Electricity is essential for modern society and a critical factor for economic and industrial growth. With the liberalization of electricity markets, electricity has become a tradeable commodity whose price fluctuates hourly. Accurate electricity price forecasting is crucial for grid management, trading, and planning operations.

The goal of this project is to compare different forecasting methods for predicting the next hour's electricity price, using past electricity prices and exogenous features related to energy generation and weather conditions. The datasets to work are `energy_dataset.csv` and `weather.csv` that contains 8 years of hourly data from 2015-01-01 to 2018-12-31. You must consider as test set the final 12 months.

The project must follow the CRISP-DM methodology and include code for its phases in the Python language:

1. Data understanding and preparation.
2. Data pre-processing.
3. Modelling: creation of models to predict electricity price 1-step ahead and 24-step ahead (one day).

Statistical models:

Because statistical models such as SARIMA/SARIMAX are computationally heavy for long hourly datasets, select a subset of the data (e.g., 1–2 consecutive years) that still contains relevant seasonal patterns.

Requirements:

- Stationarity testing (ADF)
- Differencing (if needed)
- ACF/PACF analysis
- SARIMA parameter selection
- SARIMAX using relevant exogenous variables
- Residual diagnostics (optional but recommended)
- Forecast 1-step ahead and 24-step ahead
- Plot real vs predicted values on the test set

Machine Learning models:

Use at least the following regression algorithms:

Linear Regression, Decision Tree Regressor, K-Neighbours Regressor, Support Vector Regression Bagging Regressor, Gradient Boosting Regressor, Random Forest Regressor, XGBoostRegressor, LGBMRegressor.

Requirements:

- Build supervised sliding-window datasets
- Compare several window sizes
- Use TimeSeriesSplit for model selection and tuning
- Perform hyperparameter tuning
- Produce plots comparing predictions vs. real test data

Deep Learning models:

Use LSTM and GRU networks.

Requirements:

- Use EarlyStopping
- Provide training/validation loss curves
- Perform basic hyperparameter tuning (units, layers, batch size, etc.)
- Plot real vs predicted values on the test set

4. Evaluation

   For all models (statistical, ML, and DL):

   - Make a table with MAE, RMSE, MAPE
   - Compare against baseline model: Seasonal Naïve

5. Model Selection

   Select the best-performing model and justify your choice based on:

   - Metrics performance
   - Stability across folds
   - Complexity / computational cost
   - Practical interpretability
   - Prediction quality over the full test period

Submit a written report that documents, in detail, the steps you took to arrive at your solutions. You should explain how you interpreted the most relevant graphical figures, how you cleaned and pre-processed the data, how you assessed the models, the commitments you made throughout their development and the conclusions you found.

Deadline and submission instructions

The project should be submitted to Moodle (course page) by **23:59 on Sunday, 21 December 2025**.

- Submissions received until 23:59 on 23 December 2025 incur a **10% grade penalty**.
- After 23 December 2025: Not accepted (grade = 0).

What to submit

- A single **ZIP** containing all code and the report.
- Filename format: MINDD-XXX-NumberX-NumberY.zip
  - XXX = your PL teacher's acronym (MFC – Fátima Rodrigues; AZC – Catarina Figueiredo).
  - NumberX, NumberY = the student numbers of each group member (add more numbers if your group has >2 members).
- Only **one member group** submits the project on Moodle.

Presentations and grading

- The PL teacher will assess each group during the first week of January, according to previous schedule. All group members' attendance is mandatory.