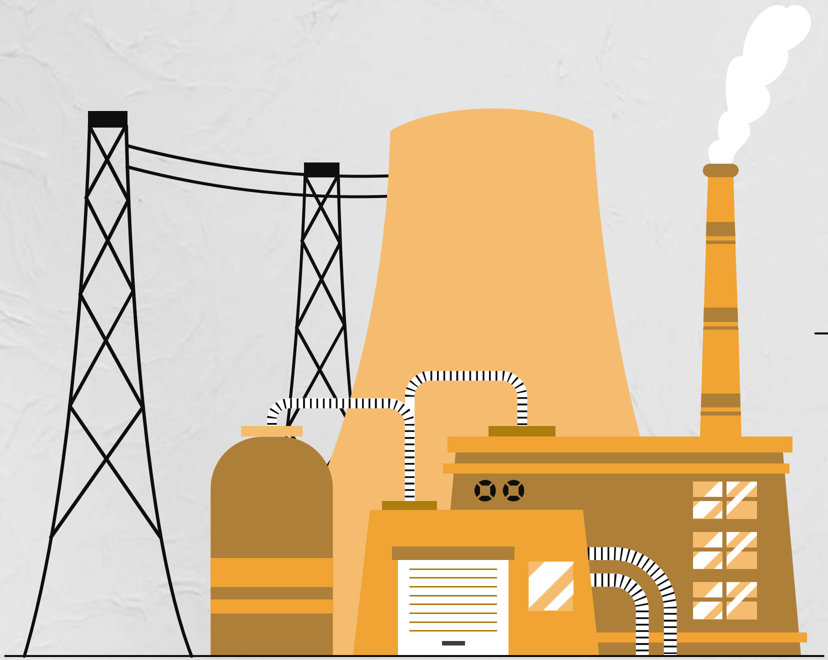


Identifying the key drivers of electricity price spikes in the Australian National Electricity Market

A Machine Learning Approach

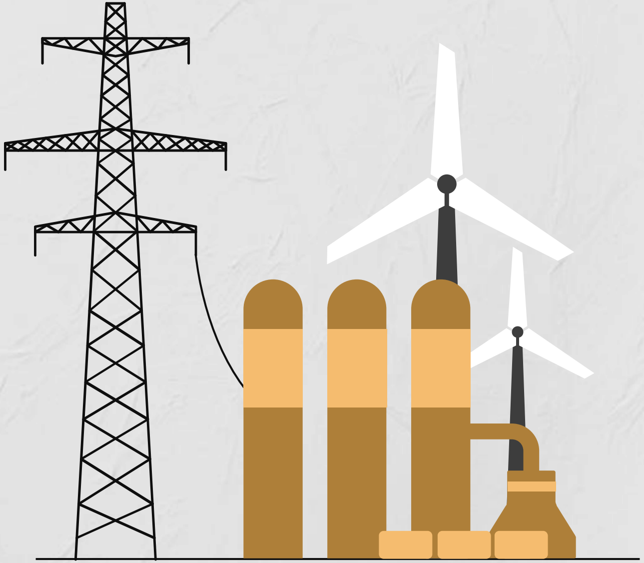
Thi Hau Nguyen

Table of contents

Link on GitHub: *https://github.com/HauNguyen8689/Introduction-to-Machine-Learning-Supervised-Learning*

01

Introduction



Electricity price spikes ?

Context:
With the next federal election set for May 3, 2025, electricity prices are a major issue.

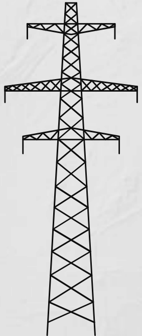
- Coalition blames renewables under Labor's Capacity Investment Scheme (CIS)
- Labor blames fossil gas generation in the National Electricity Market (NEM)

Industry Insight:
Natural gas is commonly seen as the "peaker" — the fuel that sets prices during demand spikes.

Objective:
Use machine learning to unbiasedly investigate drivers of high wholesale prices (above 95th percentile) across NEM regions.

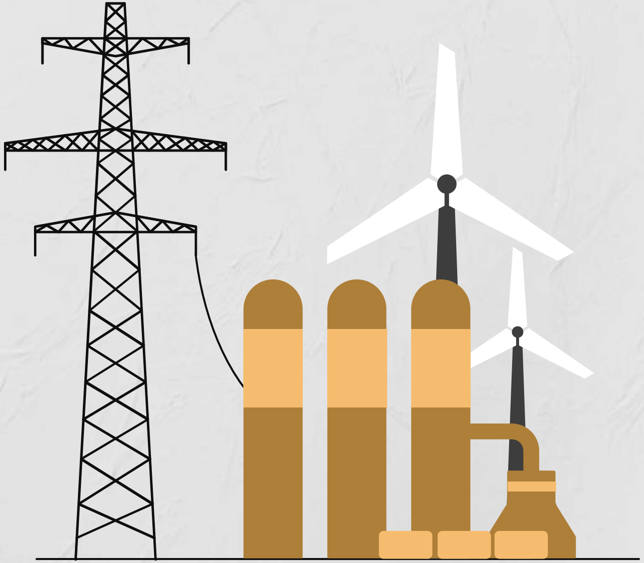
Method:

- Models: Random Forest & XGBoost
- Features: Fuel types, dispatch shares, and time-based patterns
- Outcome: Identify which factors most influence price spikes, regionally and temporally



02

Objective



NEM Overview:

- Australia's NEM spans 5 regions (NSW/ACT, QLD, VIC, SA, TAS), operating a real-time wholesale market managed by AEMO.
- Dispatch & Settlement: Every 5 minutes, matching supply and demand to ensure grid stability and efficiency.

Challenge:

The complexity of the NEM makes it difficult to pinpoint what drives price spikes.

Study Aim:

Classify high-price events (above 95th percentile) and identify conditions linked to these spikes.

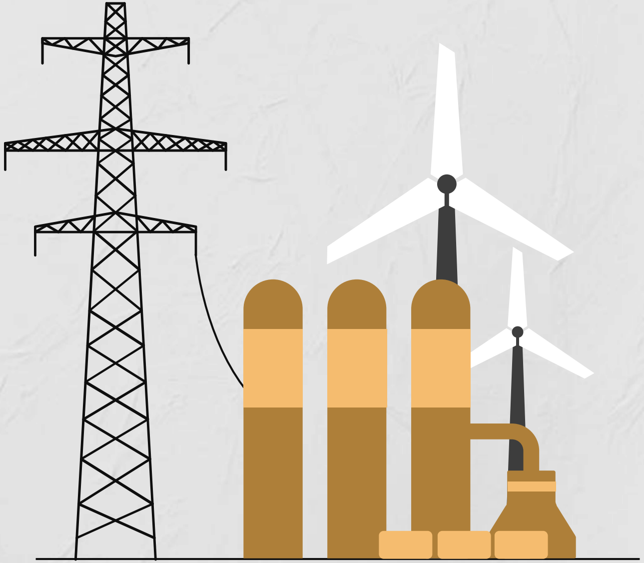
Classification Goal:

Determine which factors act as:

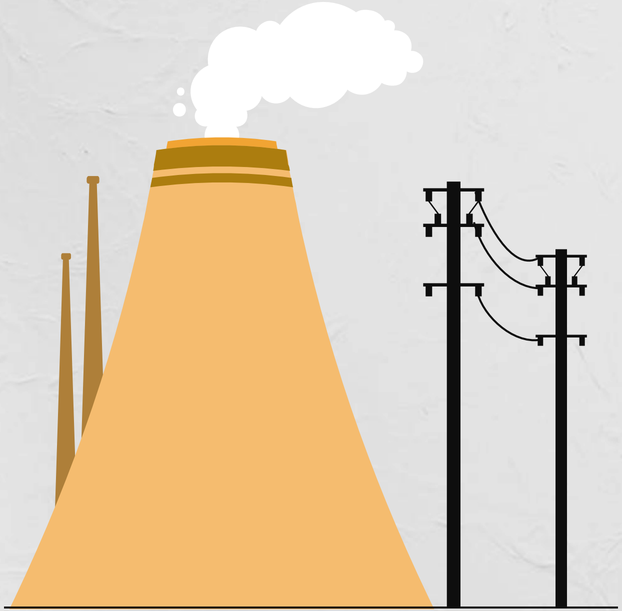
- "Peakers" – linked to price spikes
- "Non-Peakers" – linked to normal pricing

03

Data Overview and Processing



Data Sources



Source: Historical electricity data from AEMO (April 1, 2024 – March 31, 2025)

Contents:

- Regional demand and price
- Generator dispatch & capacity (by DUID)
- Generator metadata (fuel type, region)

Resolution: Originally at 5-minute intervals, aggregated to hourly — sufficient for capturing price spikes

Format: Stored in CSV, uploaded to Kaggle for easy access (*https://www.kaggle.com/datasets/baohoangnguyen/nem-hourly-dispatch*)

Data Processing

The original dataset contained approximately 3.5 million observations. It was then cleaned and aggregated by fuel type. The final dataset used for modeling consists of 276,201 observations

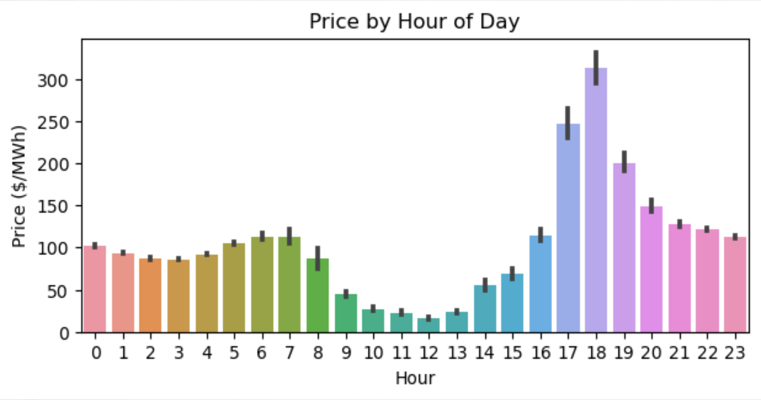
Dependent Variable (Target)

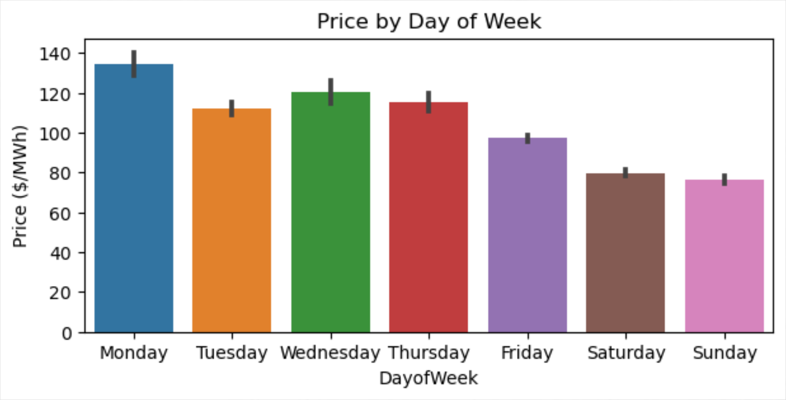
- **Peaker** = 1 for prices ≥ 95th percentile
- **Non_Peaker** = 0 for prices < 95th percentile

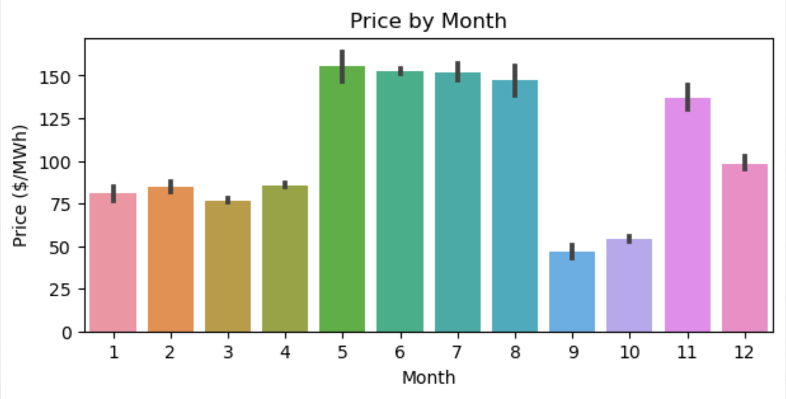
Independent Variables (Features)

- **Hour**: Hour of day (0–23)
- **DayofWeek**: Day of the week
- **Month**: Month of the year
- **Prop**: Proportion of total dispatch by each fuel type
- **FuelType**: Fuel type
- **Region**: One of the 5 NEM regions.

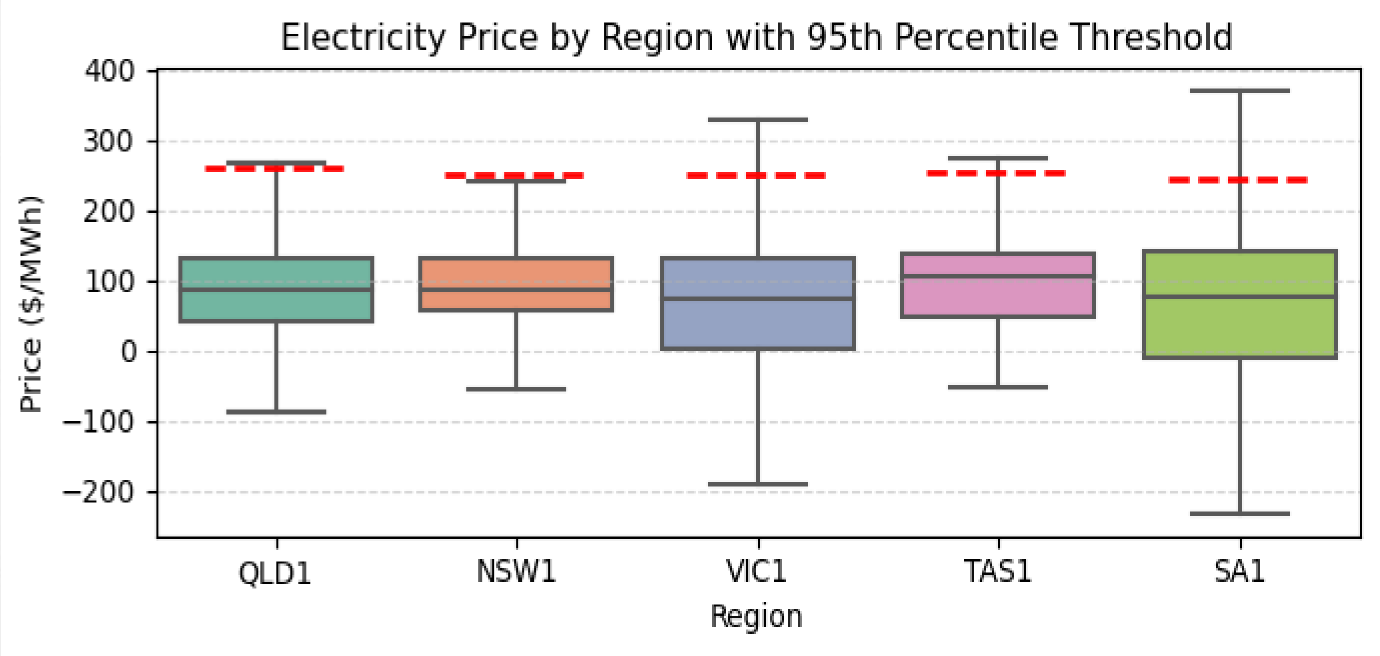
Data Visualisation





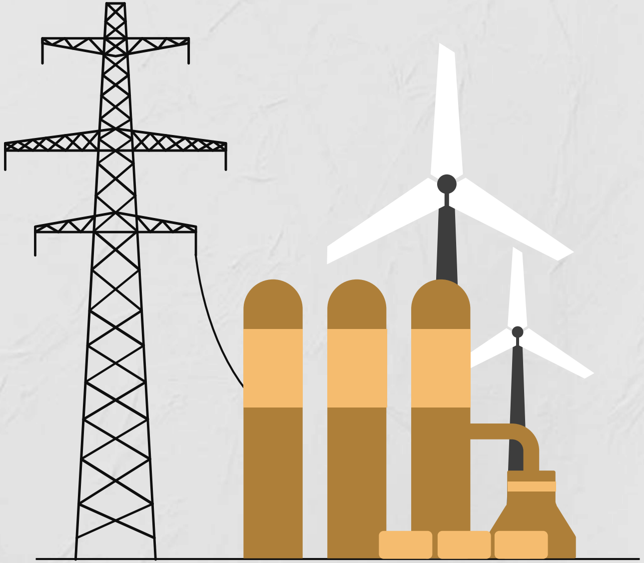


Data Visualisation (cont.)



04

Methodology



Data Splitting

- **Training dataset**: 80% of the dataset used to fit the models and perform hyperparameter tuning.
- **Testing dataset**: 20% of the dataset held out for evaluating the model's generalization ability on unseen data.

Model Selection

Random Forest:
Ensemble method that builds multiple decision trees and outputs the majority vote.

- ✓ Reduces variance
- ✓ Less prone to overfitting than a single tree

XGBoost:
Gradient boosting model that builds trees sequentially, each correcting the previous.

- ✓ High accuracy
- ✓ Fast and scalable

Hyperparameter Tuning

- Tuning the max_depth parameter for both Random Forest and XGBoost
- Test a range of depth values to select optimal depth based on performance on the test set

Model Evaluation

Primary Metric: Accuracy
Chosen for its clarity in classifying high-price events correctly

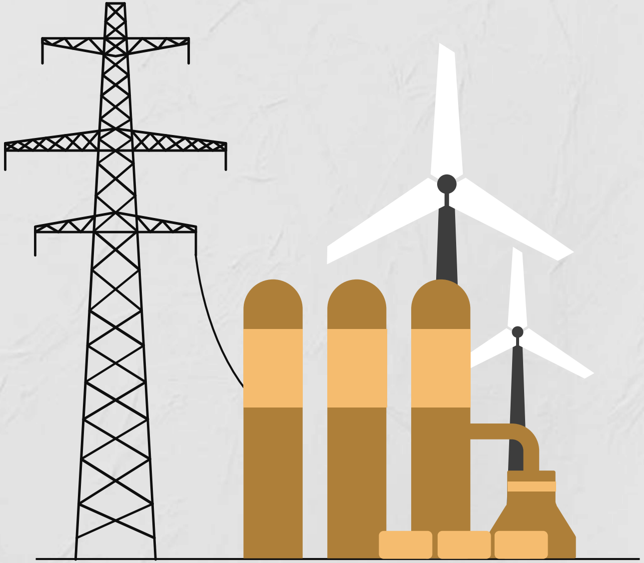
Model Selection:

- Compare performance across different max_depth values
- Best model has the highest test set accuracy

Interpretability:
Analyze feature importance to identify key drivers of price spikes

05

Interpretation and Findings



Results

```
--- NSW1 ---
Price Threshold: 261.80498208333296
Chosen Model: XGBoost
Best Params: {'max_depth': 4}
Train Accuracy: 0.9367
Test Accuracy: 0.9337
                   Feature  Importance
1                     Hour    0.218478
2                    Month    0.167240
6     FuelType_Natural gas    0.113929
5           FuelType_Hydro    0.108677
7           FuelType_Other    0.102210
10           FuelType_Wind    0.077299
3                DayofWeek    0.077114
0                     Prop    0.073869
9           FuelType_Solar    0.035374
8    FuelType_Pumped Hydro    0.025808
----------------------------------------
```

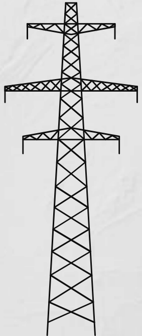
```
--- QLD1 ---
Price Threshold: 250.4841925
Chosen Model: XGBoost
Best Params: {'max_depth': 4}
Train Accuracy: 0.9344
Test Accuracy: 0.9305
                   Feature  Importance
1                     Hour    0.321931
2                    Month    0.195015
7     FuelType_Natural gas    0.104947
3                DayofWeek    0.071646
0                     Prop    0.065895
6     FuelType_Liquid Fuel    0.065540
11           FuelType_Wind    0.050052
9    FuelType_Pumped Hydro    0.049126
5           FuelType_Hydro    0.027044
10          FuelType_Solar    0.027032
----------------------------------------
```

Results (cont.)

```
--- SA1 ---
Price Threshold: 251.025565
Chosen Model: XGBoost
Best Params: {'max_depth': 4}
Train Accuracy: 0.9482
Test Accuracy: 0.9421
                 Feature  Importance
5  FuelType_Natural gas    0.279663
4  FuelType_Liquid Fuel    0.200231
8         FuelType_Wind    0.164102
2                 Month    0.129158
0                  Prop    0.073031
1                  Hour    0.069116
3             DayofWeek    0.047790
7        FuelType_Solar    0.027273
6        FuelType_Other    0.009636
----------------------------------------
```

```
--- TAS1 ---
Price Threshold: 254.68937916666667
Chosen Model: XGBoost
Best Params: {'max_depth': 4}
Train Accuracy: 0.9605
Test Accuracy: 0.9504
                Feature  Importance
2                 Month    0.483058
5         FuelType_Wind    0.174952
1                  Hour    0.105621
0                  Prop    0.105447
3             DayofWeek    0.080423
4  FuelType_Natural gas    0.050499
----------------------------------------
```

```
--- VIC1 ---
Price Threshold: 245.02795
Chosen Model: XGBoost
Best Params: {'max_depth': 4}
Train Accuracy: 0.9453
Test Accuracy: 0.9396
                Feature  Importance
7        FuelType_Other    0.283288
9         FuelType_Wind    0.188445
2                 Month    0.135840
8        FuelType_Solar    0.092249
1                  Hour    0.074687
6  FuelType_Natural gas    0.061218
5        FuelType_Hydro    0.050852
3             DayofWeek    0.045357
4   FuelType_Brown coal    0.039332
0                  Prop    0.028731
----------------------------------------
```



Feature Importance

1

Temporal Factors dominate across regions

2

Fossil Fuels — especially gas — are major price drivers

3

Wind and Solar — more often linked to lower prices

4

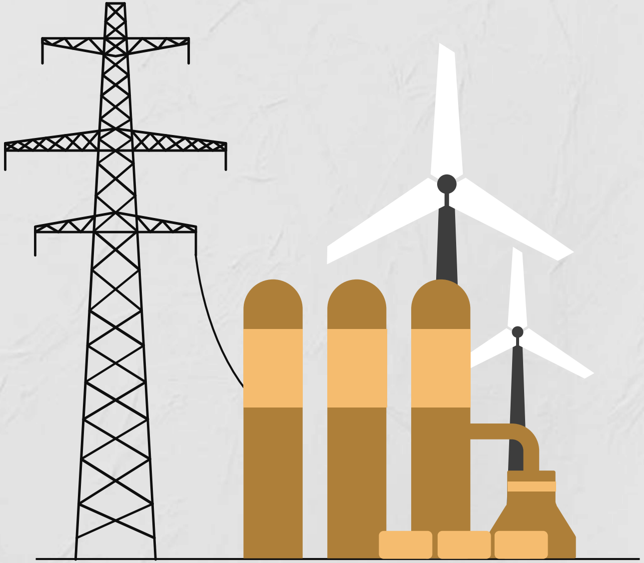
Hydro and Pumped Hydro show mixed behavior

5

Dispatch Proportion is less predictive on its own

06

Conclusion



Approach:

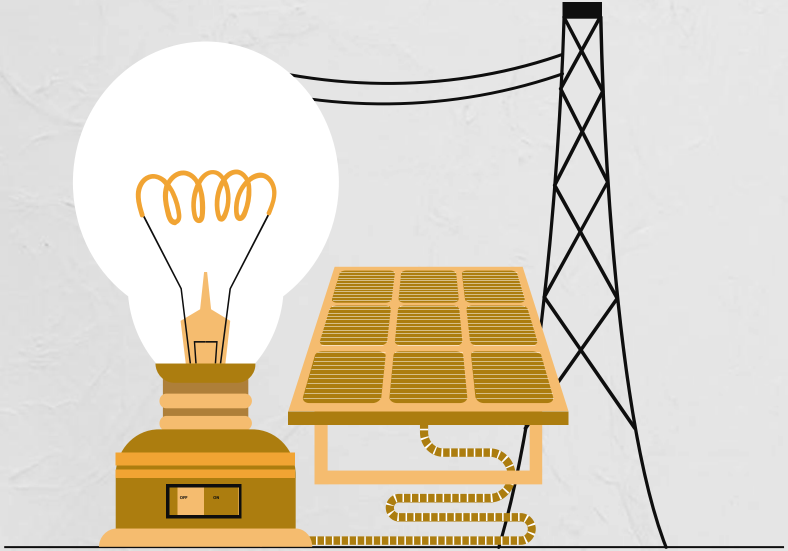
Applied supervised machine learning (Random Forest & XGBoost) to one year of hourly NEM data. XGBoost selected for highest accuracy across regions

Findings:

- Fossil fuel generators, especially natural gas, are the main drivers of high-price events
- Renewables are not the primary cause of price spikes, contrary to some public claims

Implications:

- Supports a data-driven view of electricity market behavior
- Emphasizes the need for peak demand management, grid flexibility, and better integration of dispatchable resources in Australia’s energy transition



Thanks!

Thi Hau Nguyen