Anomaly Detection in Manufacturing Process Data

Summary of steps

1. Data Preprocessing
   - Loaded the dataset
   - Dropped duplicate records(there were no duplicate records found)
   - Handled missing values using mean imputation
   - Normalized numerical features using standardscaler

2. Model Training
   - Trained on Isolation Forest model on normalized data
   - Predicted anomalies (assigned -1 for anomalies,1 for normal data)
   - Added new column anomaly_pred and anomaly_score

3. Model evaluation
   - Counted how many points were labeled as anomalies vs normal
   - Due to lack of true labels, taken anomaly percentage

4. Visualization
   - Scatter plot for feature pairs colored by anomaly status
   - Boxplots to compare features

Findings

- Identified process anomalies that might indicate potential failures .
- Model identified 150 anomalies out of 3000 which is of 5%

Potential Causes

- Abnormal readings in sensors (temperature,pressure,vibration)
- Errors in logging
- Mechanical faults or overuse or overheat

Recommended Actions

- Having preventive maintenance for flagged equipment
- Improving monitoring systems like sensors
- Cross checking in error logging

Approach

- Approach involved detecting anomalies in manufacturing process data using Isolation forest algorithm.
- Dataset was preprocessed to handle missing values and normalize features. An Isolation model is trained to identify anomalous data points.
- Several visualizations were used to interpret the results