Electricity Demand Forecasting Using AI

Project Lab Automation

Anton Valentin Dilg, Ramin Leon Neymeyer, Ramon Rennert

November 17, 2025

Technische Universität Berlin

Road Map

1. Create a dataset ☑
2. Graphical exploration ☑
3. Comparing and choosing the best model ⌛
4. Training and backtesting ⊗
5. Feature selection ⊗
6. Iterate through 3.-5. until satified with the results ⊗

Motivation

Problems:

- Thermal power plants face significant costs when deviating from planned output
- Better forecasts reduce imbalance penalties and operational inefficiencies
- Increasing variability from renewables makes accurate planning more challenging

Task

Problem Setting:

- Perspective of a coal- or gas-fired power plant submitting day-ahead generation bids

Objectives:

- Forecast the amount of electrical power, not the price
- Combine multiple datasets: grid load, generation by technology, weather data, plant production
- Build models that predict the hourly generation profile for the following day

Dataset Overview

- Includes combined heat and power (CHP) plant production, grid load, generation by source and weather
- Regional focus: 50Hertz area, specifically the CHP plant “Heizkraftwerk Berlin-Mitte”
- Data sourced from the Federal Network Agency (SMARD), freely available (CC BY 4.0)
- Time range: 2015–present, 15-minute resolution (SMARD) [2]

Power Plant & Market Data

Powerplant:

- Electrical Power generated by the plant
- 15-minute resolution
- Variable: `Power` in [MW]

Market Data:

- Day-ahead market data for the DE/LU bidding zone
- Before Oct 2018: price zone DE/AT/LU, afterwards DE/LU
- Variable: `Day-ahead Price` in [€]

Consumption & Generation Data

Consumption-related variables:

- Grid Load
- Pumped Storage
- Residual Load

Generation sources included:

- Biomass, Hydro, Wind Offshore/Onshore, Solar, Other Renewables,
- Lignite, Hard Coal, Natural Gas
- Pumped Storage, Other Conventional

Public Holidays & Temporal Features

Public holidays retrieved via API [1] and labeled per timestamp

- Variable: `Holiday` (boolean)

Additional temporal features extracted via `pandas`:

- `Hour of day` (integer)
- `Day of week` (integer)
- `Month` (integer)
- `Weekend indicator` (boolean)

Weather Data

Retrieved using Meteostat [3] (Tempelhof station)

Covers the full dataset period

Problem: Raw resolution differs

- Resampled to 15-minute intervals
- Applied linear interpolation to fill intermediate timestamps

Data Preprocessing

Missing or blank entries:

- Missing plant power → row removed
- Missing secondary features → tolerated if not critical

Daylight Saving Time issues (duplicate or missing hours):

- resolved by converting all timestamps to UTC

Final cleaned dataset exported as `.csv`

Purpose of Graphical Exploration

- Gain an intuitive understanding of temporal patterns in the dataset
- Support interpretation of model behavior and forecast plausibility
- First focus: the Power feature of the CHP plant
- Identify seasonal, weekly and daily patterns before modeling

Power Generation – Long-Term Pattern

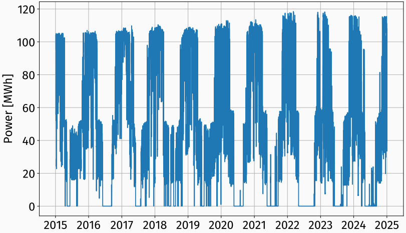

Figure 1: Power over Time.

- Strong seasonal cycle: high production in winter, low in summer
- Pattern repeats consistently across all years
- Indicates stable, long-term plant behavior driven mainly by heating demand

Power Generation – Yearly Distributions

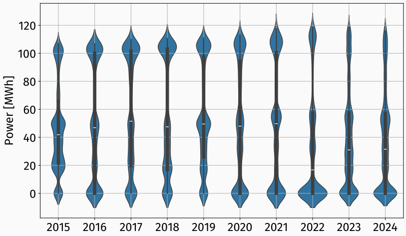

Figure 2: Distribution of Power by Year.

- Two clear modes: winter (high generation) vs. summer (low generation)
- Distribution shape stable across years → no major operational changes
- Confirms long-term stability of plant operation

Average Annual Cycle

Figure 3: Average Annual Course of Power.

- Aggregated curve shows typical yearly pattern
- Peak production in winter; sharp drop in June/July
- Seasonal heating demand is the dominant driver

Weekly & Seasonal Effects

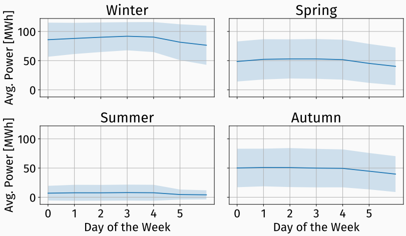

Figure 4: Average Power by Day – per Season.

- Clear reduction in power generation during weekends
- Consistent across all seasons
- Likely related to lower industrial activity
- Public holidays show similar behavior

Daily Profiles by Season

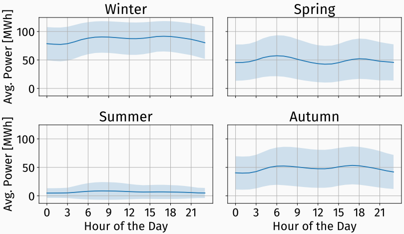

Figure 5: Average Power by Hour – per Season.

- Higher production during morning and early afternoon
- Slight reduction at night and early morning
- Daily pattern strongest in winter, smoother in summer
- Reflects typical demand-driven operation

Regional Generation Context (50Hertz)

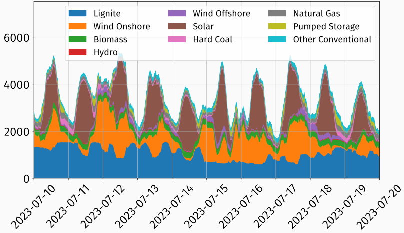

Figure 6: Power Generation by Source.

- Data reflects full 50Hertz region, not only the plant
- Lignite and wind form the regional base of electricity generation
- Solar contributes only during daylight; strong midday peaks
- Generation mix varies strongly with weather and time of day

Power Consumption & Residual Load

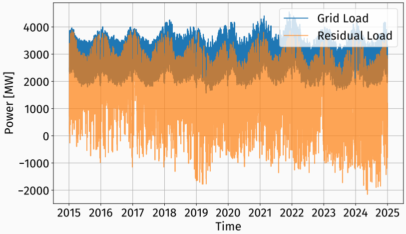

Figure 7: Grid and Residual Load Over Time.

- Grid load shows strong seasonal variation
- Residual load becomes increasingly volatile over time
- Rising share of variable renewables increases variability
- These fluctuations impact how strongly the CHP plant must adjust generation

Power & Temperature

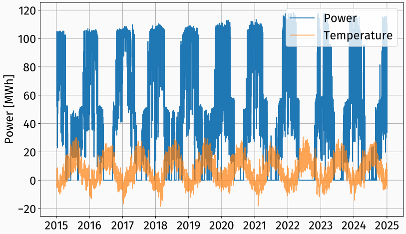

Figure 8: Power and Temperature over Time.

- Power and temperature show a strong inverse relationship
  - Cold periods → high generation
  - warm periods → low generation
- Seasonal temperature effects align with heating-driven operation

Correlation Analysis

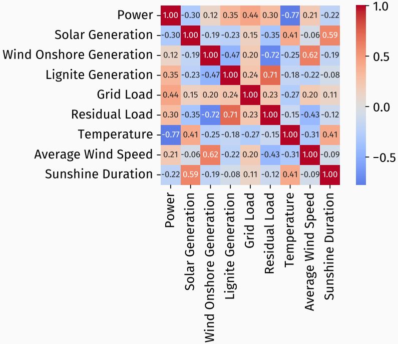

Figure 9: Correlation between Power, Generation and Weather.

- Power vs. temperature: strong negative correlation (~ 0.77)
- Weaker solar correlation due to spatial mismatch:
  - weather station (Berlin) vs. generation across entire 50Hertz area

References

M. Bandemer.
Digidates.de.
`https://digidates.de`, 2020.

Bundesnetzagentur.
Smard – strommarktdaten für deutschland.
`https://www.smard.de/home`, 2025.
Zugriff am 13. November 2025.

C. S. Lamprecht.
Meteostat python.
`https://meteostat.net/`, 2020.