Project: Industrial Anomaly Detection

1. Introduction:

Industrial operations heavily rely on machinery and equipment to maintain productivity and meet demand. However, the failure of industrial equipment can lead to significant disruptions, downtime, and financial losses. Detecting anomalies in industrial data can help in predicting and preventing equipment failures, thereby minimizing downtime and optimizing operational efficiency. This report presents the development and evaluation of an anomaly detection system tailored for industrial equipment.

2. Project Overview:

The primary objective of this project is to develop a robust anomaly detection system capable of identifying unusual behaviour in industrial equipment data. The system aims to pre-emptively detect deviations from normal behaviour patterns, enabling proactive maintenance and mitigating the risk of equipment failures. The project involves several key tasks:

a. Data Generation/Preprocessing: Real-world sensor data is collected from industrial equipment and pre-processed to handle missing values, noise reduction, and scaling. Synthetic data generation also be employed to simulate equipment behaviour if real-world data is limited.

b. Anomaly Detection Model: Anomaly detection models, such as Isolation Forest, are developed to identify anomalies in the pre-processed data. These models are trained to distinguish between normal and anomalous behaviour based on features extracted from sensor data.

c. Model Evaluation: The performance of the anomaly detection model is evaluated using various evaluation metrics such as accuracy, precision, recall, F1-score, and receiver operating characteristic (ROC) curves. These metrics assess how well the model identifies anomalies and differentiate between true positives and false positives.

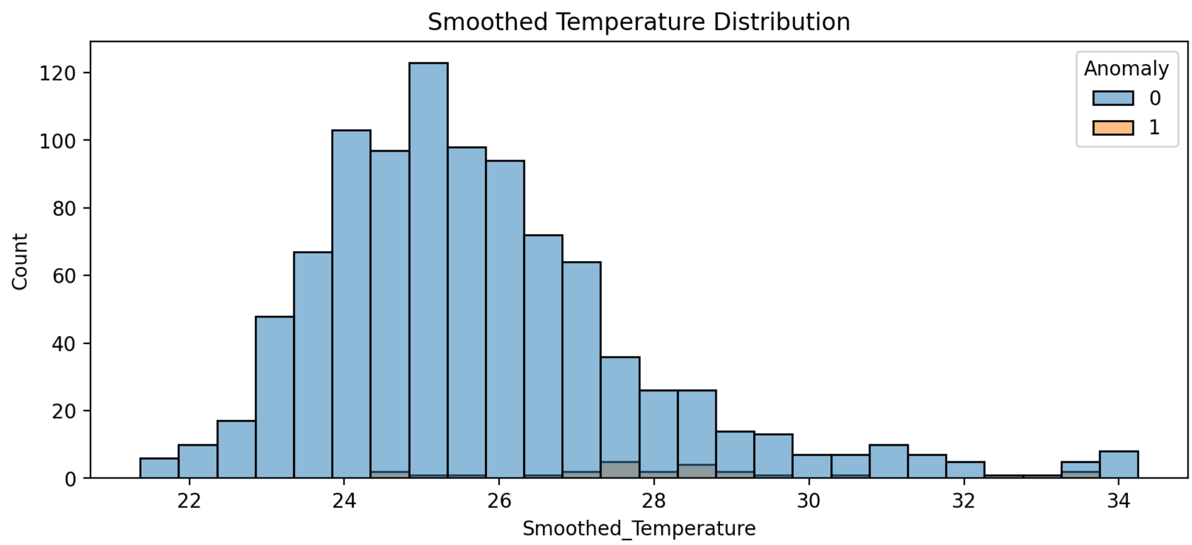
d. Visualization & Reporting: Results from the anomaly detection system are visualized using plots and charts to provide stakeholders with insights into unusual behaviour patterns. Clear visualization and reporting facilitate informed decision-making and timely intervention to prevent equipment failures.

3. Data Preprocessing:

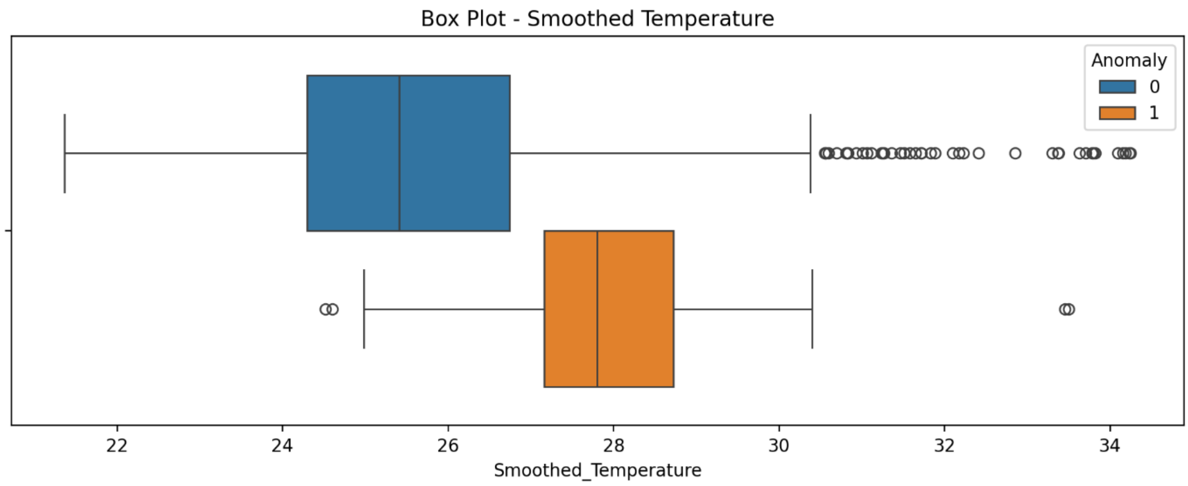
Preprocessing of sensor data is a critical step to ensure the quality and reliability of input data for the anomaly detection model. The following preprocessing steps are applied:

- Handling Missing Values: Missing values in the sensor data are addressed through deletion, depending on the extent of missingness and the impact on model performance.

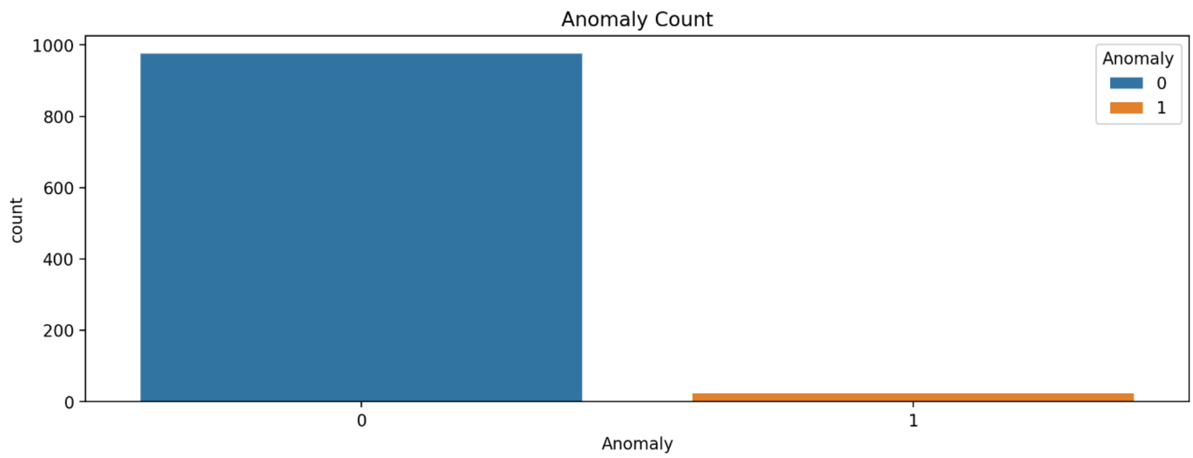
- Noise Reduction: Noise in the sensor data, which can obscure meaningful patterns, is reduced using techniques such as moving average smoothing. Smoothing helps in identifying underlying trends and patterns in the data.



- Outlier Removal: Outliers are detected in the Temperature data using Box plots, and they are removed using the IQR (Inter-Quartile Range) method.



- Over Sampling: The target variable 'Anomaly' is imbalanced. There are only 24 anomalies present out of 1000 data points. So, SMOTE (Synthetic Minority Over-sampling Technique) is used to handle data imbalance.



- Data Scaling: Sensor data is scaled to ensure consistency and comparability across different features. Standard scaling techniques are employed to normalize the data and prevent features with larger scales from dominating the model training process.

4. Feature Engineering:

Feature engineering plays a crucial role in extracting relevant information from raw sensor data and enhancing the performance of the anomaly detection model. The following feature engineering techniques are applied:

- Date-Time Features: Date and time information extracted from timestamp data are used to create additional features such as year, month, day, hour, minute, and second. These features capture temporal patterns and seasonality in the data, which are essential for detecting anomalies.

- One-Hot Encoding: Categorical features such as boiler names are encoded using one-hot encoding to convert them into numerical format. This enables the model to effectively utilize categorical information for anomaly detection.

5. Model Building:

Anomaly detection models are trained using the pre-processed data to identify deviations from normal behaviour patterns. In this project, an Isolation Forest algorithm is employed due to its ability to isolate anomalies in high-dimensional data efficiently. The following steps are involved in model building:

- Data Splitting: The pre-processed data is split into 70% training and 30% testing sets to train the model on a subset of data and evaluate its performance on unseen data.

- Pipeline Construction: A pipeline is constructed to streamline the model building process, including data scaling and model training. Pipelines ensure consistency and reproducibility in model development.

- Hyperparameter Tuning: Hyperparameters of the Isolation Forest model, such as the number of estimators and contamination rate, are tuned using grid search cross-validation to optimize model performance.

```
Best Parameters:  {'isolation_forest__contamination': 0.5, 'isolation_f
orest__n_estimators': 300}
```

6. Model Evaluation:

The performance of the anomaly detection model is evaluated using various evaluation metrics to assess its effectiveness in identifying anomalies and distinguishing between normal and anomalous behaviour. The following evaluation metrics are utilized:

- Accuracy: The overall accuracy of the model in correctly classifying instances as normal or anomalous.

```
Accuracy Score:  0.9025270758122743
```

- Precision: The proportion of true positive predictions among all positive predictions, indicating the model's ability to avoid false alarms.

```
Precision Score:  0.8939929328621908
```

- Recall: The proportion of true positive predictions among all actual positive instances, measuring the model's ability to capture anomalies.

```
Recall Score:  0.9133574007220217
```

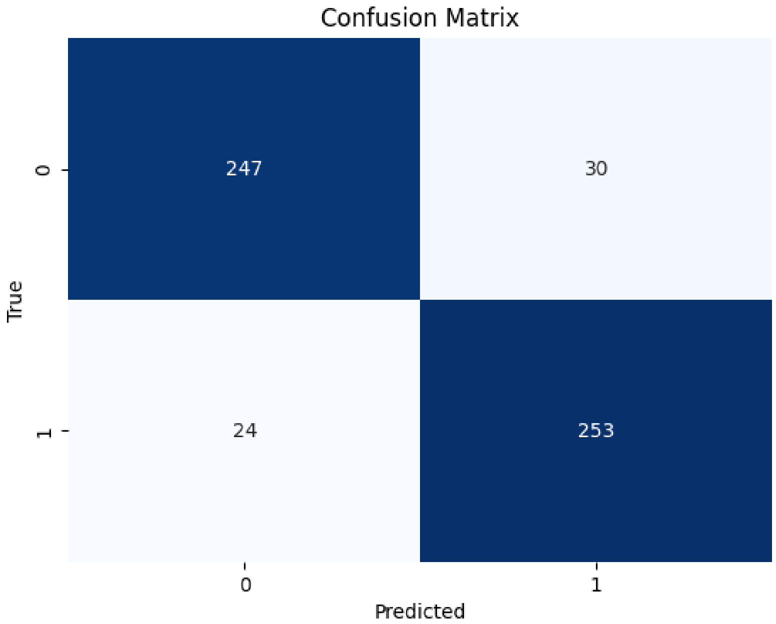
- F1-Score: The harmonic mean of precision and recall, providing a balanced measure of the model's performance.

```
F1-Score:  0.9035714285714286
```

- Classification Report: The classification report is generated to identify various metrics.

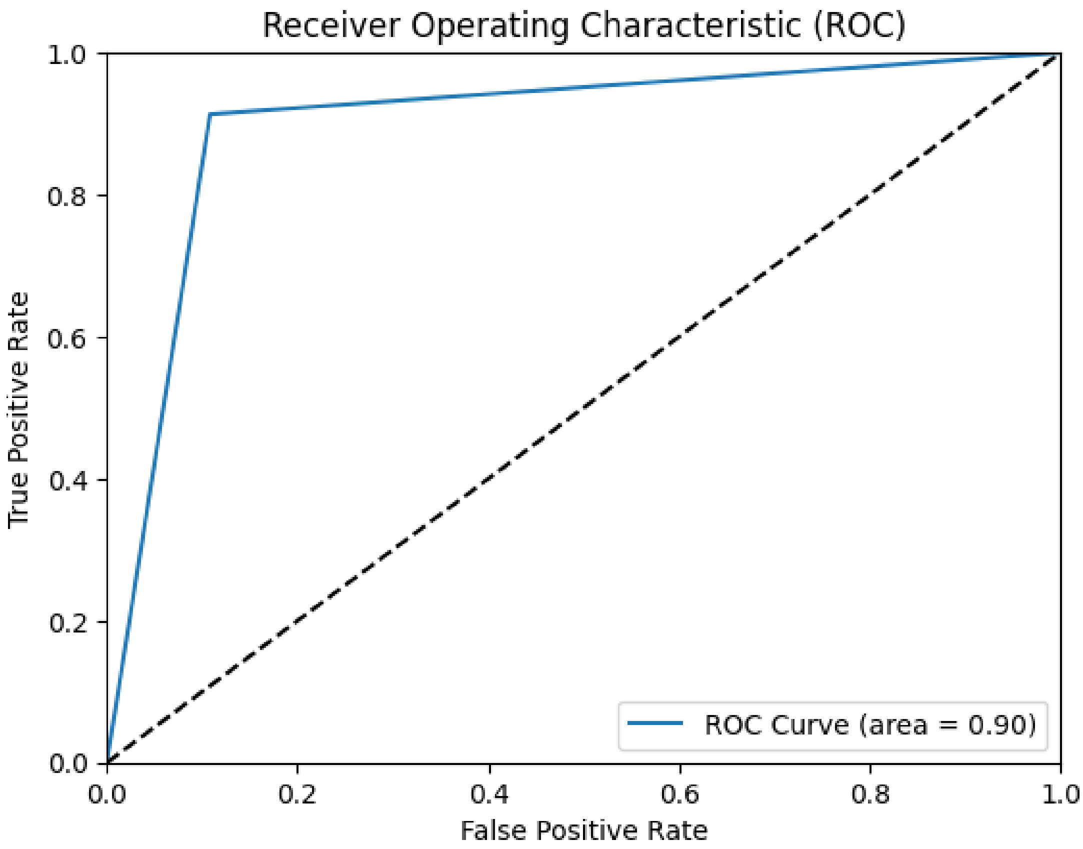
	precision	recall	f1-score	support
0	0.91	0.89	0.90	277
1	0.89	0.91	0.90	277
accuracy			0.90	554
macro avg	0.90	0.90	0.90	554
weighted avg	0.90	0.90	0.90	554

- Confusion matrix: The confusion matrix is created to determine how the model classifies the values.



- ROC Curve and AUC-ROC: The receiver operating characteristic (ROC) curve plots the true positive rate against the false positive rate at various threshold settings. The area under the ROC curve (AUC-ROC) quantifies the model's ability to discriminate between normal and anomalous behaviour.

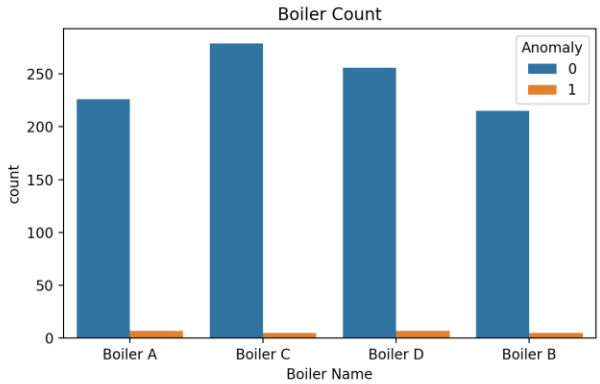
```
ROC-AUC:  0.9025270758122744
```

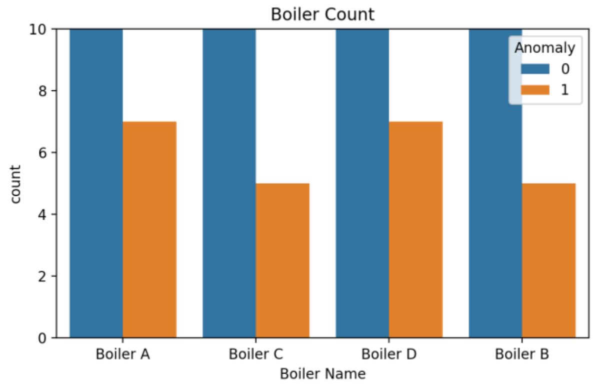


7. Visualization & Reporting:

Results from the anomaly detection system are visualized using plots and charts to provide stakeholders with actionable insights into unusual behaviour patterns. Visualizations include:

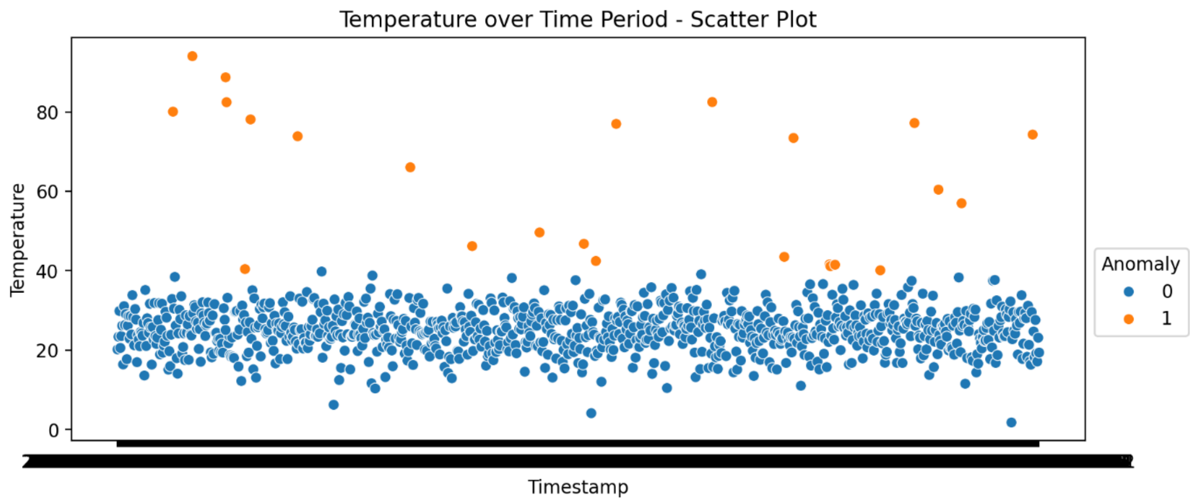
- Count plots: Displaying the distribution of anomalies across different boiler names to identify patterns and trends. Boiler A, D recorded highest number of Anomalies when compared with Boiler C and B.





- Scatter Plots: Visualizing temperature data over time to detect anomalies and observe trends in equipment behaviour. The Anomalies starts to occur when the temperature of the Boiler reaches 40 degrees and above. It reaches a maximum of 94 degrees.

```
Minimum Anamoly Temperature:  40.10736181988482
Maximum Anamoly Temperature:  94.00750069529268
```



8. Recommendations:

Based on the analysis of unusual behaviour patterns detected by the anomaly detection system, the following recommendations are suggested to prevent equipment failures or reduce downtime:

- Early Maintenance: Schedule maintenance for equipment exhibiting unusual behaviour to prevent potential failures and ensure uninterrupted operations.

- Predictive Maintenance: Utilize anomaly detection to predict equipment failures before they occur, allowing proactive maintenance and minimizing downtime.

- Process Optimization: Analyse anomalous behaviour patterns to identify underlying issues in the industrial process and optimize operations for improved efficiency and reliability.

- Training and Awareness: Train personnel to recognize and respond to anomalies promptly, enabling timely intervention and minimizing the impact of equipment failures on productivity.

9. Conclusion:

In conclusion, the developed anomaly detection system demonstrates promising results in identifying unusual behaviour in industrial equipment data, notably when the boiler temperature surpasses 40 degrees Celsius, with Boiler A and D recording the most occurrences. Therefore, prioritizing maintenance for these boilers is crucial. Additionally, implementing a fail-safe system to reduce the temperature of the boilers when anomalies are detected is recommended.

By leveraging advanced data preprocessing techniques, feature engineering, and machine learning algorithms, the system can effectively detect anomalies and provide actionable insights for preventive maintenance and process optimization. Continuous monitoring and refinement of the anomaly detection system will be essential to ensure its effectiveness in mitigating equipment failures, reducing downtime, and enhancing operational efficiency in industrial settings.