Residual Load Forecasting – Machine Learning Showcase

This project demonstrates an end-to-end machine learning workflow for forecasting the residual load in the German power grid. It was created as a showcase to highlight skills in data analysis, feature engineering, modeling, and reproducible setups with Docker.

Data sources

- Power grid data: SMARD.de (https://www.smard.de/home/downloadcenter/download-marktdaten)
- Weather data: Open-Meteo API (https://open-meteo.com/en/docs/historical-weather-api)

Project workflow

1. Data collection
   - Residual load data downloaded from *smard.de*
   - Weather data requested via *Open-Meteo API*
2. Data preparation
   - Cleaning, resampling (daily), and merging into a single dataset
   - Feature engineering (day of week, day of year, averaged climate features)
3. Modeling
   - Gradient Boosting Regressor (scikit-learn)
   - Evaluation using MAE and MAPE
   - Comparison against a simple baseline (yesterday = today)
4. Results
   - Gradient Boosting reduces MAE by ~52% compared to the baseline
   - Most important features: wind, temperature, day of week
5. Reproducibility
   - Project packaged in a Docker container for easy setup and execution

Notes

- Original datasets are not included in this repository.
  - SMARD data can be downloaded from the official website
  - Weather data can be requested via the Open-Meteo API
- This project is intended as a showcase of methods competence rather than a production-ready forecasting system.