

Power Price Forecasting Case Study

Name: Yifei Lian
Email: yifeicristlian@outlook.com

1. Data Pipeline

Sources

- **Energy-Charts API** (public, CC-BY licensed)
- Market: **DE-LU** (Germany-Luxembourg bidding zone)
- Data period: **2024-01-01 to 2024-12-31** (12,847 hourly samples)
- Variables: Day-ahead prices, load, generation forecasts

Quality Assurance

Check	Status	Details
Missing values	[OK] Pass	Forward-filled minor gaps
Timestamp validation	[OK] Pass	UTC-aligned, continuous
Value range bounds	[OK] Pass	Prices within [-500, 500] EUR/MWh
Spike detection	[OK] Pass	No anomalies beyond 3sigma

QA output: `reports/qa/` contains per-column validation logs.

2. Forecasting Model

Baseline

- **Method:** Naive seasonal (same-hour yesterday)
- **MAE:** EUR32.90/MWh
- **R2:** -0.02 (no predictive power)

Improved Model

- **Method:** LightGBM gradient boosting
- **Features:** Hour-of-day, day-of-week, month, lagged prices (1h, 24h, 168h), rolling means, load forecast, generation forecast
- **Horizon:** 7-day ahead (168 hours)

Metric	Baseline	LightGBM	Improvement
MAE	EUR32.90	EUR18.38	-44%
RMSE	EUR46.64	EUR26.78	-43%
R2	-0.02	0.67	[OK]

Key drivers (by importance): `hour`, `price_lag_1h`, `price_rolling_24h`, `day_of_week`
!Feature Importance

3. Trading Application

Clean Spark Spread (CSS)

Measures CCGT profitability:

$$\text{CSS} = P_{\text{power}} - (P_{\text{gas}} \times \text{HR}) - (P_{\text{carbon}} \times \text{CI})$$

Where:

- Heat rate (HR) = 2.0 MWh_{th}/MWh_e
- Carbon intensity (CI) = 0.4 tCO₂/MWh_e
- Gas price = EUR35/MWh (TTF proxy)
- Carbon price = EUR50/tCO₂ (EUA proxy)

Trading Buckets & Signals

Bucket	Hours	CSS	Signal
Peak	08-20 M-F	EUR22/MWh	DISPATCH
Shoulder	06-08, 20-22	EUR18/MWh	DISPATCH
Off-peak night	22-06	-EUR7/MWh	OFF
Weekend	All Sat-Sun	-EUR12/MWh	OFF

Strategy: Run generation during peak/shoulder when CSS > 0; curtail during off-peak/weekend.

Signal output: `outputs/signals/` contains hourly buy/sell/hold signals.

4. AI/LLM Integration

Trading Agent

- **Provider:** Google Gemini (`gemini-2.5-flash`)
- **Persona:** Senior Power Trader (15+ years European power markets)
- **Input:** Daily forecast summary, CSS analysis, bucket view, risk flags
- **Output:** 3-bullet execution strategy (Position, Rationale, Risk)

Example LLM Output

1. POSITION: SELL full size DE Peakload for the next 168 hours.
2. RATIONALE: Peak CSS at EUR22/MWh offers significant profitability for CCGTs, indicating strong supply response and potential for price caps.
3. RISK: Unexpected demand surge or generation outage could invert CSS.

LLM logs: `reports/llm_logs/` contains full prompt/response history.

5. Engineering

Stack

- Python 3.14, LightGBM, pandas, pytest
- Config-driven: YAML files in `config/`
- Full test suite: **62 tests passing**

Reproducibility

```
pip install -r requirements.txt
python -m src.pipeline.main      # Run full pipeline
pytest tests/                    # Run tests
```

Directory Structure

```
data/raw/          -> Source CSV files
data/clean/        -> QA-validated data
data/features/     -> Model-ready features
outputs/signals/   -> Trading signals
reports/           -> Metrics, figures, LLM logs
```

Appendix: Figures

Figure	Description
feature_importance.png	LightGBM feature weights
model_comparison.png	Baseline vs improved
error_distribution.png	Residual histogram
metrics_comparison.png	MAE/RMSE/R2 comparison