -** Hourly energy (watt-hours) for **laundry/utility room** appliances (washing machine, tumble-dryer, refrigerator, lighting).
- **Sub_metering_3 -** Hourly energy (watt-hours) for **water heater** and **air-conditioning** units.
- **power_factor -** Ratio of active to apparent power, indicating efficiency of electrical usage during the hour.
- **hour -** Hour of day (0–23) extracted from `datetime`.
- **weekday -** Day of the week (0 = Monday … 6 = Sunday).
- **month -** Month of the year (1–12).
- **is_holiday** – Boolean flag from a **Holiday API** marking French national holidays.
- **is_weekend** – Boolean flag: 1 if the observation falls on Saturday or Sunday, otherwise 0.
- **temp** – Hourly outdoor temperature (°C) fetched from the **Meteostat API**.
- **rhum -** Hourly relative humidity in % (from Meteostat API).
- **wspd -** Hourly average wind speed in m/s (from Meteostat API).

DAILY DATASET (ENGINEERED)

- **datetime** – Date stamp in `YYYY-MM-DD` format indicating the specific day of observation.
- **daily_energy_kWh** – **Target variable.** Total household energy consumption for the day (kilowatt-hours).
- **Global_active_power_kW** – Daily mean of global active power drawn by the household (kilowatts).
- **Global_reactive_power** – Daily mean of global reactive power (kilowatts).
- **Voltage** – Daily averaged supply voltage (volts).
- **Sub_metering_1** – Daily total energy (watt-hours) for **kitchen appliances** such as dishwasher, oven, and microwave.
- **Sub_metering_2** – Daily total energy (watt-hours) for **laundry/utility room** appliances including washing machine, tumble-dryer, refrigerator, and lighting.
- **Sub_metering_3** – Daily total energy (watt-hours) for **water heater** and **air-conditioning** units.
- **power_factor** – Daily ratio of active to apparent power, indicating electrical efficiency.
- **weekday** – Day of the week as an integer (0 = Monday … 6 = Sunday).
- **month** – Month of the year as an integer (1 = January … 12 = December).
- **is_weekend** – Boolean flag: `1` if the day falls on Saturday or Sunday, otherwise `0`.
- **is_holiday** – Boolean flag: `1` if the day is a French national holiday (via Holiday API), otherwise `0`.
- **temp** – Daily average outdoor temperature in °C (from Meteostat API).
- **rhum** – Daily average relative humidity in % (from Meteostat API).
- **wspd** – Daily average wind speed in m/s (from Meteostat API).

MONTHLY DATASET (ENGINEERED)

- **monthly_energy_kWh** – **Target variable.** Total household electricity consumption for the month (kilowatt-hours).
- **Global_active_power_kW** – Monthly mean of global active power drawn by the household (kilowatts).
- **Global_reactive_power** – Monthly mean of global reactive power (kilowatts).
- **Voltage** – Monthly average supply voltage (volts).
- **Sub_metering_1** – Monthly total energy (watt-hours) used by **kitchen appliances** (dishwasher, oven, microwave, etc.).
- **Sub_metering_2** – Monthly total energy (watt-hours) used in the **laundry/utility room** (washing machine, tumble dryer, refrigerator, lighting).
- **Sub_metering_3** – Monthly total energy (watt-hours) for **water heater** and **air-conditioning**.
- **power_factor** – Monthly ratio of active to apparent power, reflecting overall electrical efficiency.
- **month** – Integer representing the calendar month (1 = January … 12 = December).
- **temp** – Monthly average outdoor temperature in °C (retrieved via the Meteostat API).
- **rhum** – Monthly average relative humidity in % (from Meteostat API).
- **wspd** – Monthly average wind speed in m/s (from Meteostat API).

MODEL TRAINING AND EVALUATION

To provide accurate energy forecasts across different time horizons, **three independent machine-learning models** were developed—one each for **hourly**, **daily**, and **monthly** predictions.
Each was trained and validated on its respective engineered dataset, with careful selection of the algorithm based on data size, pattern complexity, and prediction goals.

HOURLY ENERGY PREDICTION MODEL

- **Algorithm: Random Forest Regressor (scikit-learn)**
- **Reason for Choice:**
  - Hourly energy usage shows **non-linear relationships** between electrical parameters (voltage, sub-metering) and weather.
  - Random Forest handles **high-dimensional, non-linear** interactions well, requires minimal scaling, and provides feature-importance insights.
  - Robust to outliers and effective with moderate data volume.
- **Key Features Used:** Sub-metering readings, hour of day, power factor, Voltage and temperature.
- **Training Details:**
  - Training/validation / testing split: **60 % / 20 % / 20%**
- **Performance Metrics:**
  - **MAE (Mean Absolute Error):** ~0.17 kWh
  - **RMSE (Root Mean Squared Error):** ~0.27 kWh
  - **R^2 Score:** ~0.90
  - **Interpretation:** The model explains **90 % of variance**, making it reliable for fine-grained hourly predictions.

DAILY ENERGY PREDICTION MODEL

- **Algorithm: Artificial Neural Network (ANN) – TensorFlow / Keras**
- **Reason for Choice:**
  - Daily energy demand is influenced by **seasonality** (month, weekend/holiday) and complex weather–usage patterns.
  - Thus the daily dataset contains a **rich mix of feature types**:
    - **Continuous numeric** (e.g. sub-meter readings, temperature).
    - **Cyclic** representations of time (month encoded as sine and cosine to preserve its circular nature).
    - **Binary categorical** indicators (weekend/holiday flags).
  - Such heterogeneous inputs create **complex, non-linear relationships** between weather, seasonal patterns, and household usage that simple linear models or decision trees may struggle to capture fully.
  - An ANN can **learn non-linear feature interactions automatically** without manual cross-term engineering, making it well-suited to:
    - Detect **seasonal cycles** from the month sine/cosine encoding.
    - Combine **binary signals** (is_weekend, is_holiday) with continuous weather variables.
    - Adapt to subtle shifts caused by temperature, humidity, and sub-meter behaviors.
- **Key Features Used:** Sub-metering readings, power factor, is_weekend, is_holiday, month (in sine /cosine) and temperature.
- **Splitting Details:**
  - Training/validation / testing split: **80 % / 10 % / 10%**

- **Network Architecture:**
  - **Input layer:** 9 engineered features (scaled numeric + cyclic + binary).
  - **Hidden layers:** Dense layers ([32] neurons) with ReLU activation to model high-dimensional interactions.
  - **Output layer:** Single neuron, linear activation for continuous kWh output.
  - **Optimizer:** Adam, Loss: Mean Squared Error.
- **Performance Metrics:**
  - **MAE:** ~0.28 kWh
  - **RMSE:** ~0.41 kWh
  - **R^2 Score:** ~0.85

MONTHLY ENERGY PREDICTION MODEL

- **Algorithm: Random Forest Regressor**
- **Splitting Details : Training /Testing split: 80% / 20%**
- **Feature Set:**
  - **Sub_metering_1, Sub_metering_2, Sub_metering_3** – monthly aggregated appliance-level consumption.
  - **Power_factor** – efficiency of power usage.
  - **Temperature (temp)** – average monthly outdoor temperature.
  - **Month** – numeric month indicator to capture seasonal effects.
- **Reason for Choice:**
  - Monthly consumption shows **non-linear relationships** between appliance loads, seasonal temperature changes, and efficiency.
  - Random Forest is an **ensemble of decision trees** that handles small datasets well, naturally models **non-linear interactions**, and is robust to outliers and multicollinearity.
  - It requires minimal feature scaling and provides feature-importance scores for interpretability.

- **Model Tuning:**

  - **GridSearchCV** was used to systematically search over hyperparameters.
  - **Best Parameters:**
  - `'max_depth': 10,`
  - `'max_features': 0.5,`
  - `'min_samples_leaf': 1,`
  - `'min_samples_split': 2,`
  - `'n_estimators': 100`
  - **Best Cross-Validation Score: 0.958** (R^2)

- **Performance on Final Test Set:**

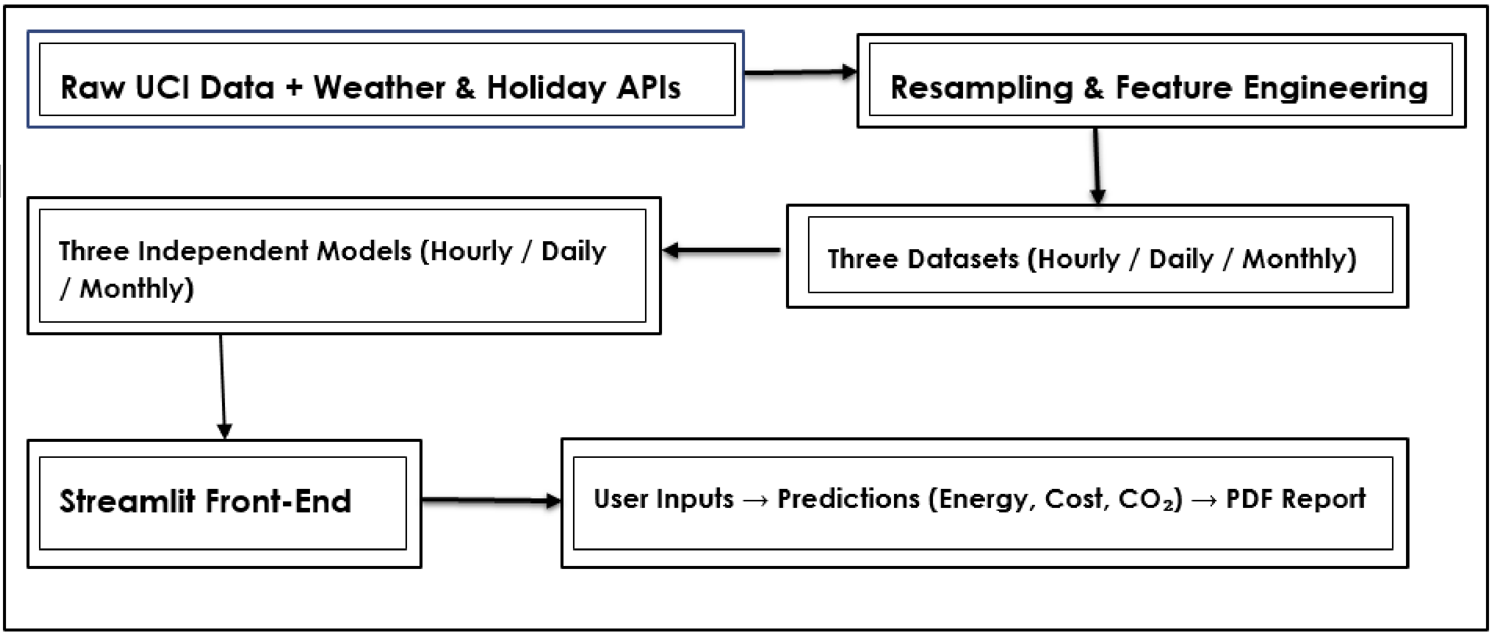
  - **MAE:** 25.17 kWh
  - **RMSE:** 30.61 kWh
  - **R^2 Score: 0.9699**

- **Interpretation** of **Metrics:**

  - The **very high R^2** (≈ 0.97) indicates the model explains most of the variance in monthly energy usage—expected when predicting totals that strongly follow seasonal patterns (month and temperature are powerful predictors).
  - The relatively **higher MAE/RMSE** in absolute kWh terms is partly due to:
    - **Small dataset (~245 records):** less data means the model can capture overall trends (good R^2) but individual month errors appear larger when expressed in kWh.
    - **Wide monthly consumption range:** a few high-consumption months inflate error magnitudes even if percentage error is moderate.
  - Despite the higher absolute errors, the model still provides **reliable trend forecasting** for planning monthly electricity usage.

System Architecture / Workflow

The Enervision workflow begins with the **raw UCI power-consumption data**, enriched by **weather** and **holiday APIs** to add seasonal context.
This data is **resampled and feature-engineered** into three granular datasets—**hourly, daily, and monthly**—with added metrics such as power factor, cyclic time encodings, and synthetic monthly records.
Each dataset trains its own model: **Random Forest** for hourly and monthly predictions, and an **ANN** for daily patterns.
A **Streamlit interface** lets users submit inputs, which are routed to the correct model to generate **energy, cost, and CO_2 forecasts**.
Results are displayed instantly and can be exported as a **PDF report**, creating a full end-to-end AI energy-prediction platform.



Application Deployment

- **Frontend:** The Enervision platform is delivered as an interactive **Streamlit web application** designed to resemble a modern multi-page website, with a custom theme and optional CSS styling.
- **Hosting:** The app is hosted on **Streamlit Community Cloud**, ensuring that it runs entirely in the browser without any local setup for the user. (It can also be launched locally with `streamlit run app.py`.)
- **Backend:** The backend is powered by **Python 3** and key libraries such as **scikit-learn**, **TensorFlow/Keras**, **pandas**, **numpy**, and **reportlab** for on-demand PDF generation.

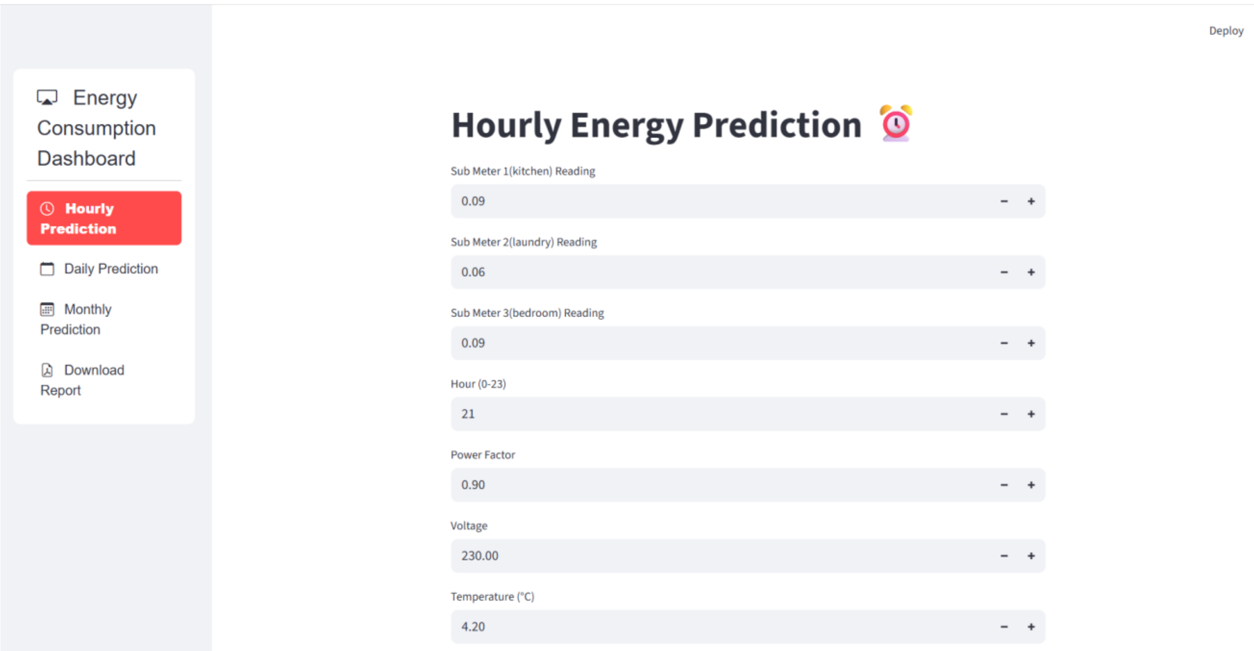
Through this interface, the system predicts **hourly, daily, and monthly electricity consumption**, estimates **costs and CO_2 emissions**, and provides a **downloadable PDF report** that consolidates all predictions.

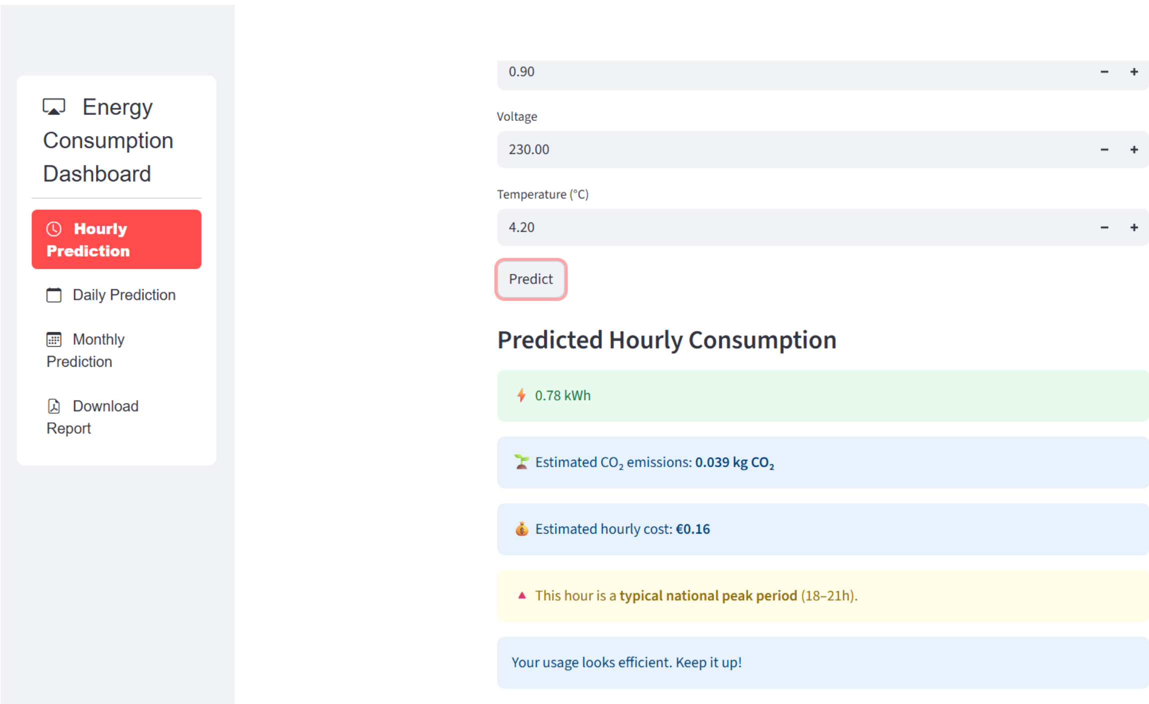
User Guide

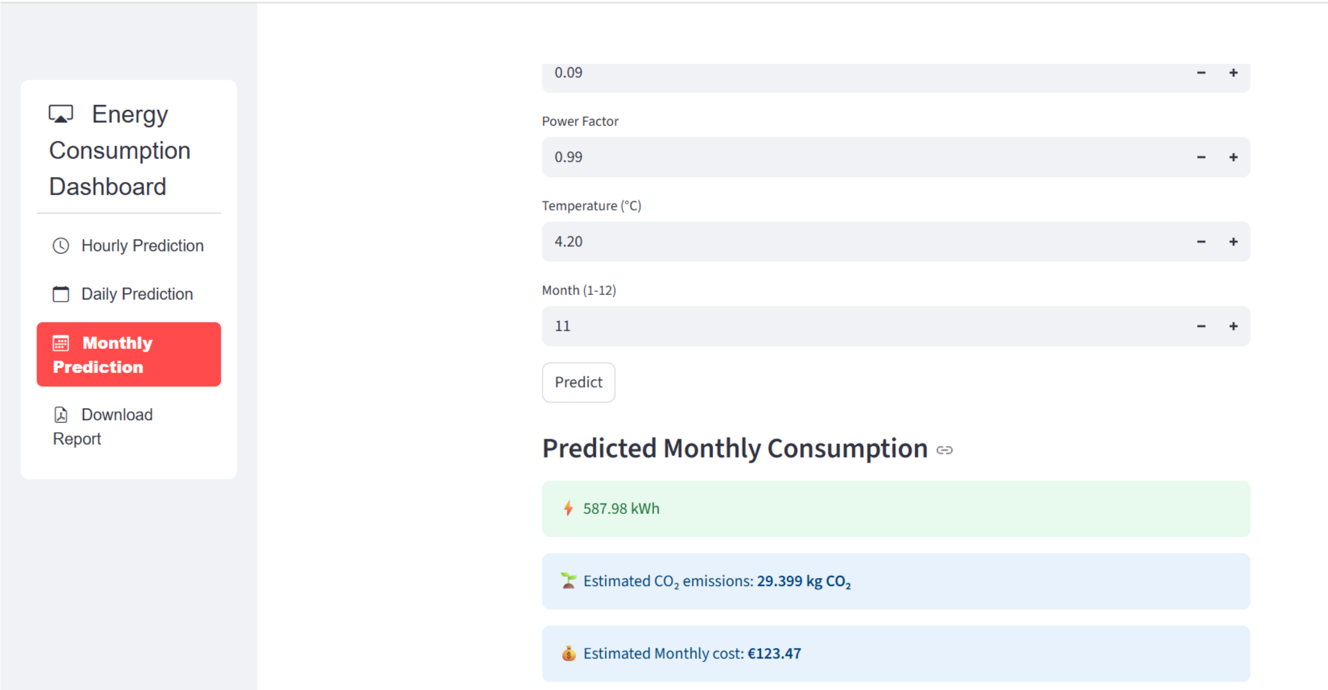
Follow these simple steps to use the application:

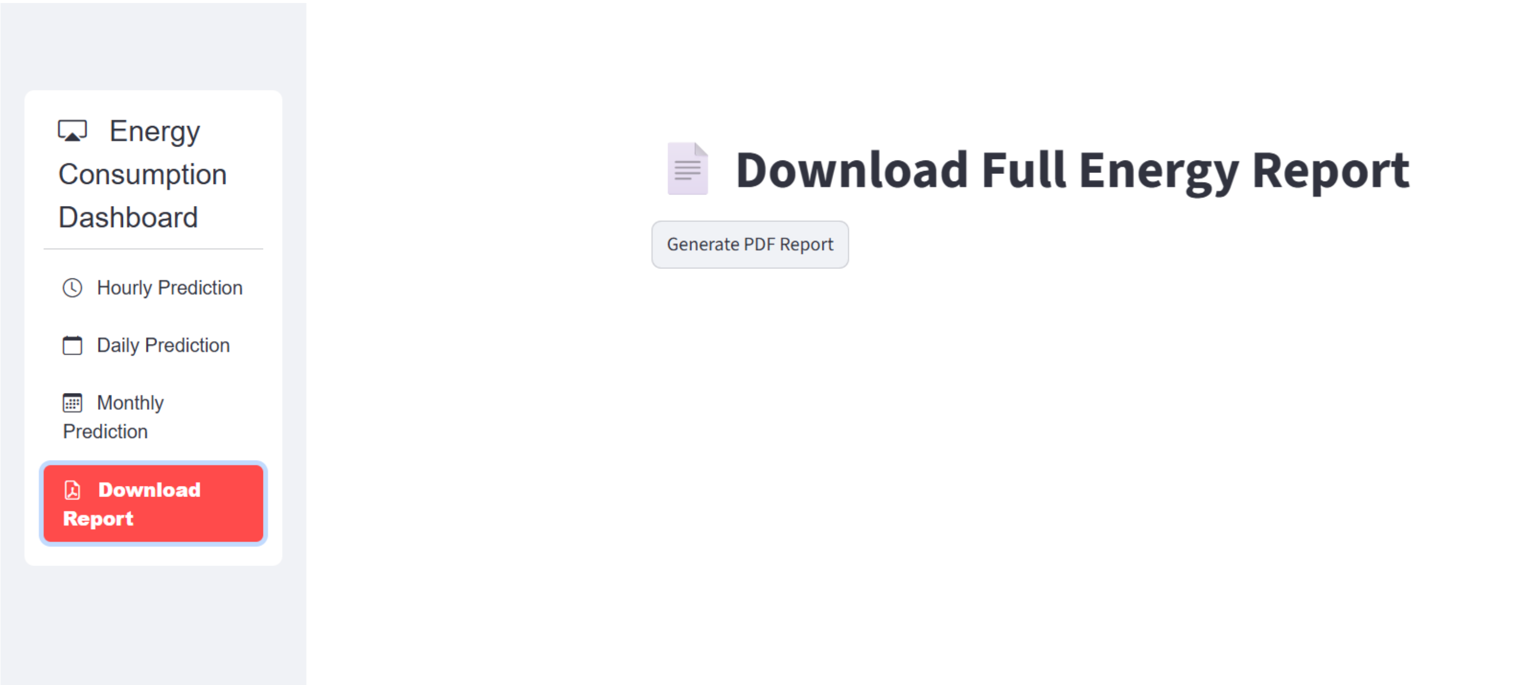
1. **Select a Prediction Tab** – Choose *Hourly*, *Daily*, or *Monthly* from the sidebar or top menu.
2. **Enter Input Values** – Provide current sub-meter readings, temperature, humidity, wind speed, and other required parameters.
3. **Run Prediction** – Click **Predict** to view estimated energy consumption (kWh), associated cost, and carbon footprint.
4. **Download Summary** – Click **Download PDF Report** to save a complete prediction summary for reference.

SCREENSHOTS:









REFERENCES

Datasets and tools used:

- **UCI “Individual Household Electric Power Consumption” dataset**
- **Meteostat Weather API**
- **Holiday API (France)**
- **Python packages:** `pandas`, `numpy`, `scikit-learn`, `tensorflow`, `streamlit`, `reportlab`, **etc.**