Recommendations for Energy Systems Related Datasets

Compiled by Jose Machicao

1. Energy Demand and Consumption Prediction

These datasets are ideal for regression tasks where the goal is to predict energy demand or consumption.

a) UCI Energy Efficiency Dataset

- **Description:** Features such as building attributes (e.g., area, orientation) and climate factors (e.g., temperature, humidity) are used to predict energy demand for heating and cooling.
- **Tasks:** Regression for heating and cooling loads.
- **Link:** UCI Repository

b) Household Power Consumption

- **Description:** Records of a single household's electricity consumption over several years, including active and reactive power, voltage, and sub-metering.
- **Tasks:** Time-series prediction of power consumption.
- **Link:** UCI Repository

c) Smart Grid Stability Dataset

- **Description:** Data on smart grid stability under different conditions, with features like voltage and frequency measurements.
- **Tasks:** Predict stability or classify system states.
- **Link:** UCI Repository

---

2. Renewable Energy Prediction

Focus on datasets related to wind, solar, or other renewable energy sources.

a) Solar Power Data for Integration Studies

- **Description:** Large-scale solar generation profiles for different locations, collected by the National Renewable Energy Laboratory (NREL).
- **Tasks:** Predict solar generation based on weather data.
- **Link:** NREL Datasets

b) Wind Turbine SCADA Data

- **Description:** Supervisory control and data acquisition (SCADA) data from wind turbines, including wind speed, rotor speed, and power output.
- **Tasks:** Predict turbine power output or classify faults.
- **Link:** Kaggle Wind Turbine Dataset

c) Global Solar Atlas and Wind Atlas

- **Description:** Datasets providing solar irradiation and wind potential globally.
- **Tasks:** Predict renewable energy potential based on geographic and meteorological data.
- **Link:** World Bank Solar and Wind Atlas

---

3. Building Energy Management

Datasets focusing on energy efficiency in buildings.

a) ASHRAE – Great Energy Predictor III

- **Description:** Hourly energy consumption for commercial buildings, including features like outdoor temperature and building characteristics.
- **Tasks:** Predict building energy usage.
- **Link:** Kaggle Dataset

b) REDD (Residential Energy Disaggregation Dataset)

- **Description:** Appliance-level energy usage data for residential buildings.
- **Tasks:** Predict or classify energy usage patterns by device.
- **Link:** REDD Dataset

c) Open Building Dataset

- **Description:** Data on building energy usage, including occupancy patterns and environmental sensors.
- **Tasks:** Predict or classify energy efficiency.
- **Link:** Open Building Dataset

---

4. Smart Grid and Energy Markets

Datasets for modeling smart grids and market operations.

a) Electricity Load and Prices Dataset

- **Description:** Historical electricity prices and demand data for energy markets.
- **Tasks:** Time-series forecasting or price classification.
- **Link:** Kaggle Dataset

b) Pecan Street Dataset

- **Description:** Residential electricity and solar PV data, collected from homes equipped with smart meters.
- **Tasks:** Predict consumption or solar generation.
- **Link:** Pecan Street

c) Open Power System Data

- **Description:** Provides data on electricity production, consumption, and market prices in Europe.
- **Tasks:** Forecast demand, classify market conditions, or model energy system operations.
- **Link:** Open Power System Data

---

5. Fault Detection and Classification

These datasets are useful for learning about fault detection or classification in energy systems.

a) Electric Motor Temperature Dataset

- **Description:** Measurements from an electric motor, including temperature and voltage, used to classify operational states.
- **Tasks:** Fault detection or classification.
- **Link:** Kaggle Dataset

b) Condition Monitoring of Power Transformers

- **Description:** Data on dissolved gases in transformer oil for condition monitoring.
- **Tasks:** Classify transformer faults.
- **Link:** IEEE DataPort

---

6. Climate and Environmental Factors

These datasets include weather and environmental data that influence energy systems.

a) NOAA Weather Data

- **Description:** Hourly and daily weather observations globally, useful for renewable energy predictions.
- **Tasks:** Predict solar/wind energy output based on weather.
- **Link:** NOAA Climate Data

b) ERA5 Climate Reanalysis Data

- **Description:** High-resolution climate data, including temperature, wind speed, and solar radiation.
- **Tasks:** Predict energy demand or renewable generation.
- **Link:** ERA5 Dataset