MadeAI 2025 - Modelling, Data Analytics and AI in Engineering
7-11 July 2025, Porto, Portugal

Energy Flow Forecasting using LightGBM

Zahra Zamanian[1,*], Noémie Vlaminck[2], Caroline Sainvitu[2], François Vallée[3], Laurent Dewasme[1] and Alain Vande Wouwer[1]

[1]Systems, Estimation, Control and Optimization (SECO), University of Mons, Belgium, E-mail: zahra.zamanian; laurent.dewasme; alain.vandewouwer@umons.ac.be
[2]Cenaero, Belgium, E-mail: noemie.vlaminck; caroline.sainvitu@cenaero.be
[3]Electrical Power Engineering, University of Mons, Belgium, E-mail: francois.vallee@umons.ac.be

Abstract: Accurate forecasting is a critical issue for energy transactions, grid stability, and operational planning since energy is necessary for both economic growth and the transition to low-carbon sectors. Energy management in Europe is made more complex by changing regulations, the integration of renewable energy, and market dynamics. To maintain energy security, stabilise prices, and balance supply, interconnectors such as the 140-kilometer NEMO Link between Belgium (BE) and Great Britain (GB) are essential. However, the complex or even stochastic nature of the features governing supply, demand, and transmission limitations make energy transaction forecasting particularly difficult.

Energy systems are frequently analyzed using classical statistical tools (CST), such as trend analysis, correlations, and causal relationships. However, when working with complex temporal connections and a large number of variables, these methods could not always offer enough insights [1]. In this context, machine learning (ML) methods, particularly Gradient Boosting Decision Trees (GBDT), offer practical tools for feature selection and time series forecasting. Among GBDT algorithms, LightGBM [2] has emerged as a promising candidate due to its computational efficiency, ability to handle large datasets, and integrated feature selection process. For an initial analysis, LightGBM is particularly useful for identifying key features that influence energy flow, providing a structured approach to feature selection that enhances model interpretability and predictive accuracy. Its ability to rank feature importance allows analysts to identify the most influential variables, enhancing model efficiency and generalization.

To improve energy trading strategies, cost stability, and grid reliability, this work focuses on developing a 48-hour-ahead energy flow forecasting strategy for the NEMO Link using collected market, environmental, and capacity data. To do this, we used datasets covering the period of January 2021 to June 2022 with hourly measurements of 41 variables, provided by N-SIDE[1] with the help of the Cenaero research centre.[2] Data preprocessing was first achieved to capture temporal dependencies. Then, time-lagged features at 1-hour, 2-hour, 1-day, 2-day, and 1-week intervals were incorporated, with an additional 1-month lag for the target variable, yielding 245 features. The feature selection was based on LightGBM feature importance scores, with a threshold $\Gamma = 18$ to retain the top 33 most relevant features, balancing predictive accuracy and model efficiency. Fig.1 illustrates the dataset split; 80% of the data was used for training and 20% for validation. Positive values indicate exports from BE to GB, while negative values represent imports from GB to BE. Flow fluctuations are typically influenced by market demand, supply conditions, and transmission capacity.

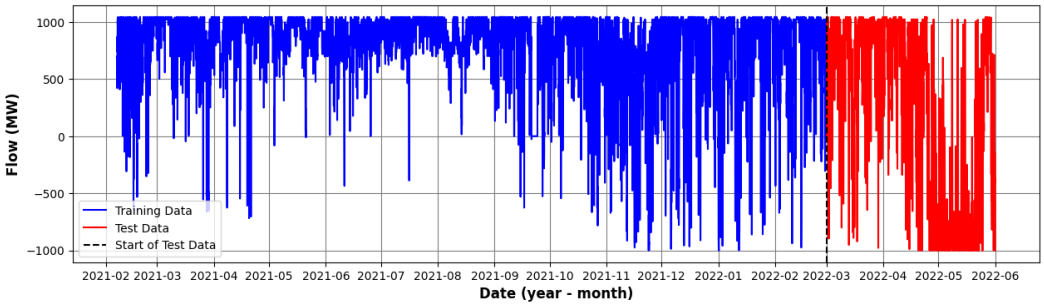

Figure 1: Output flow training (blue) and testing (red) data.

[1]N-SIDE is a global company with offices in Belgium, the USA, and Japan, which develops software solutions for complex challenges involving time series. This study was conducted in collaboration with its Belgian office.

[2]Cenaero is a Belgian non-profit applied research centre supporting the TRAIL grand challenge[3].

[3]Trusted AI Labs (TRAIL) is a Belgian initiative uniting universities and research centers in the Walloon and Brussels regions to advance AI research and innovation.

A LightGBM model was trained for forecasting, utilizing its histogram-based gradient calculation and leaf-wise tree growth to improve efficiency and predictive performance. Hyperparameter tuning was achieved using the Optuna framework. A systematic search across the hyperparameter space was conducted to select, for instance, the number of trees and learning rate, with cross-validation ensuring optimal selection.
Given a data set $(x_i, y_i)_{i=1}^{N}$, where x represents input features, y the target variable, and N is the number of samples, the objective is to estimate a function $F(x)$ that minimizes the expected loss function:

$$F^* = \arg\min_F \mathbf{E}_{x,y}[L(y, F(x))]$$

To forecast the next 48 hours with the trained model, a recursive multi-step approach was considered. Fig.2 illustrates the performance of a 48-hour-ahead recursive multi-step forecasting approach, where the model predicts each future step sequentially using prior forecasts as inputs. The predicted values are compared with actual energy flow measurements on unseen data, effectively serving as a time series cross-validation method, and a 95% confidence interval (CI) is provided to quantify the uncertainty. This validation process ensures the model's consistency and generalization ability when forecasting multiple steps ahead. The findings demonstrate consistent prediction performance, with actual values typically staying within the CI boundaries. The LightGBM model is a valuable tool for operational planning in energy trading and grid management, as evidenced by its attained root mean square error (RMSE) of 169 MW, which indicates that it accurately captures fluctuations in energy flow.

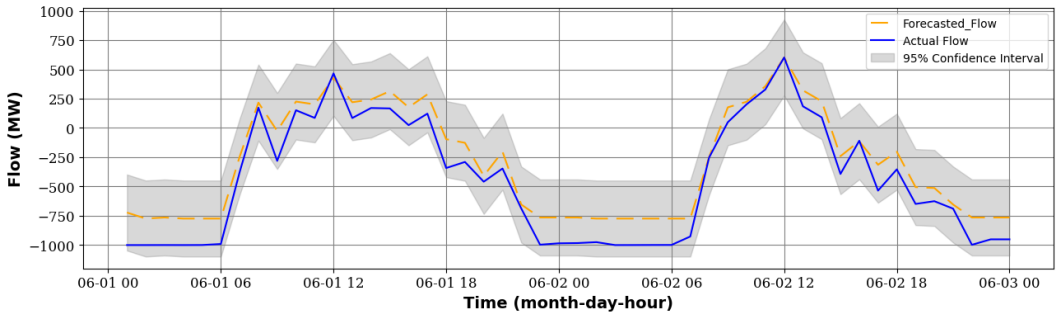

Figure 2: The forecasted target for 48 hours ahead with 95% CI.

This study underscores the importance of energy forecasting in addressing contemporary energy challenges and demonstrates that LightGBM provides a valuable tool for high-dimensional time series forecasting due to its effective feature selection and capability to capture key temporal dependencies. As part of our ongoing research, we aim to develop and compare alternative forecasting models, including Long Short-Term Memory (LSTM) networks and conventional time series models such as Autoregressive Moving Average model with eXogenous inputs (ARMAX). This will allow us to assess the role of nonlinearity, model complexity, prediction accuracy, and interpretability in energy forecasting.

Keywords: Energy Forecasting, Time Series Forecasting, Machine Learning.

* Presenting author

References

[1] Suganthi, L. and Samuel, A.A., 2012. Energy models for demand forecasting—A review. Renewable and sustainable energy reviews, 16(2), pp.1223-1240.

[2] Ke, G. et al, 2017. Lightgbm: A highly efficient gradient boosting decision tree. Advances in neural information processing systems, 30.