Forecasting Daily Electricity Generation in Spain

Capstone Forecasting Project

Time Series Analysis

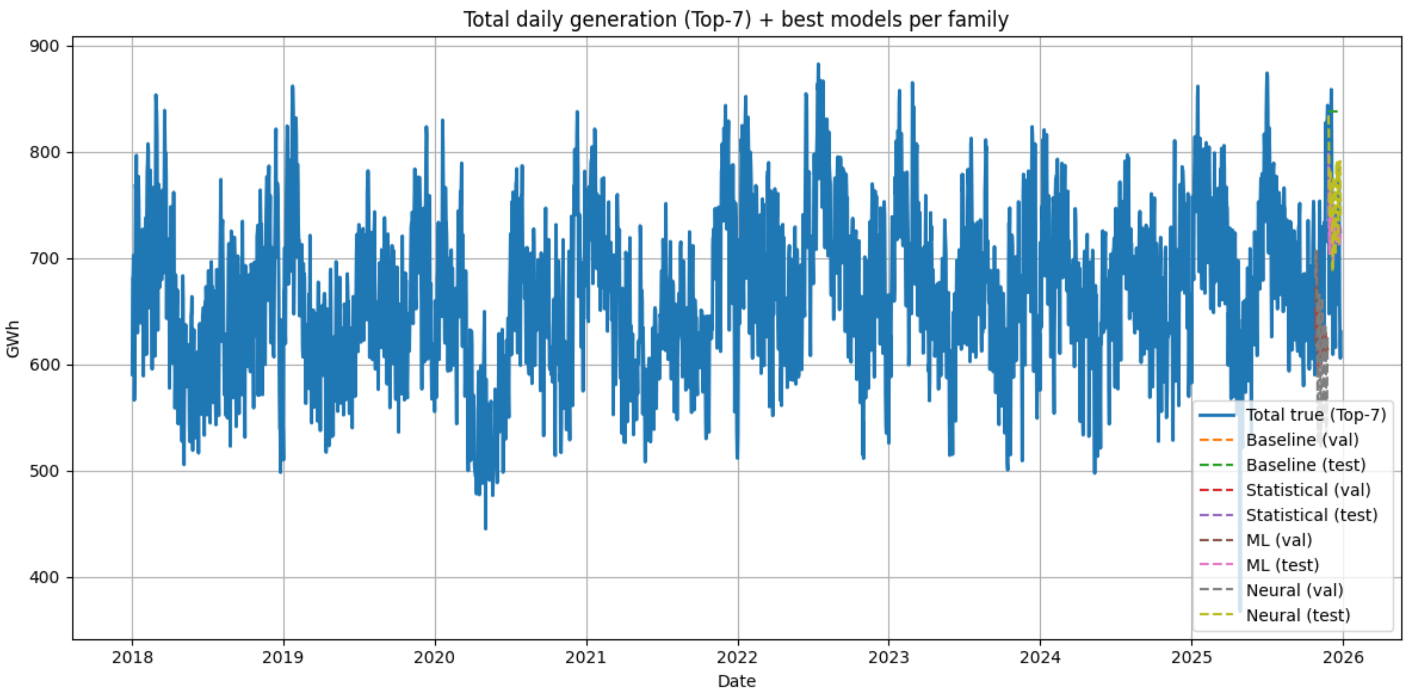
Team: Spanish Forecasters

Team & Contributions.

- Enrique Ruiz:
  - Data parsing/cleaning
  - Modelling
  - Evaluation
  - Plots
  - Business value
- Mireia Montoya:
  - Documentation
  - Slide deck
  - Interpretation of results
  - Presentation speaking roles
  - Plots

Business Problem

- **Problem:** Electricity generation varies due to the weather, demand or seasonality. This creates uncertainty for system planning.
- **Goal:** Forecast daily electricity generation for Spain by production technology for January 2026.



Why Forecasting Matters?

- Accurate forecast support:
  - Operational planning and balancing
  - Resource scheduling
  - Markt decisions and risk management
- Without forecast:
  - Reactive decisions
  - Higher operational risk
  - Potential extra costs

Solution Overview

1. **Business understanding:** define the forecasting goal and the value.
2. **Data understanding:** inspect the dataset, frequency, trends, seasonality, missing values.
3. **Data preparation:** clean, reshape, and build a final time series dataset.

4. **Feature engineering:** create lag features, calendar features…

5. **Modelling:** train model families used in class (baselines, statistical, ML, neural)

6. **Evaluation:** time-based validation/test split, metrics and plots.

7. **Output:** forecast January 2026, export predictions, save metrics for reporting.

Dataset Info

- Source: Spanish Electric Network (REE)
- Frequency: daily
- Target: electricity generation (GWh)
- Level: production technology

Focus:

- Top-7 technologies by average generation

Data Preparation

From Raw Data to Panel Dataset

1. Convert raw data to panel format:
   - **unique_id:** technology
   - **ds:** date
   - **y:** daily generation
2. Clean dates and numeric values
3. Ensures complete daily calendar
4. Validate and handle missing days

Model Families Tested

- **Baselines:**
  - Naive
  - Moving average
- **Statistical:**
  - AutoARIMA
- **Machine Learning:**
  - Random Forest
  - Gradient Boosting
  - Lag, calendar & rolling features
- **Deep Learning:**
  - NBEATS
  - NLinear

Evaluation Strategy

Time-Based Evaluation

- No shuffling (time series CV)
- Split:
  - Train
  - Validation (H = 31 days)
  - Test (H = 31 days)
- Metrics:
  - MAPE
  - MAE
  - RMSE
  - OPE
  - R2

Model Selection

- Best MAE on validation
- Baselines excluded from final forecast

Results & Forecast

January 2026 Forecast

- Best overall model selected across technologies
- Daily forecast for Jan 2026
- Outputs:
  - CSV (and Parquet if available)

Uncertainty

- Empirical prediction intervals
- Based on validation residual quantiles (95%)

Value & Conclusion

- **Value Generated:**
  - Better planning of energy balance and reserves
  - Improved risk-aware decision-making
  - Useful even with uncertainty
- **Conclusion:**