MACHINE LEARNING - BASED PRICE FORECASTING FOR GB ANCILARRY SERVICES

A Data-Driven Approach to Forecasting Market Prices for DC, DM, and DR Services

AI driven Solution for Renewable Energy

Project Summary

As Great Britain moves toward a low-carbon grid, maintaining frequency stability is increasingly complex. This project develops machine learning models to forecast day-ahead prices for Dynamic Frequency Response (DFR) services—Dynamic Containment (DC), Dynamic Moderation (DM), and Dynamic Regulation (DR)—using historical market and system data.

Built with XGBoost and time-aware validation, the models integrate demand forecasts, renewable output, and frequency signals to predict price dynamics. High-performing models like DR-H (R^2 = 0.73) and DM-H (R^2 = 0.65) capture real-time system responses, while others rely on historical patterns. This approach enables more strategic BESS bidding and supports future integration into a live dashboard for real-time market insights.

IMPERIAL HACKATHON 2025

Group 16 - April 2025

Scan me!

1. Introduction

The UK power grid sector accelerates its transition toward net-zero emissions, maintaining stability in increasingly decarbonised grid has become a growing challenge. National Energy Sistem Operator (NESO) of the Great Britain plays a critical role in ensuring frequency stability through a suite of ancillary services. Among these, Dynamic Frequency Response (DFR) services including Dynamic Containment (DC)-post fault service designed to arrest significant frequency deviations following large-scale generation losses with frequency range –0.8 Hz to +0.5 Hz, Dynamic Moderation (DM)-pre fault service that provides a stabilising influence on the system by gently nudging frequency back toward 50 Hz during relatively volatile periods or +- 0.2 Hz , and Dynamic Regulation (DR) have been emerged as key mechanisms to rapidly respond to frequency deviations and balance supply and demand in real time[1]. These services are particularly important to maintaining system frequency around 50Hz and mitigating during periods of low system inertia and increasing renewable energy penetration, which pose new challenges to frequency stability.

Since the launch of the Enduring Auction Capability (EAC) platform in November 2023, these services have been procured via a daily co-optimised auction, allowing dynamic bidding and participation across multiple services. In response, battery energy storage system (BESS) operators have adapted by using more flexible strategies, such as participating in the Balancing Mechanism without Final Physical Notifications (FPNs) and engaging in Net Imbalance Volume (NIV) chasing[2].

This project aims to develop a data-driven, day-ahead DFR price forecasting model using historical price, demand, and volume data, while also exploring how BESS behaviours influence market outcomes. The goal is to support grid stability, improve market transparency, and contribute to the broader clean energy transition.

2. Data Description and Pre-Processing

The dataset spans from 2023 to 2024 and consolidates a wide range of electricity market and system-level data essential for day-ahead price forecasting in Great Britain. It includes hourly Day-Ahead Prices from both the N2EX and EPEX exchanges (£/MWh), as well as 30-minute generation data across diverse fuel types—including wind, hydro, biomass, coal, nuclear, and gas-fired power (CCGT and OCGT). Additional features include Maximum Export Limit Below Physical Notification (MEL below PN), system frequency metrics for grid stability, and ancillary services data at four-hour intervals, covering volume forecasts, accepted volumes, and pricing for Dynamic Containment (DC), Dynamic Regulation (DR), and Dynamic Moderation (DM). National demand forecasts and price signals are also included to provide a comprehensive modelling foundation.

To prepare the dataset for modelling, files were concatenated by data category to ensure continuity. Time columns were standardized into Python datetime objects and reformatted into the UK date format (DD/MM/YYYY). Missing values were handled using a combination of forward fill (ffill) and backward fill (bfill) methods. This approach ensures that gaps in the data are filled using the most recent available values or the next available values, preserving temporal consistency while avoiding the introduction of unrealistic or arbitrary assumptions. All datasets were then resampled into 4-hour blocks based on Electricity Forward Agreement (EFA) periods using the mean value, aligning the data with typical market operations and smoothing short-term fluctuations to better capture broader trends relevant to price forecasting. The datasets were merged via an outer join on timestamps to retain all available data points, with missing values post-merge addressed using the same fill strategies to preserve temporal consistency and data integrity.

A comprehensive feature engineering process (***see attachment 1***) was conducted to prepare the dataset for machine learning. Temporal features such as EFA blocks, day of the week, month, and weekend indicators were derived to capture time-based patterns. Market price features include the average and spread between N2EX and EPEX day-ahead prices. For ancillary services like DC-H and DR-L, lagged prices (4-hour and 24-hour), rolling averages (3 and 7 periods), and shortfall indicators (forecast vs accepted volume) were

1. National Grid ESO (2023). Frequency Response Services Market Reports. [Online]. Available: https://www.nationalgrideso.com
2. Ofgem (2023). Ancillary Services and Balancing Mechanisms Overview. [Online]. Available: https://www.ofgem.gov.uk

incorporated. Cross-service features, such as price differentials (DM-H vs. DR-H) and DR-H negative bid flags, were also included, along with Fourier terms to capture annual seasonality.

The renewable share was calculated as the proportion of wind and hydro generation relative to total major sources. System-level indicators include a post-EAC flag (from November 2023), frequency deviation from 50 Hz, low-frequency event flags, and a low-inertia proxy based on thermal generation share. Other features include DC-H to DM-H volume ratios, lagged renewable percentage, peak hour flags, and interaction terms like frequency × low inertia and renewable × market price. Demand-side metrics, including national demand forecasts, their lags, and demand-price interactions, were also added. Missing values from lagged features were forward filled, with any remaining nulls dropped to ensure the dataset was clean and ready for model training.

3. Modelling Approach and Hyperparameter Tuning

DFR Forecasting model simulation used **XGBoost (Extreme Gradient Boosting)** as the core algorithm for regression tasks, integrated within a scikit-learn Pipeline that combines robust preprocessing and model training. XGBoost is an optimized and highly efficient gradient boosting algorithm known for its speed, scalability, and strong predictive performance—particularly in structured data problems. It is well-suited for electricity price forecasting, where the relationships between features and target variables can be highly nonlinear. In this pipeline, XGBRegressor is used with the *'reg:squarederror'* objective function, which minimizes the squared difference between actual and predicted values, ideal for continuous regression tasks. To prepare the data, the pipeline includes **RobustScaler**, which scales features based on the interquartile range, making it resistant to outliers that are often present in real-world electricity market data. This helps stabilize the model's learning process and improves generalization.

For hyperparameter tuning, the notebook applies **BayesSearchCV** from the scikit-optimize library, which leverages Bayesian optimization to efficiently explore the hyperparameter space. This method is more sample-efficient than traditional grid or random search, especially for complex models like XGBoost. The tuning process is guided by a **time-aware cross-validation strategy** using *TimeSeriesSplit* with 3 folds—ensuring that the model is validated in a temporally consistent manner. The performance metric used for optimization is **Negative Mean Squared Error** (n*eg_mean_squared_error*), aligning with the model's squared error objective and ensuring accurate evaluation of forecasting performance. This end-to-end approach ensures the model is both robust and finely tuned for high-precision forecasting in the electricity market.

Table 1. Hyperparameter Tuning Results

Service	learning_rate	max_depth	n_estimators
DC-H	0.07	5	150
DC-L	0.07	5	110
DM-H	0.09	3	160
DM-L	0.09	3	160
DR-H	0.18	6	679
DR-L	0.18	6	679

4. Result and Evaluation

The DM-H and DM-L price forecasting models show distinct behavior based on their top features and importance values. The **DM-H model is driven by market fundamentals**, with *DM-H_shortfall* (38%) as the most important feature, indicating that supply shortages sharply increase prices. *post_EAC* (8%) suggests sensitivity to structural or regulatory events. rolling_7_DM-H_price (6.6%) and lag_1_DM-H_price (6.4%) indicate that temporal patterns and recent price trends continue to play a role, capturing price persistence. Interestingly, DR-L_market_signal (5.8%) emerges as the fifth contributor, suggesting that conditions in the DR-L market—perhaps related to demand-side flexibility—have indirect but meaningful influence on DM-H pricing.

In contrast, the DM-L model prioritizes temporal features. *lag_1_DM-L_price* (23%), *lag_6_DM-L_price* (14%), and *rolling_7_DM-L_price* (10%) dominate, showing strong reliance on past pricing behavior and cyclical trends. Volume Requirements Forecast - DM-L - GB (MW) (9%) highlights the role of anticipated demand, while EFA_block (5%) points to time-of-day effects influencing price behavior. This suggests that while DM-H reacts more to market shocks and shortfalls, **DM-L is shaped primarily by historical pricing and expected demand**.

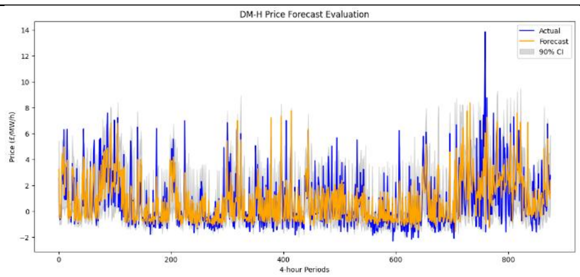
The DR-H model is largely shaped by real-time system-level signals, emphasizing its sensitivity to operational dynamics. Its most influential feature, *Ancillary Volume Accepted - DR-H* - GB (MW) (58%), reflects how accepted balancing volumes significantly drive price volatility. This is closely followed by and *lag_1_DR-H_price (12%)* and *DR-H_shortfall (11%)*, underscoring the model's responsiveness to both supply constraints and immediate past prices. While DR-H also accounts for non-linear demand interactions and renewable generation share, their impact is relatively minor.

The DR-L model leans more on historical price behavior, with a focus on temporal consistency and cyclical trends. Its top *features—rolling_7_DR-L_price* (22%), *lag_1_DR-L_price* (11%), and *lag_6_DR-L_price* (11%)—highlight a pattern-driven forecasting approach. Features such as *Ancillary Volume Accepted - DR-L* - GB (MW) (6%) and *demand_price_interaction* (6%) are still relevant, but their influence is more evenly distributed. This suggests that, unlike DR-H, the DR-L model is less reactive to real-time system events and instead prioritizes trend continuity and recent historical pricing.

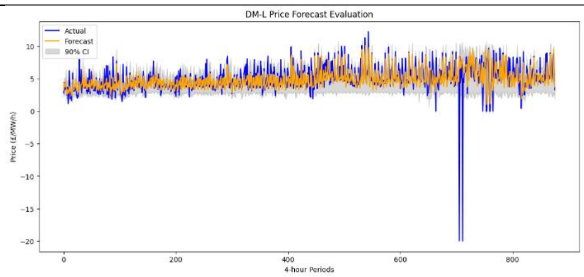
In terms of performance, **DR-H achieves a strong R^2 of 0.74**, effectively capturing price spikes and system-driven volatility. Meanwhile, **DR-L's R^2 is 0.39**, reflecting its strength in stable conditions but limited ability to respond to sudden market changes.

The DC-H model demonstrates stronger predictive performance with an **R^2 of 0.52**, **indicating its ability to moderately capture both short-term trends and dynamic market shifts.** It relies heavily on temporal price features, such as *rolling_7_DC-H_price* (11%), *lag_1_DC-H_price* (10%), and *lag_6_DC-H_price* (10%), supported by structural signals like *EFA_block* (8%) and DC-H_to_DM-H_vol_ratio (6%). This mix shows DC-H's sensitivity to both recent price history and inter-market volume relationships.

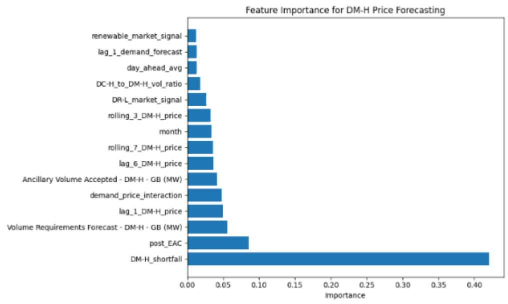
The DC-L model achieves a lower **R^2 of 0.41**, **suggesting it struggles more with capturing price volatility**. However, it emphasizes demand forecasting and renewables, with *Volume Requirements Forecast - DC-L* - GB (MW) (17%) as its most critical feature. Temporal features like *lag_6_DC-L_price* (9%) and *lag_1_DC-L_price* (8%) remain important, but the model also gives weight to *national_demand_forecast* (7%), indicating stronger influence from demand than in DC-H. **DC-H is more trend-focused and structurally balanced**, performing better in moderately volatile conditions, while **DC-L is more demand-sensitive**, but less effective during sudden price shifts.

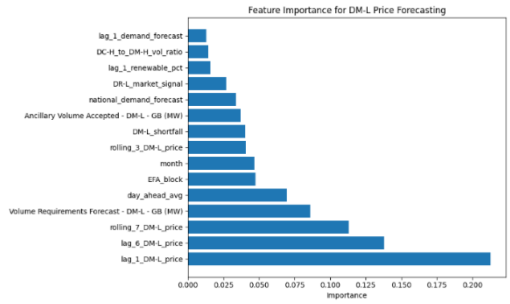


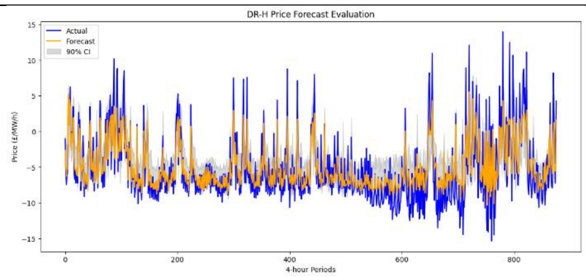
MAE = £0.88/MW/h, RMSE = £1.27/MW/h, R² = 0.65



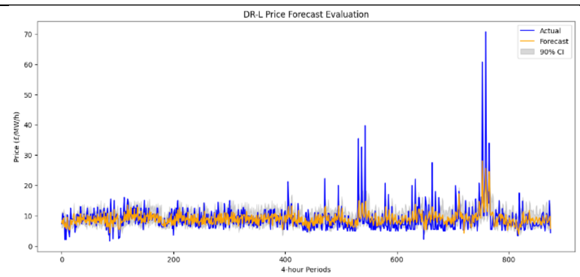
MAE = £0.88/MW/h, RMSE = £1.72/MW/h, R² = 0.36



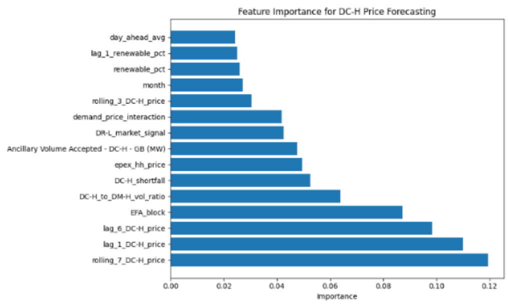


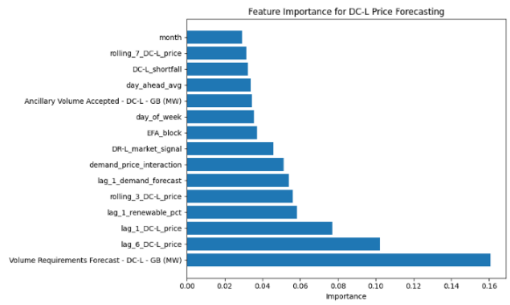


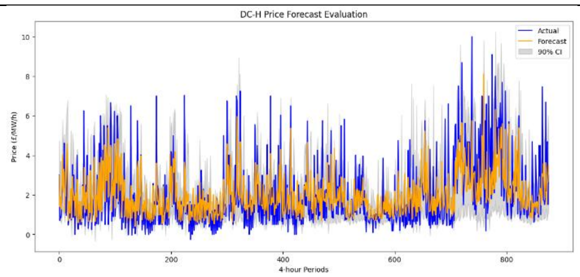
MAE = £1.77/MW/h, RMSE = £2.36/MW/h, R² = 0.74



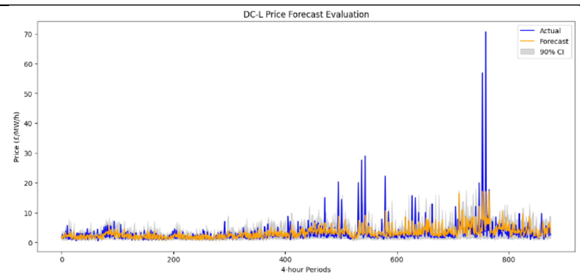
MAE = £2.16/MW/h, RMSE = £3.59/MW/h, R² = 0.39



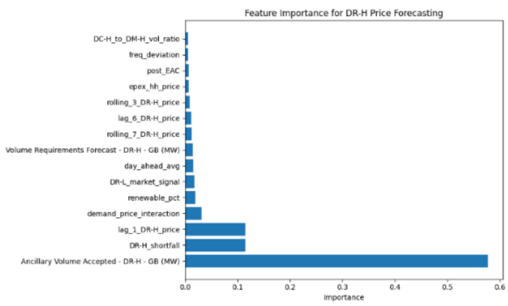




MAE = £0.81/MW/h, RMSE = £1.16/MW/h, R² = 0.52



MAE = £1.22/MW/h, RMSE = £3.04/MW/h, R² = 0.41



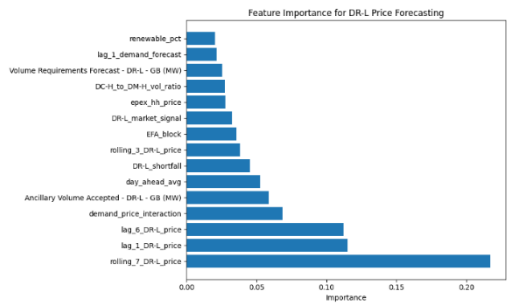


Figure 1. Prediction result, error metrices, and feature importance

In the context of this model, the 90% Confidence Interval (CI) represents the range within which we expect the true ancillary service price to fall 90% of the time, based on the model's predictions. For each forecasted price point, the model generates a lower bound (typically the 5th percentile) and an upper bound (typically the 95th percentile). The interval between these bounds constitutes the 90% CI, indicating there is a 90% probability that the actual price will lie within this range.

It is important to note that this interval is technically a prediction interval rather than a traditional confidence interval. While the terms are often used interchangeably in machine learning, a prediction interval accounts for both model uncertainty and data variability, making it more suitable for estimating future observations. In contrast, a traditional confidence interval reflects uncertainty around the estimated mean and not around individual future outcomes.

Practically, the 90% CI is useful in both risk management and model diagnostics. For example, if the 90% CI for demand response-high (DR-H) prices is [£10, £50], a market participant might use the lower bound of £10 for conservative bidding and the upper bound of £50 to estimate potential worst-case costs. Additionally, if actual observed prices frequently fall outside the predicted confidence interval, this could indicate that the model requires further tuning, such as adjustments to feature engineering or hyperparameter optimization.

5. Conclusion

The performance and behavior of the six price forecasting models (DM-H, DM-L, DR-H, DR-L, DC-H, DC-L) highlight key contrasts in how different markets respond to temporal patterns, system signals, and demand conditions.

- DR-H is the most responsive to real-time system operations, with a high R^2 of 0.74, driven by features like *Ancillary Volume Accepted* and *shortfall*. It excels in capturing price spikes and volatility.
- DM-H also performs well ($R^2 = 0.65$), balancing system-level indicators (e.g., post_EAC, shortfall) with lagged price trends, making it robust in dynamic conditions.
- In contrast, DR-L ($R^2 = 0.39$) and DM-L ($R^2 = 0.36$) depend heavily on historical price behavior, such as lag and rolling price averages. While they perform reliably during stable conditions, they are less effective under market shocks.
- DC-H achieves moderate accuracy ($R^2 = 0.52$) with a strong emphasis on temporal trends (lag and rolling prices) and inter-market volume relationships, making it suitable for structured but changing environments.
- DC-L, with $R^2 = 0.41$, is more influenced by demand forecasting and renewable integration, relying on volume requirements and lagged renewable share. However, it struggles to capture abrupt changes.

Models like DR-H and DM-H that integrate system responsiveness with temporal signals perform best in volatile conditions. DR-L, DM-L, and DC-L, while more stable under calm conditions, lack responsiveness. DC-H sits in the middle, offering a blend of both structural and temporal insights. Together, these models underscore the need to tailor forecasting approaches based on market behavior and volatility level.

6. Future Work

Categories	Actions
Feature Engineering Enhancements	1. Introduce additional real-time system indicators such as frequency response, reserve margins, and localized imbalance prices to better capture sudden market shifts—especially for models like DR-L and DC-L that currently underperform during high volatility.

	2. Real-Time Data: Incorporate live frequency and inertia data from NESO APIs. 3. Bidding Strategy Simulation: Add features reflecting competitor behavior (e.g., historical bid patterns). 4. Weather Integration: Include wind/solar forecasts to refine renewable generation estimates.
Model Enhancements	1. Dataset: Add the dataset to run deep learning model, because the total of this dataset is not that much for deep learning model. 2. Hybrid Models: Combine XGBoost with LSTMs to capture temporal dependencies and nonlinear interactions. 3. Structural Break Analysis: Explicitly model the impact of EAC platform changes (Nov 2023) using intervention analysis. 4. Uncertainty Quantification: Expand probabilistic forecasting to include scenario-based simulations (e.g., Monte Carlo).
Operational Improvements	1. Dynamic Thresholds: Implement adaptive thresholds for low inertia/frequency alerts based on grid conditions. 2. BESS-Specific Features: Integrate battery degradation metrics or state-of-charge dynamics from domain knowledge (BESS Overview.pdf). 3. Market Linkages: Model interactions between DFR and wholesale markets using game theory.
Validation	1. Out-of-Sample Testing: Validate on post-2024 data to assess model generalizability. 2. Economic Impact Analysis: Quantify revenue gains from optimized bidding strategies using forecasted prices.
Sustainability	Negative Price: Calculate the negative price to predict the day after data predicted (D+1).

7. Recommendation

Through this project, we are planning to develop the dashboard for **live price forecasting** by integrating real-time data inputs, enhancing user interactivity, and enabling service-specific predictions and evaluation metrics (**Figure 2**). The goal is to support decision-making in electricity markets by offering a **transparent, adaptive, and responsive tool** for monitoring and anticipating price dynamics.

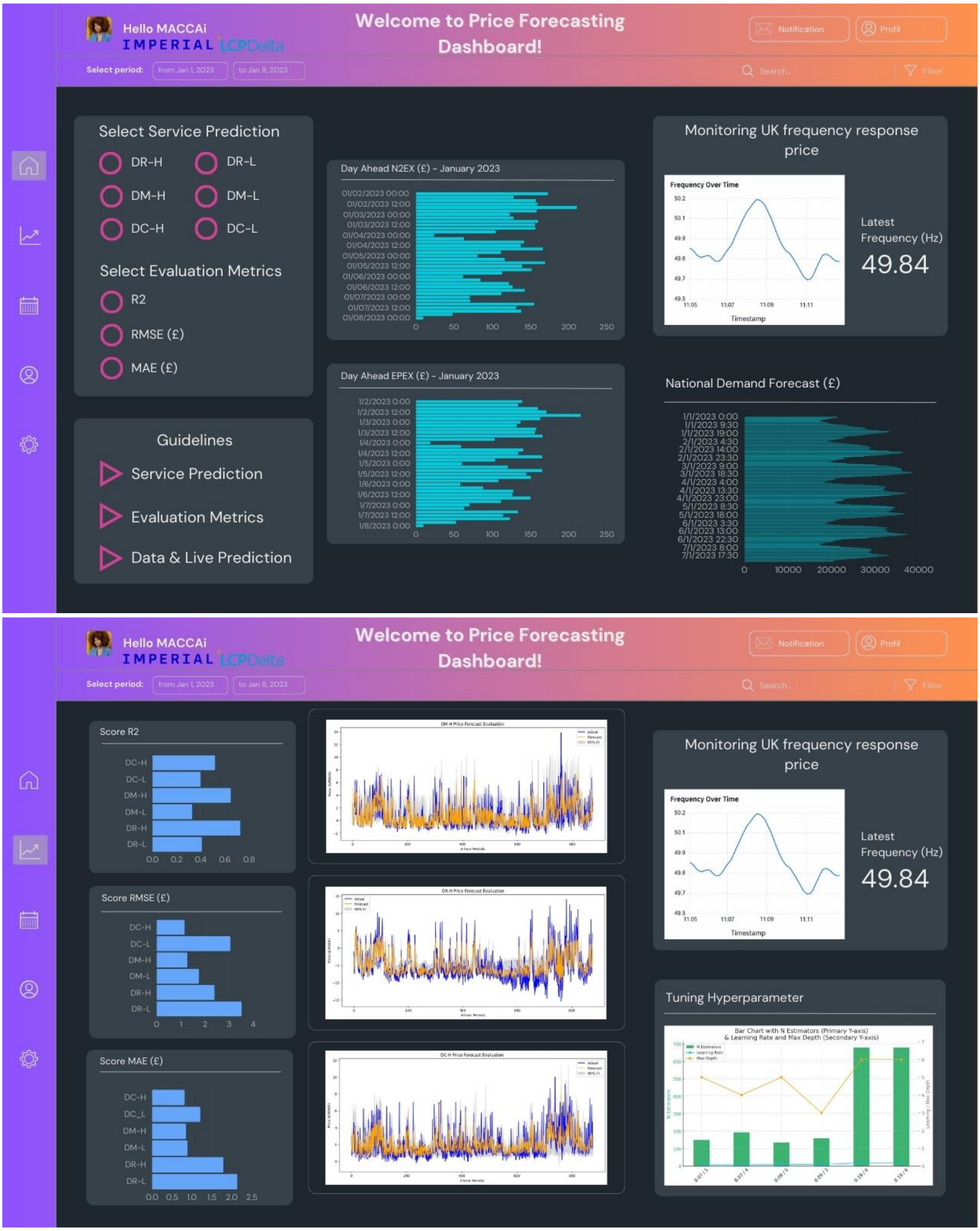
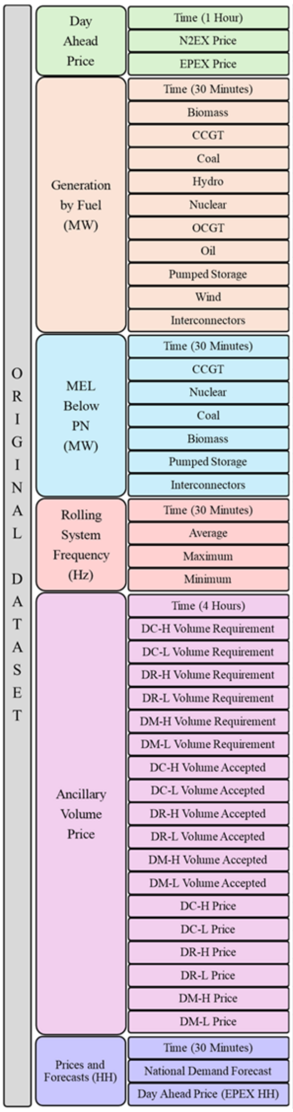


Figure 2. Dashboard plan

Attachment

I. Original Data Usage and Feature Engineering



Feature Created	Feature Source	Engineering Description
EFA_block	timestamp	Derived 4-hour block number from timestamp (EFA convention).
day_of_week	timestamp	Extracted day of the week (0 = Monday, 6 = Sunday).
Month	timestamp	Extracted calendar month from timestamp.
is_weekend	day_of_week	Flag for weekends based on day_of_week.
day_ahead_avg	Day-Ahead Price (N2EX & EPEX)	Mean of N2EX and EPEX day-ahead prices.
day_ahead_spread	Day-Ahead Price (N2EX & EPEX)	Difference between N2EX and EPEX day-ahead prices.
epex_hh_price	Day Ahead Price (EPEX half-hourly)	Granular half-hour EPEX price used for higher resolution.
renewable_pct	Wind, Hydro, and Total Generation	Share of renewable (wind + hydro) in total generation.
lag_1_renewable_pct	renewable_pct	Lag of renewable generation share by 1 period.
low_inertia	Thermal Generation Share	Binary flag when thermal generation share < 30%.
post_EAC	Timestamp	Binary flag for periods on or after Nov 2023.
DC-H_to_DM-H_vol_ratio	Volume Forecasts for DC-H & DM-H	Ratio between forecasted volumes for DC-H and DM-H.
lag_1_DM-H_price_for_DR-H	Ancillary Price - DM-H	Lagged DM-H price (used in DR-H features).
lag_1_DR-H_price_for_DM-H	Ancillary Price - DR-H	Lagged DR-H price (used in DR-H features).
DM-H_minus_DR-H	Ancillary Price - DM-H & DR-H	Lagged spread between DM-H and DR-H prices.
DR-L_market_signal	day_ahead_avg × DR-L volume	Interaction capturing DR-L market exposure.
is_peak_hour	timestamp	Binary flag for peak hours (16:00–20:00).
renewable_market_signal	renewable_pct × day_ahead_avg	Interaction between renewable share and day-ahead prices.
national_demand_forecast	Prices & Forecasts dataset	Forecast of national electricity demand.
lag_1_demand_forecast	national_demand_forecast	1-period lag of national demand forecast.
demand_price_interaction	demand × day_ahead_avg	Interaction term linking demand and pricing behavior.
DC_High_trigger	Rolling System Frequency - HH Max	Binary flag if frequency exceeds 50.5 Hz, triggering DC High.
DC_Low_trigger	Rolling System Frequency – HH Min	Binary flag if frequency drops below 49.2 Hz, triggering DC Low.
DC_High_mag	Rolling System Frequency - HH Max	Magnitude of frequency excursion above 50.5 Hz (DC High).
DC_Low_mag	Rolling System Frequency - HH Min	Magnitude of frequency excursion below 49.2 Hz (DC Low).
DM_High_trigger	Rolling System Frequency - HH Max	Trigger for DM High when frequency > 50.2 Hz.
DM_Low_trigger	Rolling System Frequency - HH Min	Trigger for DM Low when frequency < 49.8 Hz.
DR_High_trigger	Rolling System Frequency - HH Max	Trigger for DR High when frequency > 50.1 Hz.
DR_Low_trigger	Rolling System Frequency - HH Min	Trigger for DR Low when frequency < 49.9 Hz.
FFR_trigger	Rolling System Frequency - HH Min	Flag for Static FFR trigger when frequency falls below 49.7 Hz.
lag_1_{service}_price	Ancillary Price - {service}	Lagged price by 1 period (4 hours).
lag_6_{service}_price	Ancillary Price - {service}	Lagged price by 6 periods (24 hours).
rolling_3_{service}_price	Ancillary Price - {service}	3-period rolling average (12 hours lookback).
rolling_7_{service}_price	Ancillary Price - {service}	7-period rolling average (28 hours lookback).
{service}_shortfall	Volume Forecast vs Accepted	Difference between forecasted and accepted volumes.
Volume Requirements Forecast - {service} - GB (MW)	Forecast dataset	Forecasted demand for ancillary service.
Ancillary Volume Accepted - {service} - GB (MW)	Accepted dispatch volume	Actual accepted volume for ancillary service.