Anomaly Detection in Temperature Data Using Logistic Regression

AMAN SHARMA

Problem Statement

•Anomalies in temperature data can indicate faults or critical events.

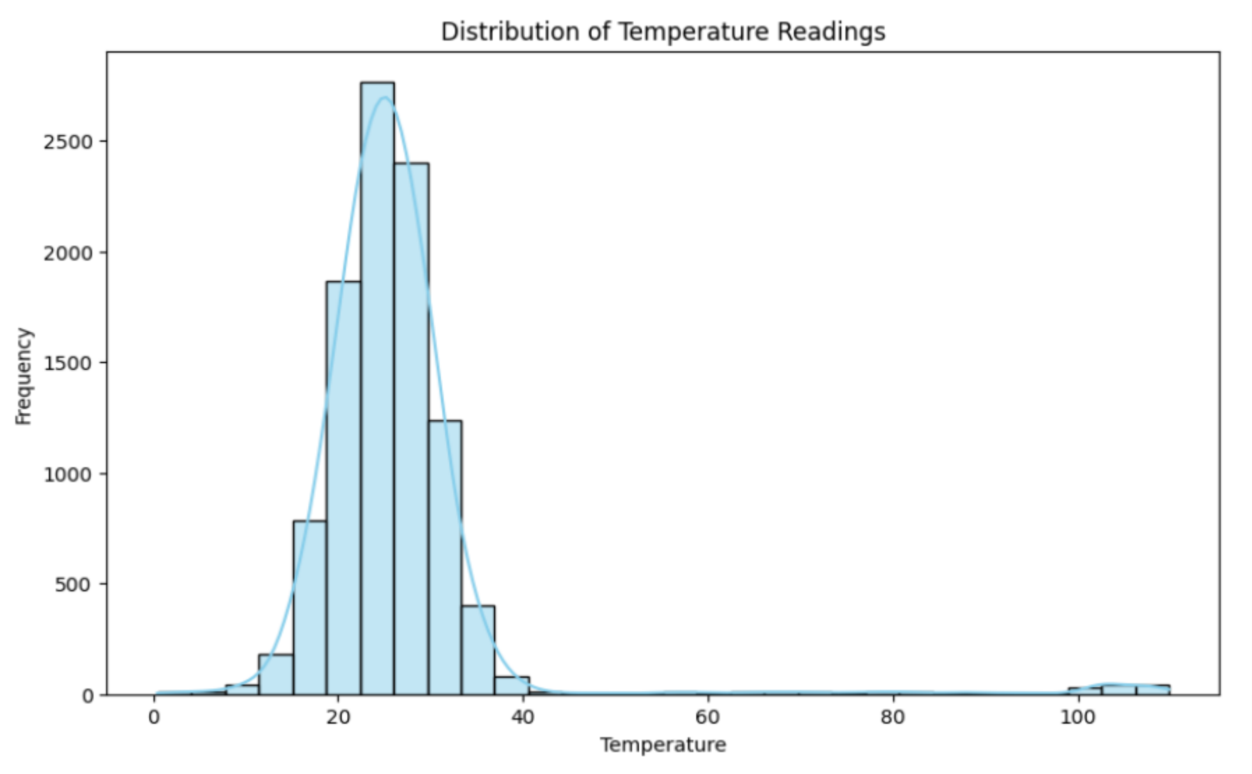
•Objective: Build a model to detect anomalies in the data using machine learning.

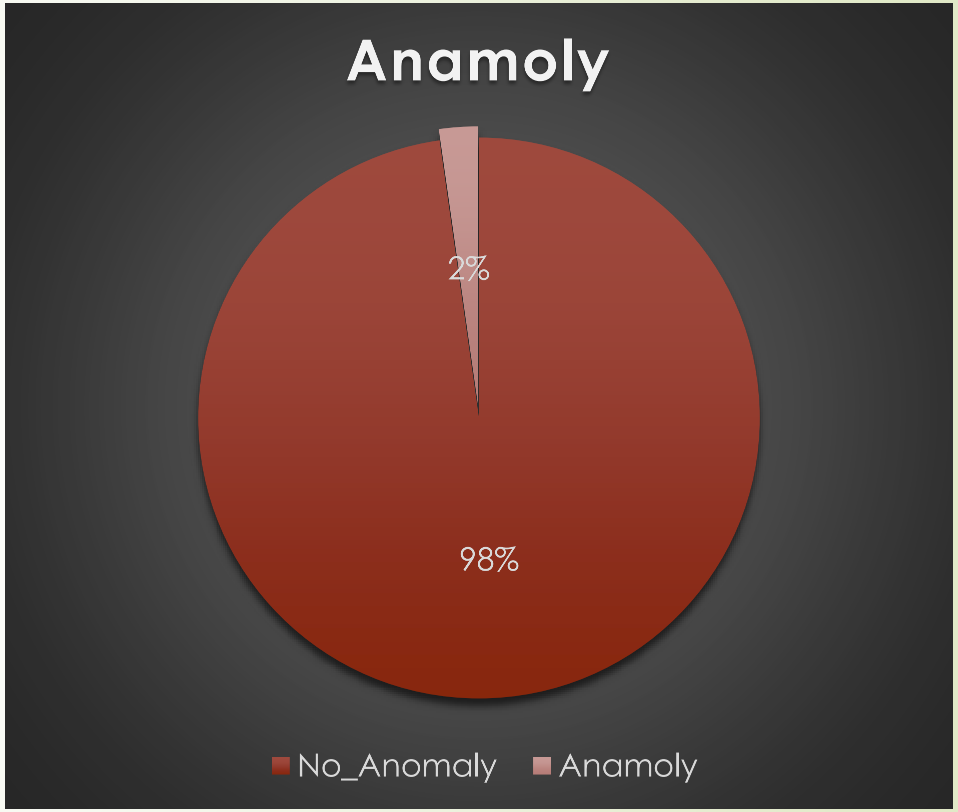
Dataset Overview

- Total entries: 10,000
- Features: Timestamp ,Temperature, Anomaly, Location.

	Timestamp	Temperature	Anomaly	Location
0	2024-06-01 19:50:28	24.476332	0	SensorB
1	2024-06-01 19:50:28	18.253966	0	SensorA
2	2024-06-01 19:50:28	14.953520	0	SensorB
3	2024-06-01 19:50:28	17.667181	0	SensorA
4	2024-06-01 19:50:28	17.079826	0	SensorA

Exploratory Data Analysis



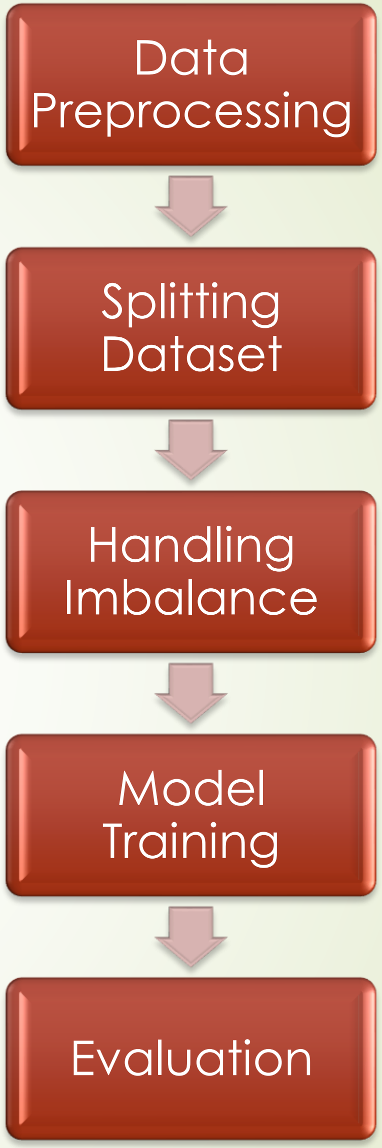


- These extreme values of temperature is anomalies , we can't remove it.

- The dataset is heavily imbalanced, with significantly more normal readings than anomalies.

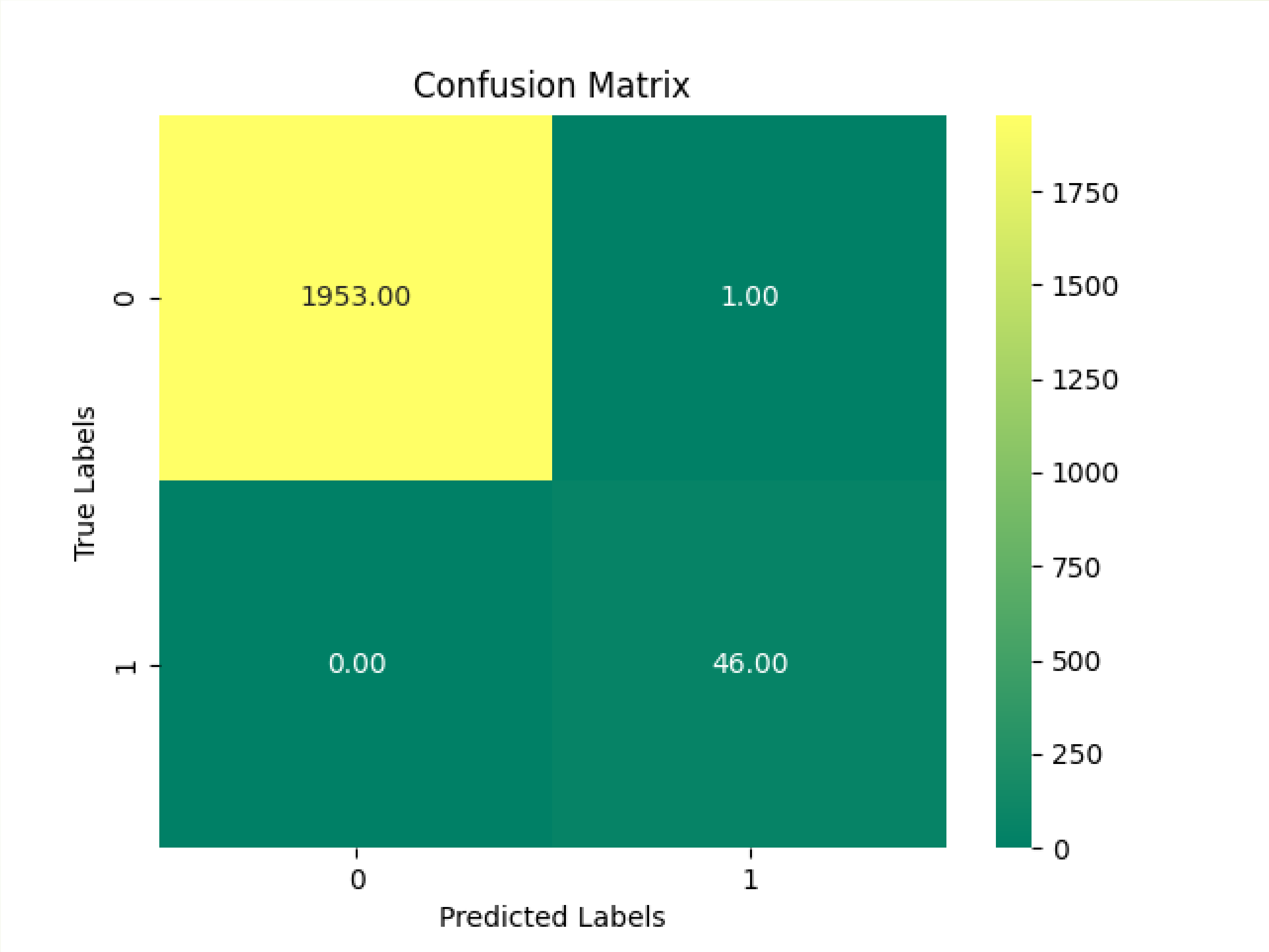
Methodology

•**Step 1:** Data Preprocessing – Focused on "Temperature" feature only.
•**Step 2:** Data Splitting – 80% for training, 20% for testing.
•**Step 3:** Handling Imbalance – Use Hybrid sampling.
•**Step 4:** Model Training – Built a Logistic Regression model.
•**Step 5:** Model Evaluation – Evaluated performance with metrics like Accuracy, F1-Score, and ROC-AUC.



Model Performance

- **Metrics :**
- Accuracy: 1.00.
- Precision: 0.98.
- Recall: 1.00.
- F1-Score: 0.99.
- ROC-AUC: 1.00.



Key Findings

- Logistic regression was highly effective for this dataset.
- Handling imbalance with hybrid sampling(oversampling and under sampling) improved results.
- Simple "Temperature" feature was sufficient for anomaly detection.

Limitations

- Results depend heavily on the "Temperature" feature.
- Possible overfitting due to high accuracy across all metrics.
- May not generalize well to other datasets or real-world scenarios.