Executive Report

Industrial heat production remains one of the largest contributors to carbon emissions in manufacturing, driven mainly by fuel-based boilers, furnaces, and heaters. Many facilities continue to rely on coal, diesel, and other fuels that not only raise CO_2 output but also increase long-term operational costs. The goal of this project was to identify which facilities would benefit the most from replacing fuel-based systems with electrified alternatives.

Our analysis used clustering techniques and a decision tree classifier to determine which facilities present both high carbon emissions and high heat demand. The clustering model identified unique facility groups based on heat demand and fuel ratios. Meanwhile, the decision tree achieved strong predictive performance in classifying "Highly Recommended" facilities for electrification.

The final recommendation from our prediction reveals that 5,040 facilities are strong candidates for electrification, representing the greatest potential decrement in carbon emissions and long-term cost optimisation. Implementing these recommendations would generate measurable sustainability improvements while supporting industrial decarbonization.

o Business/Engineering Challenge

Industrial manufacturing requires huge amounts of process heat, typically supplied by equipment such as boilers, furnaces, and heaters. A heat proportion is generated using fossil fuels, resulting in high CO_2 emissions and fuel-related costs.

This creates two major engineering and business challenges:

1. High Carbon Emissions

Coal-based and diesel-based units produce unnecessarily high CO_2 relative to the heat they generate. Certain unit types have the emissions profile of industrial operations.

2. Rising and Unstable Operating Costs

Fuel prices vary significantly, making operations dependent on fossil fuels vulnerable to unpredictable cost spikes. On the other hand, electrification provides:

- Higher thermal efficiency
- Lower long-term energy cost fluctuation
- Reduction in fuel-handling infrastructure requirements

However, electrifying all facilities at once is unrealistic due to infrastructure demands and upfront investment requirements. Thus, the main challenge is prioritising which facilities should be electrified first.

Key Target Metrics

To solve this challenge, the project focused on three indicators:

- Annual CO_2 emissions (tons)
- Total annual heat demand (kWh)
- Annual cost of individuals ($)
- Estimated heat production under fossil fuel vs electricity

These metrics define which facility has a high environmental impact, high heat intensity, or high expense.

The business challenge is not just to understand the current emissions profile, but to use data-driven methods to recommend where electrification yields the highest environmental and economic return.

- **Solution Overview (The Model)**

The data team did an analytical pipeline combining engineering knowledge with machine learning techniques.

1. Data Cleaning & Integration

- Removed duplicates and invalid records
- Standardised fuel types, unit types, and facility identifiers
- Collect heat demand, CO_2 emission values, and cost values for individual fuel types
- Managed the datasets for different data scientists

This produced a consistent dataset suitable for both methods of modelling.

2. Exploratory Data Analysis

- Identified top 3 CO_2-emitting fuels (coal, diesel, LPG)
- Found that **boilers** were the single largest contributors to industrial emissions
- Observed that heat demand strongly correlates with CO_2 emissions
- Give insights for initial dashboards from heat, cost, and emissions characteristics

3. Clustering - K-Means as Model Winner

K-means was used to cluster facilities based on: Heat demand and fuel ratio in the facilities.

This produced groups such as:

- Natural Gas Dominant
- Other Fuel Heavy Coal dominant
- Mixed Gas-Other
- Mixed Diesel-Gas
- LPG_NGL Dominant

These clusters reveal which facilities are structurally predisposed to benefit from electrification.

4. Predictive Modelling — Decision Tree as Model Winner

The Decision Tree was selected as the final predictive model because it

- Provides clear, rule-based classification depending on the decision
- Reflects engineering relationships (e.g., coal → high CO_2)
- Easy to explain to management and technical teams,
- Consistently predicted electrification priority.

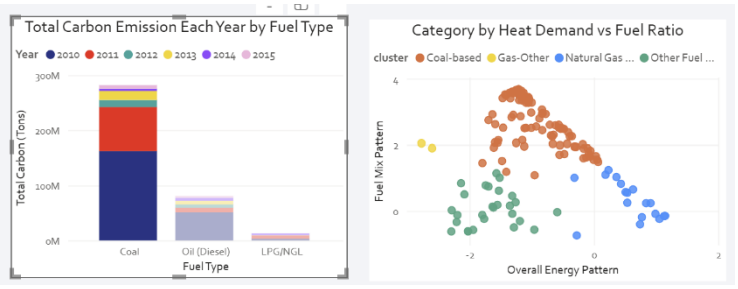
This model has a high accuracy of 97%, a high recall of 98%, and a high precision 99%. These factors are essential because:
1. Indicating that the facilities do not need to be electrified as it is recommended, will keep the high number of carbon production.
2. Indicating that the facilities need to be electrified when they are not will lead to wasteful investment.

The model outputs three classes:

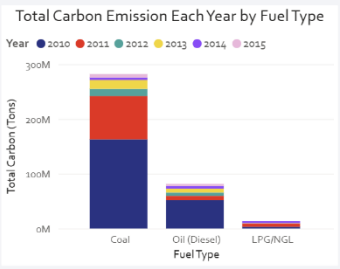
- Highly Recommended (True Values from certain prediction)
- Moderate (Unsure facilities from false prediction)
- Not Necessary (False Values from certain prediction)

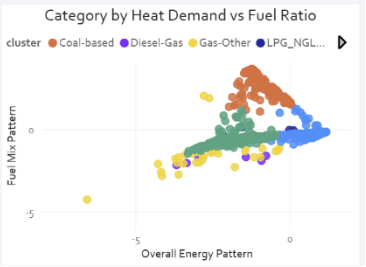
This becomes the foundation for the final electrification strategy.

- **Key Findings and Insights (CLO3)**



1. Coal is the dominant source of industrial CO_2 emissions and heat.
Coal produces the majority of high-intensity heat and carbon output. Facilities using coal-based boilers generated the largest emission volumes across the dataset.

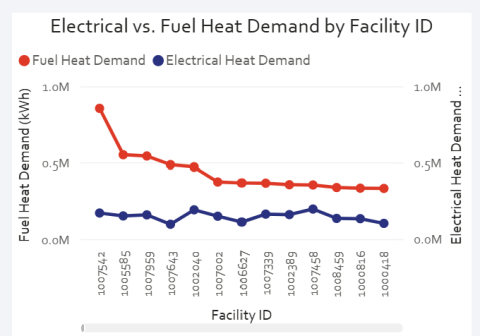




2. The "coal-based" cluster is the strongest electrification target.

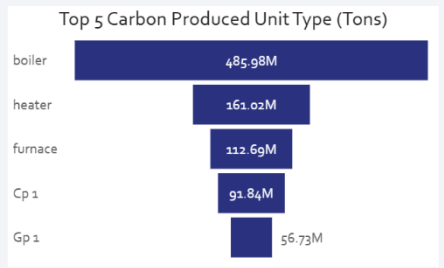
This cluster points out the highest demand, heavy fossil-fuel dependence, and significantly high CO_2 emissions

Facilities in this cluster represent the most immediate potential for carbon reduction.



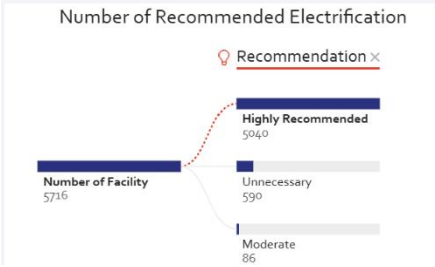
3. Cost comparison shows electricity is favourable in high heating demand facilities.

Over time, electric heat becomes more cost-stable, less dependent on fuel price, and more energy efficient.



4. Boilers and thermal heaters drive most industrial emissions.

Boilers contributed 485 million tons of CO_2, exceeding other equipment. They should be prioritised for electrification.



5. Model Recommendation Summary

The Decision Tree identified highly recommended for electrification: 5040 facilities, moderate: 590 facilities, and not necessary: 86 facilities

- **Recommendations & Implementation Plan**

1. Electrify facilities in the coal-based cluster first.

These facilities have an impact on carbon reduction, energy savings, and long-term cost efficiency. However, changing all at once is not practical. The chosen facilities must be considered with the factors regarding cost.

2. Prioritise coal-based boilers and diesel heaters.

These have the highest emissions per unit of heat and present the largest opportunity for improvement. There are alternatives that can potentially generate as much energy as fossil fuel-

based such as electric boilers, electric furnaces, industrial heat pumps, etc. Electrification of high-priority facilities can reduce CO_2 emissions by 40–70%, depending on fuel type.

3. How to handle the old equipment.

The boiler must be replaced with the new electric boilers, leading to the elimination of coal-based boiler usage. However, sometimes factories keep the old boiler but remove the coal-feeding system and keep it available for emergency use. It works only if the old boiler is in good condition and there is still a way to fuel it, e.g. LPG.

Coal systems usually cannot be kept because coal-handling equipment is removed.

4. Prepare electrical infrastructure upgrades

Electrification may require transformer upgrades, additional grid capacity from power plants, and power quality stabilisation.

5. Require Source

- Required Resources fall into 4 engineering categories.

A. Equipment Cost: Electric boilers, heat pumps, control panels, and transformers.

B. Installation & Infrastructure includes electrical wiring, power distribution upgrade, and decommission of the coal boiler.

C. Engineering Labour: mechanical modifications.

D. Maintenance: Electric boilers have lower annual maintenance, no fuel handling and fewer moving parts.

Electrification provides financial benefits from reduced fuel costs, better energy efficiency, and lower maintenance requirements. Coal and diesel boilers operate at 70–80% efficiency, while electric boilers typically work at 95–99% efficiency, leading to a 20–30% reduction in energy losses. This efficiency gain decreases annual operating costs.

ROI is an evaluation using the difference between fossil-fuel costs and electricity-based heat costs, combined with maintenance savings and CO_2 reduction benefits.

$$ROI = \frac{\text{Annual Cost Savings}}{\text{Total Electrification Investment}}$$

Annual cost savings include:

- Fuel savings from replacing coal/diesel with high-efficiency electric systems
- Maintenance savings, typically reduced by 30%, due to the simplicity of electric boilers
- CO_2 reduction benefits lower their overall tax burden.