Analyzing Electricity Market data using Machine Learning

The electricity market is influenced by various factors such as demand, renewable energy generation, weather conditions, and fuel prices. Accurate forecasting and analysis are crucial for market participants to optimize operations and make informed decisions.

Objective

Work with Spanish historical electricity market data (find it on the ESIOS-REE webpage) to:

- **Supervised Learning (Regression):** Predict the day-ahead spot electricity price in Spain.
- **Unsupervised Learning (Clustering):** Group daily electricity price profiles into clusters to identify patterns.
- **Principal Component Analysis (PCA):** Reduce data dimensionality and test its effect on your models.

Tasks and guidelines

TASK 1) EDA: Exploratory Data Analysis (EDA) of potential input variables for your forecasting model.

TASK 2) Develop a machine learning model (supervised) to forecast the Marginal electricity price in Spain. → **Kaggle** competition.

To obtain the target data, access the Esios website and download historical spot electricity prices. Possible input data can also be searched on this website. Some general guidelines for developing your model are detailed below.

1) Understand your data set. Use descriptive statistics and data visualization tools.
2) Data analysis of the data. Important! Prepare your dataset. Clean it and transform it, if needed.
3) Think about possible feature selection for your model.
4) Split the dataset into training and test (the test set must be the one provided by Kaggle).
5) Train your model using different ML.
6) Evaluate your model, adjust hyperparameters and retrain it if needed.
7) Validate your model using the following regression metrics: RMSE, MAE and R2.

- Granularity: hourly.
- Provide the test error metrics results using the test data from 01/09/2025 – 30/09/2025 (1 month).

TASK 3) Create a clustering model to group the daily SPOT electricity price profiles in different clusters.

Identify patterns in the electricity prices data based on historical data and other features, if necessary. Are they similar? Can you identify any patterns among the different groups detected?

TASK 4) Train and test the regression model using PCA for the input data. What are the main differences between the 2 model results?

Submission Requirements:

Jupyter Notebook:

- Include all your code for data analysis, model development, and evaluation. Inside this file, SEPARATE the different Tasks.
- Ensure the notebook is well-organized with clear explanations and comments.

Report (PDF):

- Summarize your methodology, findings, and conclusions. It should have an index and sections explaining the main key findings. Include key visualizations and tables.
- Discuss any challenges faced and how you addressed them.

Deadline: Final Report to be submitted before the end of the day on December 19th. Be sure to cite external resources or libraries you use. Your work should be original and reflect your understanding of machine learning concepts applied to real-world data.