WEBER Tom, MICHEL Merlin

WATTWISE

Energy Forecasting App

https://streamlit-app-231204006378.europe-west1.run.app

May 12, 2025, University of Liège

PROJECT OVERVIEW

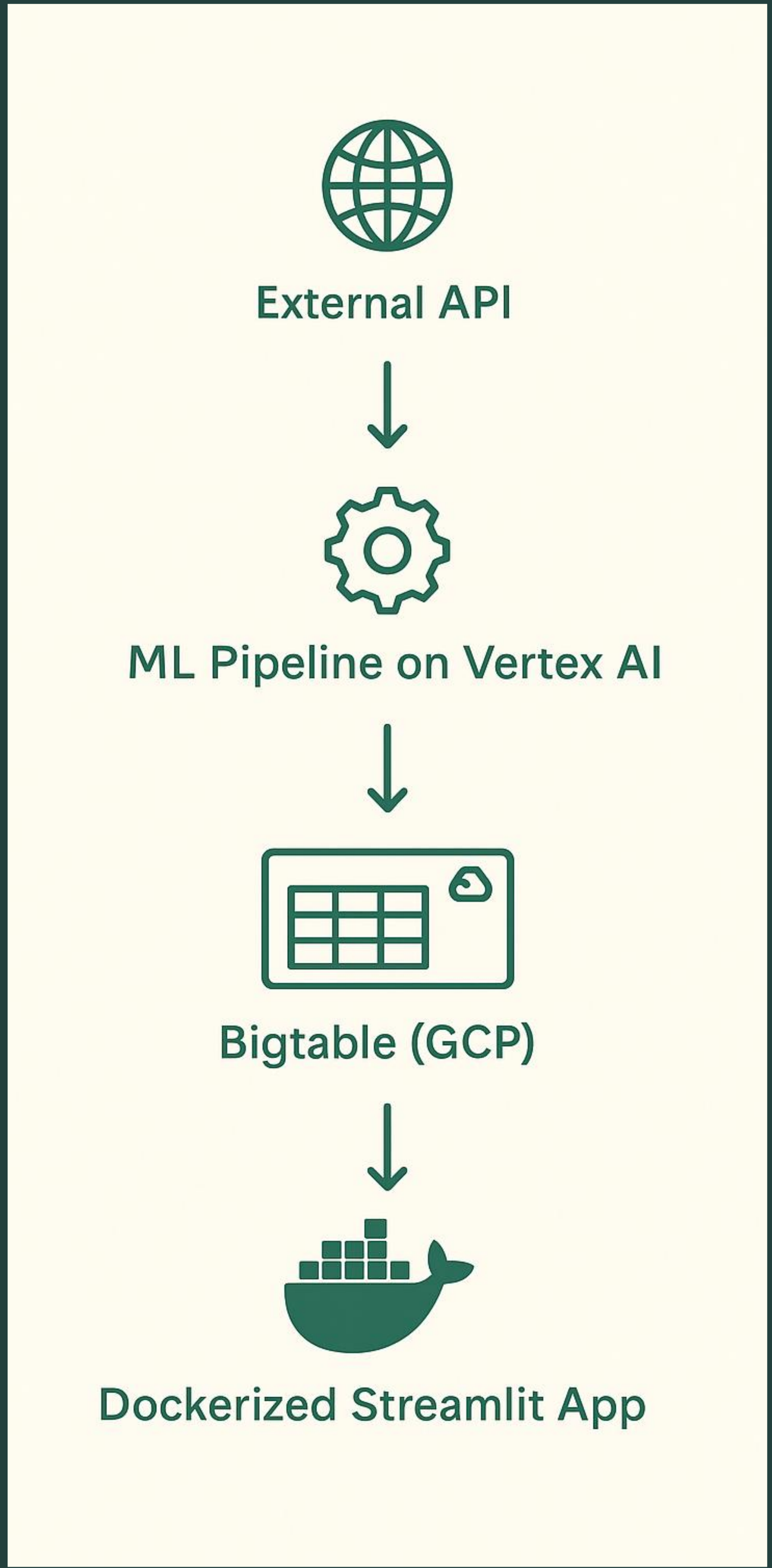
Design an automated, cloud-native system that retrieves meteorological and energy data, trains a forecasting model, stores predictions in a scalable database, and presents results in a user-friendly web interface.

01 End-to-end pipeline using GCP (Google Cloud Platform)

02 Scheduled data retrieval and model retraining

03 Scalable data storage with Bigtable

04 Real-time web interface using Streamlit and Cloud Run

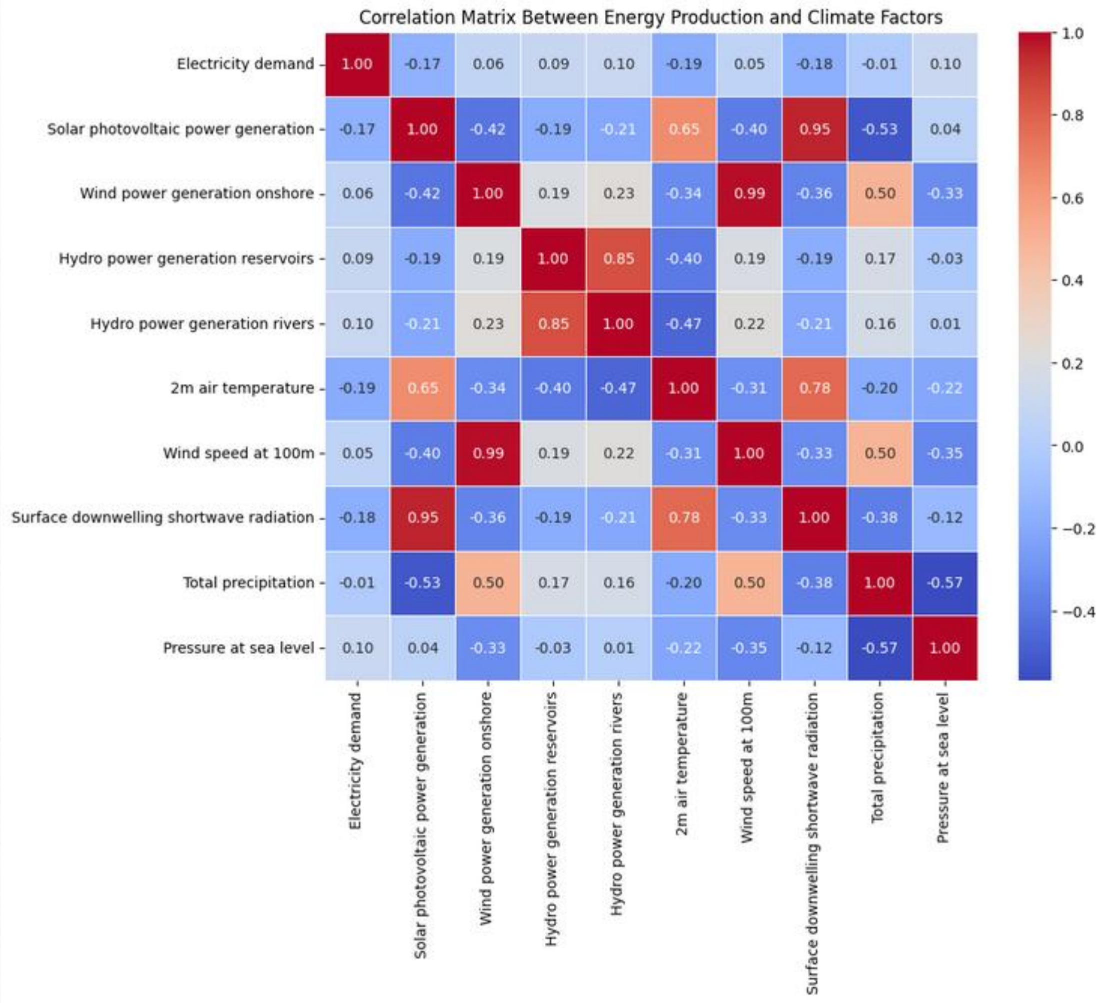

External API

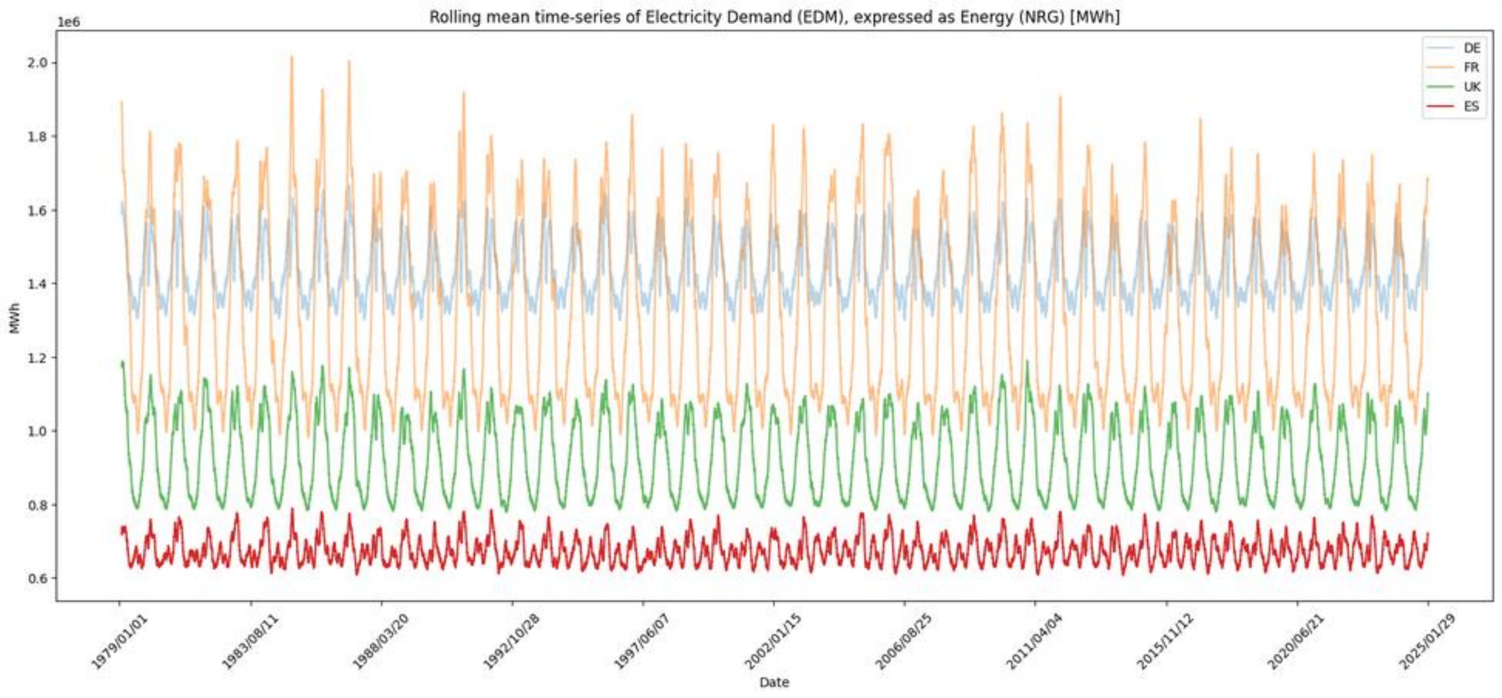
ML Pipeline on Vertex AI

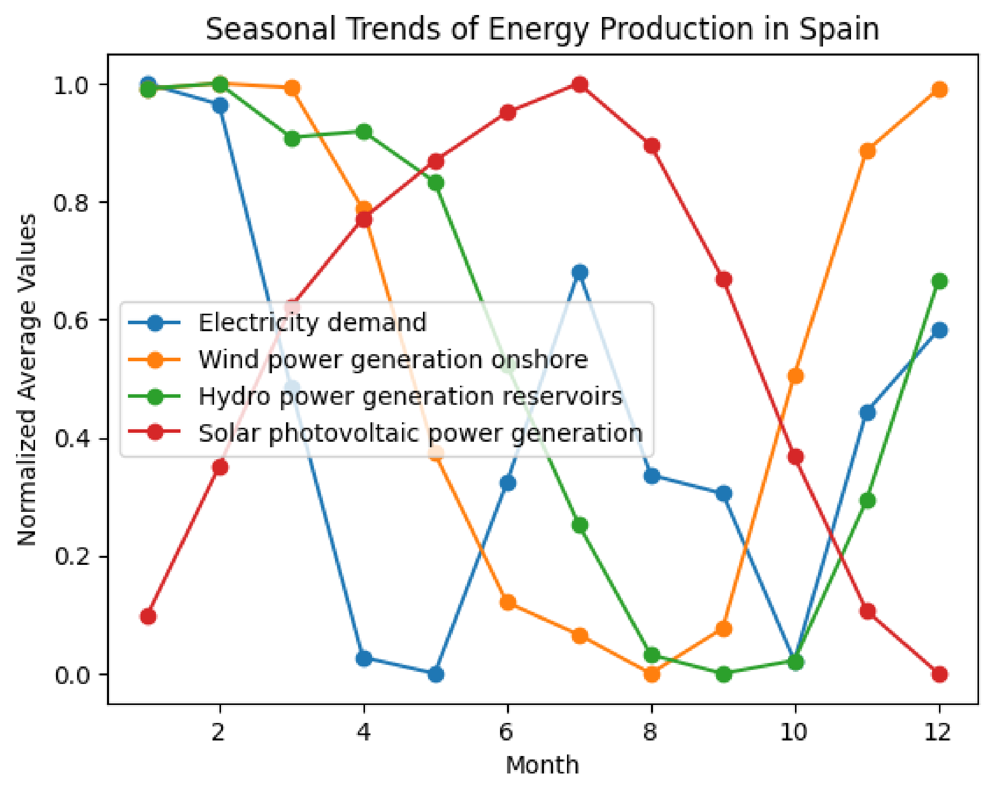
Bigtable (GCP)

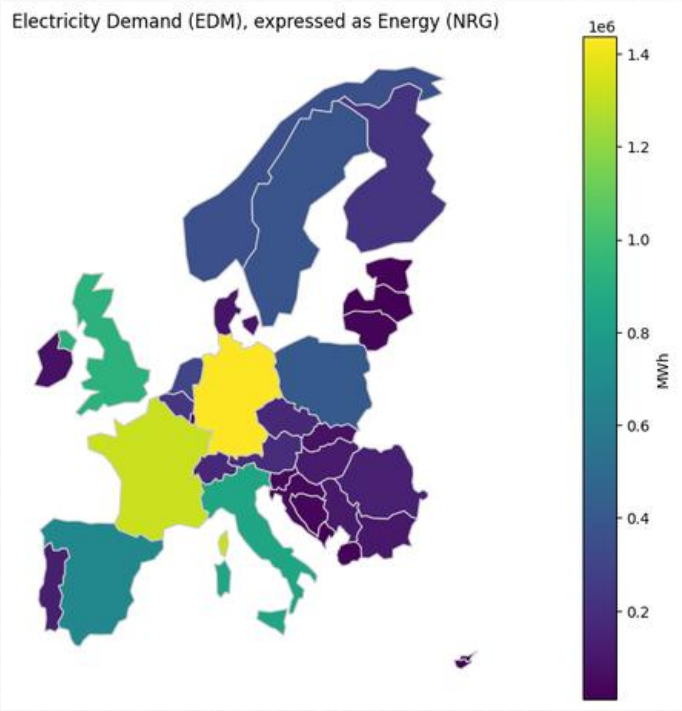
Dockerized Streamlit App

EDA


Correlation Matrix Between Energy Production and Climate Factors
Electricity demand 1.00 -0.17 0.06 0.09 0.10 -0.19 0.05 -0.18 -0.01 0.10
Solar photovoltaic power generation -0.17 1.00 -0.42 -0.19 -0.21 0.65 -0.40 0.95 -0.53 0.04
Wind power generation onshore 0.06 -0.42 1.00 0.19 0.23 -0.34 0.99 -0.36 0.50 -0.33
Hydro power generation reservoirs 0.09 -0.19 0.19 1.00 0.85 -0.40 0.19 -0.19 0.17 -0.03
Hydro power generation rivers 0.10 -0.21 0.23 0.85 1.00 -0.47 0.22 -0.21 0.16 0.01
2m air temperature -0.19 0.65 -0.34 -0.40 -0.47 1.00 -0.31 0.78 -0.20 -0.22
Wind speed at 100m 0.05 -0.40 0.99 0.19 0.22 -0.31 1.00 -0.33 0.50 -0.35
Surface downwelling shortwave radiation -0.18 0.95 -0.36 -0.19 -0.21 0.78 -0.33 1.00 -0.38 -0.12
Total precipitation -0.01 -0.53 0.50 0.17 0.16 -0.20 0.50 -0.38 1.00 -0.57
Pressure at sea level 0.10 0.04 -0.33 -0.03 0.01 -0.22 -0.35 -0.12 -0.57 1.00
Electricity demand
Solar photovoltaic power generation
Wind power generation onshore
Hydro power generation reservoirs
Hydro power generation rivers
2m air temperature
Wind speed at 100m
Surface downwelling shortwave radiation
Total precipitation
Pressure at sea level
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4


Rolling mean time-series of Electricity Demand (EDM), expressed as Energy (NRG) [MWh]
1e6
2.0
1.8
1.6
1.4
1.2
1.0
0.8
0.6
MWh
DE
FR
UK
ES
1979/01/01
1983/08/11
1988/03/20
1992/10/28
1997/06/07
2002/01/15
2006/08/25
2011/04/04
2015/11/12
2020/06/21
2025/01/29
Date


Seasonal Trends of Energy Production in Spain
1.0
0.8
0.6
0.4
0.2
0.0
Normalized Average Values
Electricity demand
Wind power generation onshore
Hydro power generation reservoirs
Solar photovoltaic power generation
2
4
6
8
10
12
Month


Electricity Demand (EDM), expressed as Energy (NRG)
1e6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
MWh

DATA RETRIEVAL

Data Sources

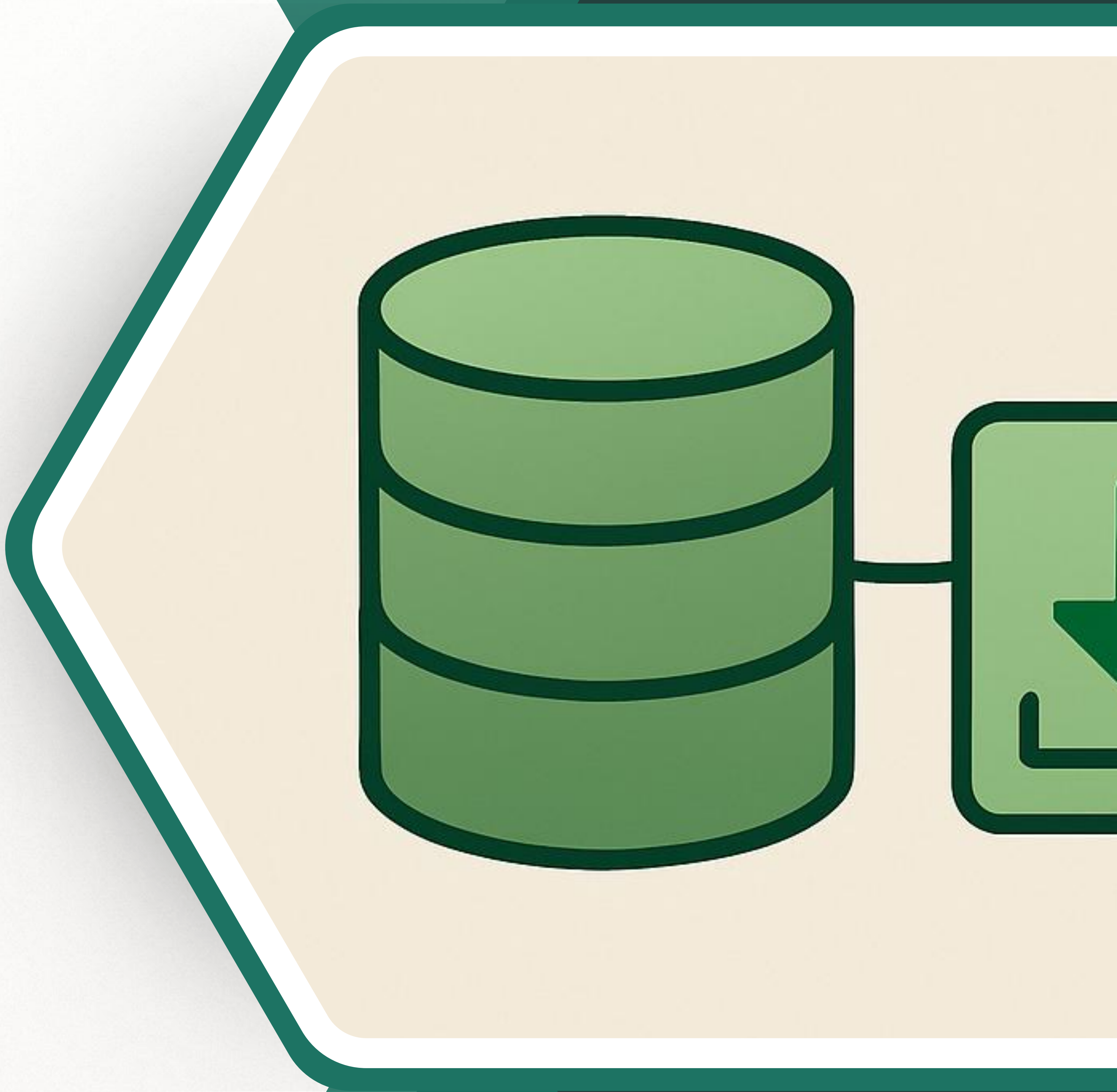
- External API for weather and energy usage
- Structured as daily time series by country

Components

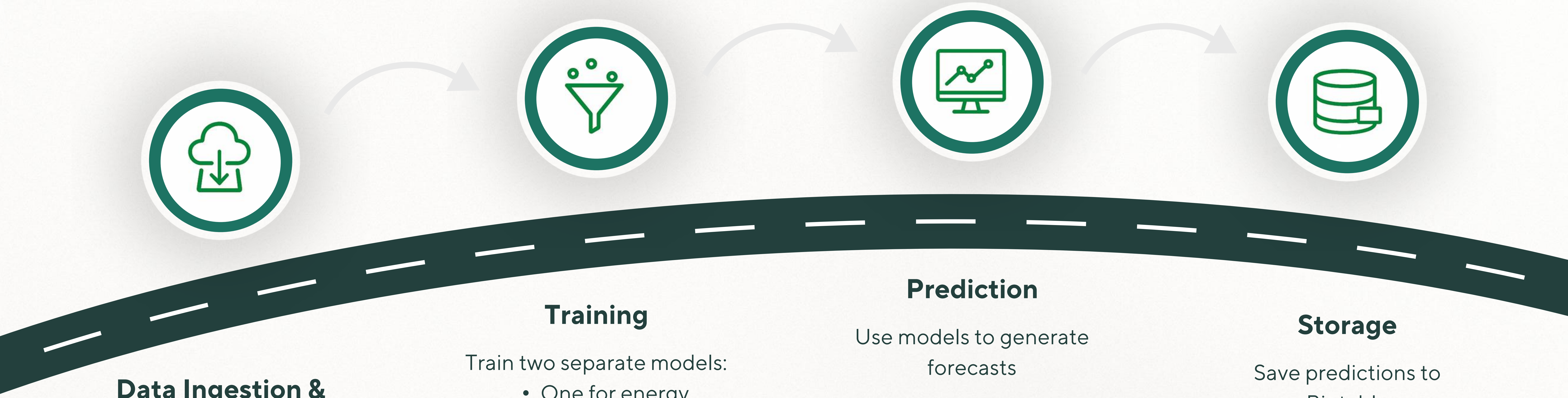
- data_ingestion: Downloads data and stores raw datasets in Cloud Storage
- data_preprocessing: Cleans, normalizes, and splits features & targets
- Data is split into:
- Energy features and target (Electricity Demand)
- Meteo features and targets (e.g. Wind Speed, Air Temperature)

Storage

- Preprocessed datasets stored in Google Cloud Storage
- Final predictions and features stored in Bigtable (keyed by COUNTRY#YYYY-MM-DD)



ML PIPELINE



Data Ingestion & Preprocessing

Load raw data from external APIs

Training

Train two separate models:

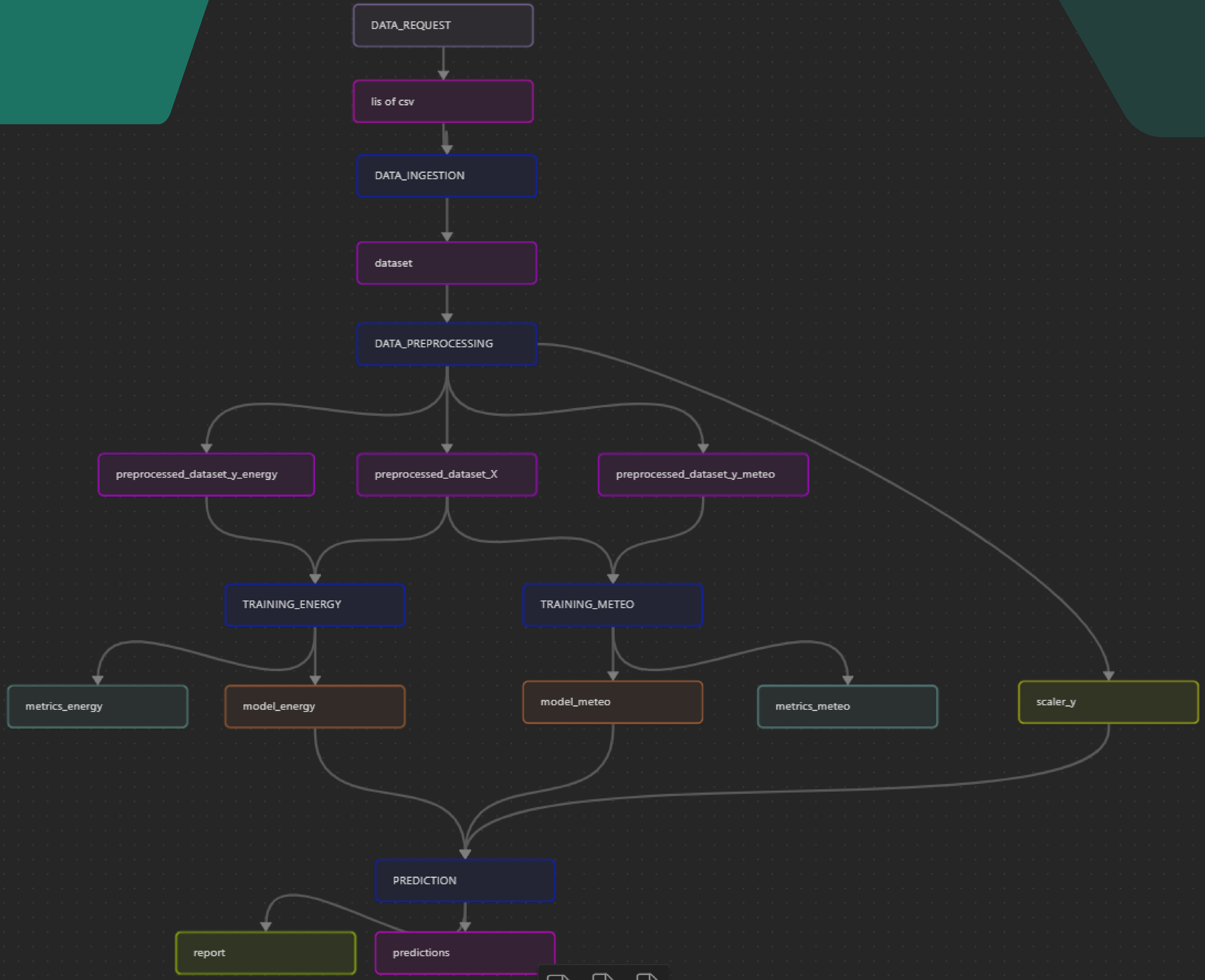
- One for energy prediction
- One for weather features

Prediction

Use models to generate forecasts

Storage

Save predictions to Bigtable



DATA_REQUEST

lis of csv

DATA_INGESTION

dataset

DATA_PREPROCESSING

preprocessed_dataset_y_energy

preprocessed_dataset_X

preprocessed_dataset_y_meteo

TRAINING_ENERGY

TRAINING_METEO

metrics_energy

model_energy

model_meteo

metrics_meteo

scaler_y

PREDICTION

report

predictions

SCALABLE BIGTABLE SCHEMA

Why ?

- **Scalable** for multiple countries and time ranges
- **Optimized** for time-series storage and fast reads
- **Easy integration** with ML pipeline and web frontend

Design

- **Two tables** :
  - input_data
  - predictions
- **One row per country per day** : {country_code}#{date} e.g. ES#2025-04-01, FR#2025-05-01
- **cf1 column family**

Access

- **Vertex AI ML pipeline (for writing input/output)**
- **Streamlit frontend (for reading data by country & date)**

WEB APPLICATION

01 Flask vs streamlit ?

02 Data retrieval

03 Features

DEPLOYMENT ON CLOUD RUN

Cloud Run

- Fully managed, serverless platform
- Scales automatically with demand
- No server provisionning or maintenance
- Pay-per-use billing model

Secure Integration

- Uses a dedicated service account with Bigtable read access
- No credentials stored inside the container
- Access managed via GCP IAM roles

Web App Access

- Cloud Run assigns a public HTTPS URL
- App listens on port 8080, required by Cloud Run
- Streamlit UI is accessible from any device

Fast Iteration Workflow

1. Modify code
2. Run gcloud builds submit
3. Re-deploy with gcloud run deploy
4. New version is instantly live

CI/CD

- pre-commit-hooks to check quality of code using Ruff
- Test some functionality of code using Pytest
- Issue running Google cloud command using python

DEMO

https://streamlit-app-231204006378.europe-west1.run.app