Anomaly Detection in Cyclone Preheater Data

Aditya Raj Pandey

March 23, 2025

1 Introduction

I analyzed a cyclone preheater dataset containing 6 variables recorded every 5 minutes over 3 years (377,719 records). My goal was to identify periods of abnormal operation using anomaly detection techniques.

2 Data Preprocessing

I examined the data structure and found approximately 1,300-1,600 non-numeric values in each column that needed addressing. I converted the time strings to datetime format and handled non-numeric values by converting them to NaN and replacing them with column medians to minimize outlier impact.

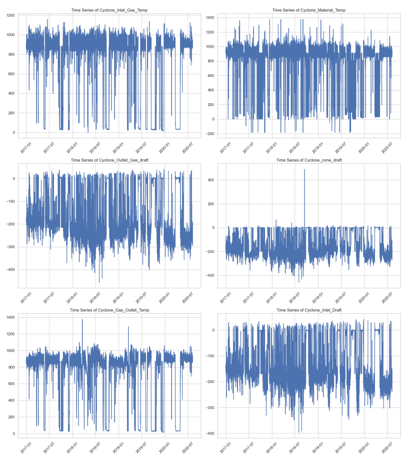

Figure 1: Time series plots of all variables in the cyclone preheater dataset

I analyzed variable distributions and examined correlations to understand data characteristics. The visualization of distributions and correlations revealed several important patterns:
Box plots confirmed the presence of outliers in most variables, which could represent anomalies of interest or data errors.

3 Feature Scaling

I applied Min-Max scaling to normalize all features to a [0, 1] range. This step was crucial because:

- DBSCAN is sensitive to feature scales
- Variables had significantly different ranges
- Min-Max scaling preserves data relationships while standardizing ranges

4 Anomaly Detection with DBSCAN

I chose DBSCAN for anomaly detection because:

- It makes no assumptions about cluster shape
- It automatically determines the number of clusters based on data density

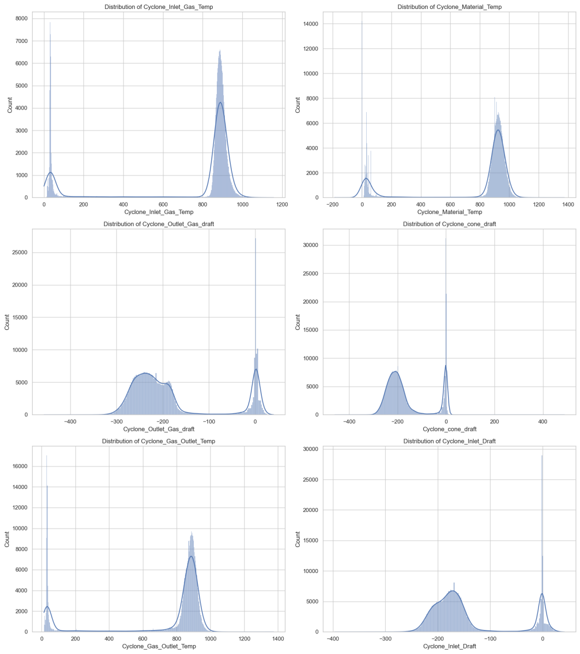

Figure 2: Distributions of all features

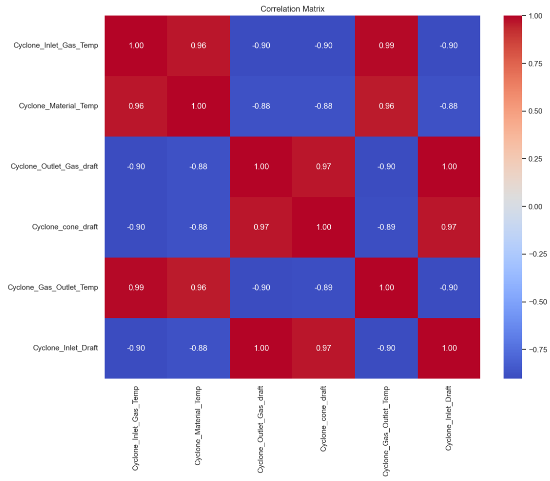

Figure 3: Correlation matrix of variables

- It naturally identifies points in low-density regions as anomalies
- It's robust to outliers

DBSCAN's effectiveness depends on two parameters: `eps` (maximum distance between neighbors) and `min_samples` (minimum points to form a dense region).

4.1 Parameter Optimization

Due to the dataset size, I divided it into batches of approximately 10,000 points each. I used the k-distance method to find optimal `eps` values:

I tested multiple `min_samples` values with different `eps` values and evaluated them using silhouette scores:

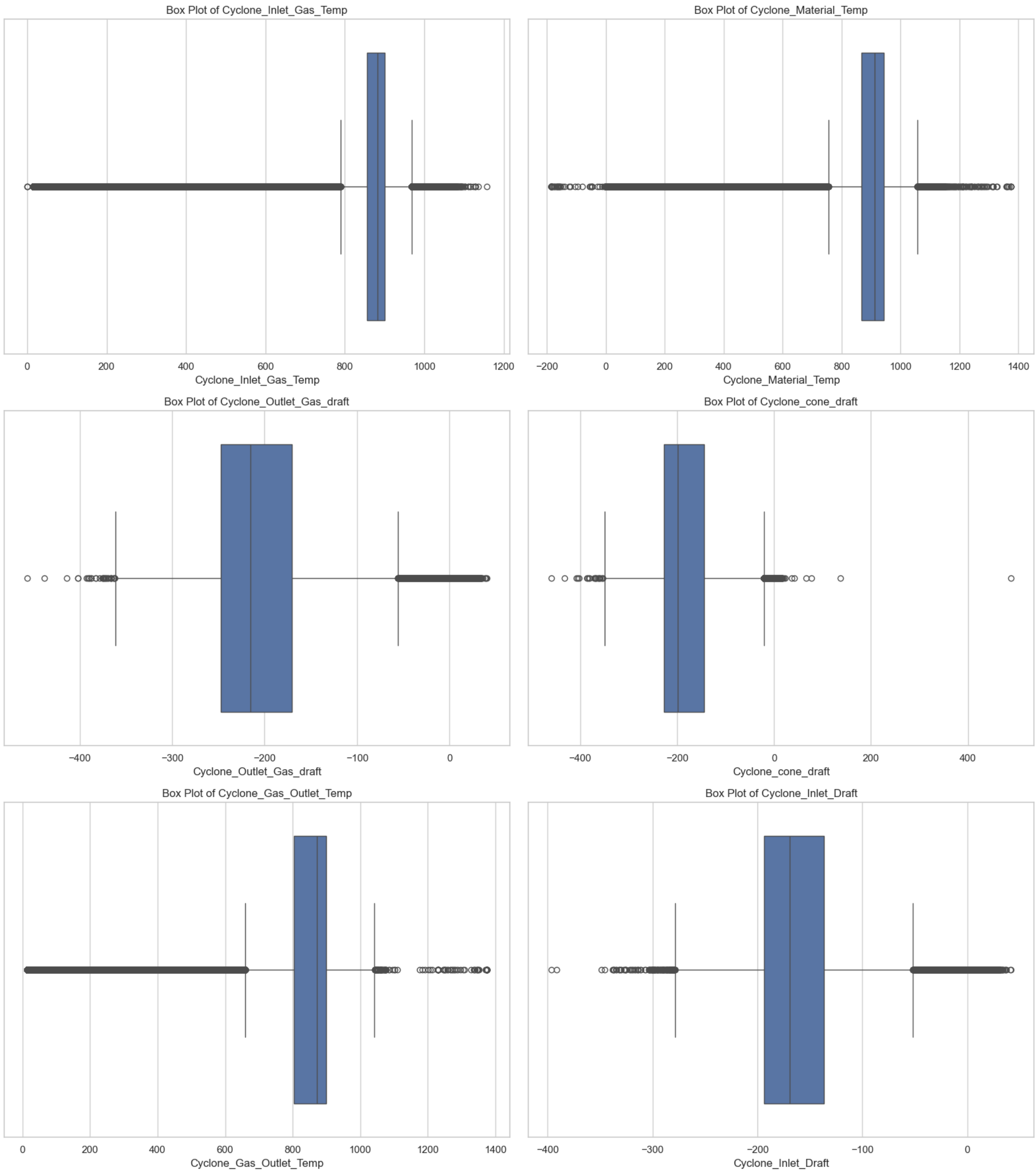

Figure 4: Box plots showing outliers for each variable

5 Results and Anomaly Analysis

After applying DBSCAN with optimized parameters, I identified anomalies as points labeled as noise (cluster -1). These anomalies were distributed across multiple time periods of varying durations.

Individual anomaly plots for each variable revealed distinct patterns. Temperature variables (Cyclone_Inlet_Gas_Temp, Cyclone_Material_Temp, and Cyclone_Gas_Outlet_Temp) showed similar anomaly patterns, while pressure variables (drafts) displayed their own distinct patterns.

Key observations from the anomaly analysis:

- Temperature anomalies often coincided with sharp drops in values
- Draft anomalies frequently appeared as spikes above or below the normal operating range
- Some anomalies occurred simultaneously across multiple variables, suggesting system-wide events
- Other anomalies were isolated to specific variables, indicating potential sensor or subsystem issues

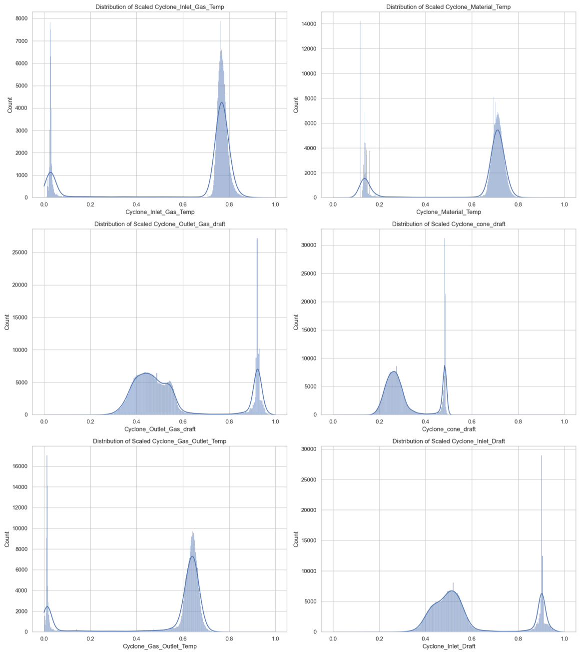

Figure 5: Distributions after Min-Max scaling

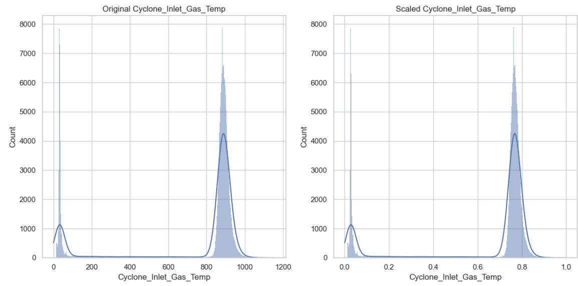

Figure 6: Original vs. scaled data

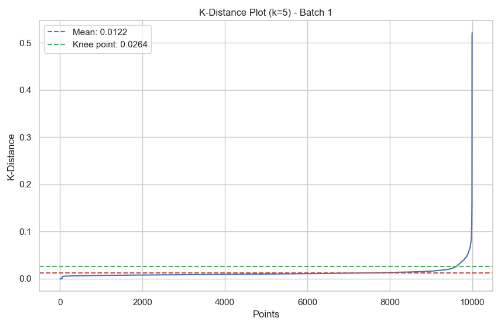

Figure 7: K-distance plot (first batch)

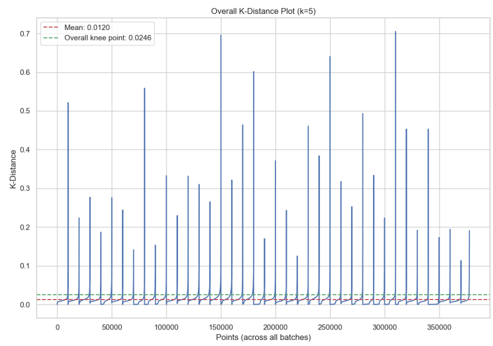

Figure 8: Overall K-distance plot

Figure 9: Parameter testing using silhouette scores

6 Conclusion

I successfully identified anomalous time periods in the cyclone preheater dataset by:

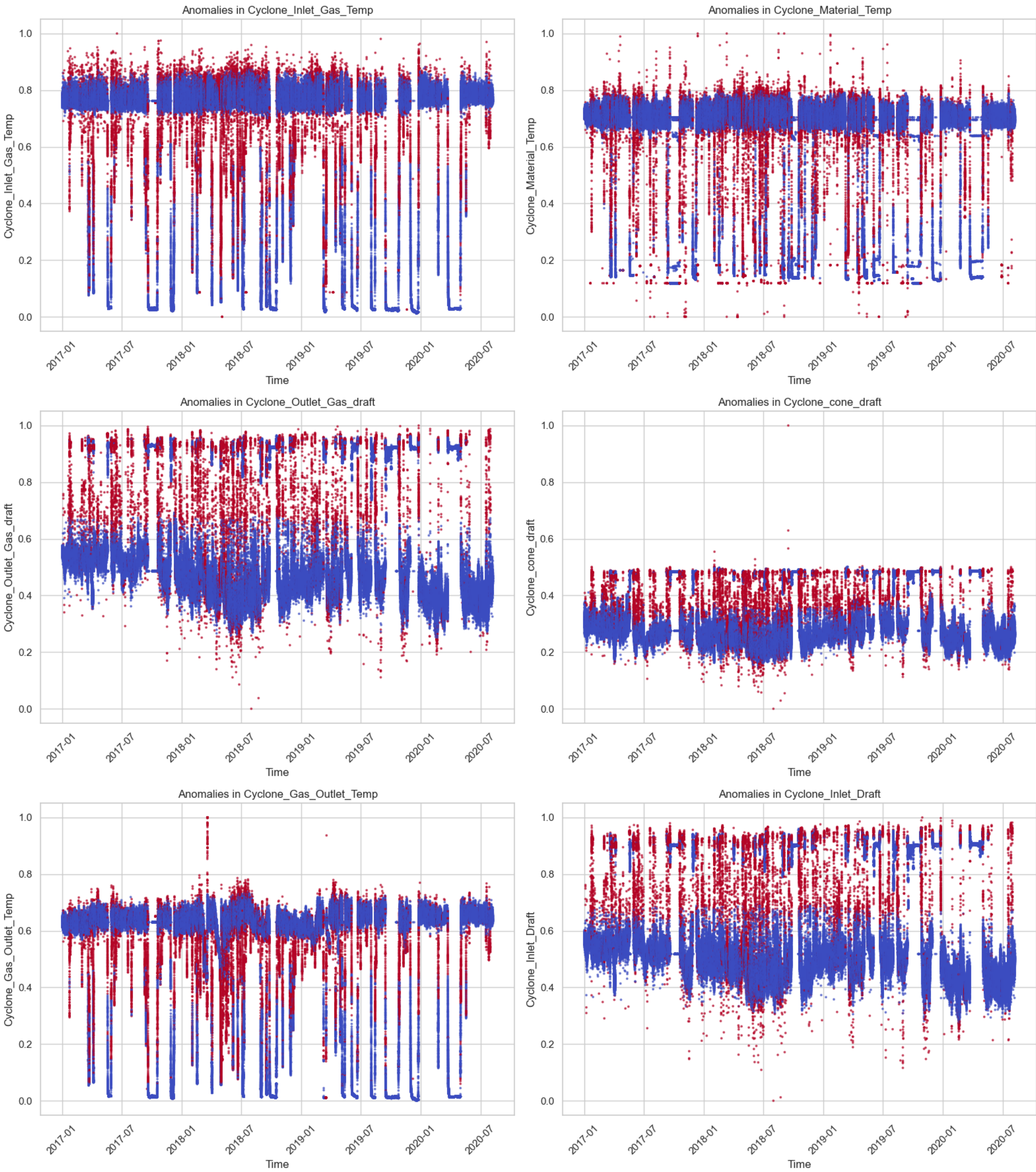

Figure 10: Detected anomalies (red) for all variables

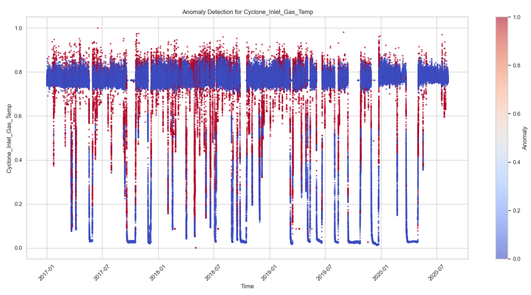

Figure 11: Inlet Gas Temperature

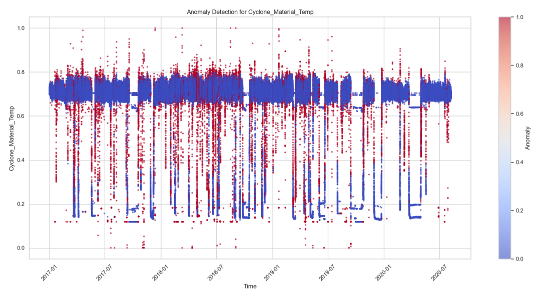

Figure 12: Material Temperature

- Preprocessing the data and handling non-numeric values
- Applying Min-Max scaling for feature normalization
- Using DBSCAN with optimized parameters for anomaly detection

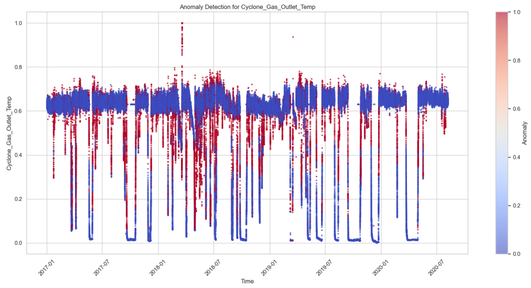

Figure 13: Gas Outlet Temperature

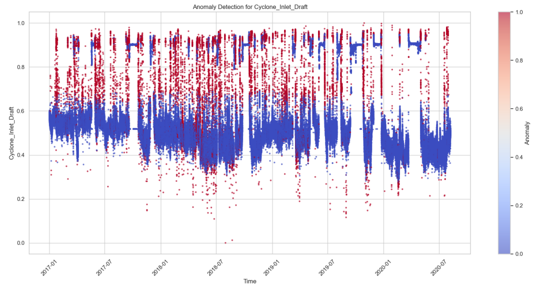

Figure 14: Inlet Draft

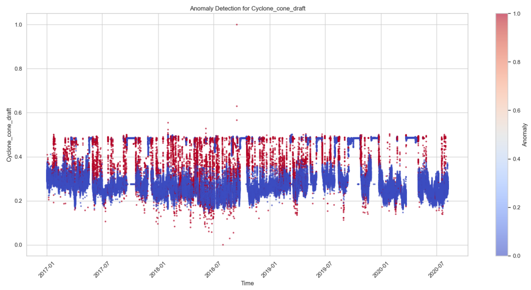

Figure 15: Cone Draft

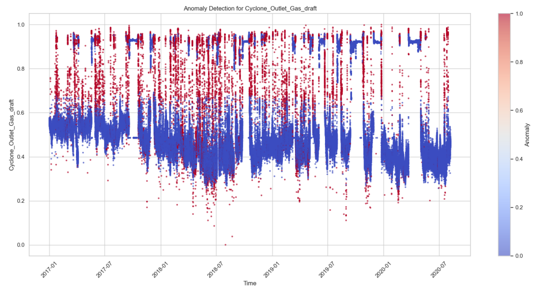

Figure 16: Outlet Gas Draft

- Visualizing abnormal operation periods

The detected anomalies showed temperature and pressure deviations from normal operating conditions, with some periods showing breakdowns in the usual variable correlations. These results provide valuable insights for further investigation of system malfunctions or unusual operating conditions.

Future work could include root cause analysis of the identified anomalies, comparison with other anomaly detection methods, and development of monitoring thresholds for early detection of abnormal operations.