Time Series Anomaly Detection for IoT Sensors

1. Problem Understanding

Industrial machines are monitored using sensors that collect data continuously over time. Sometimes the sensor values behave abnormally due to machine wear, damage, or failure.

The goal of this project is to **detect unusual patterns (anomalies)** in time-series sensor data that may indicate equipment problems **without using labelled anomaly data**.

This approach is realistic because, in real industries, failure labels are rarely available.

2. Dataset Used

NASA Bearing Dataset

Why this dataset?

- Contains real vibration sensor data
- Represents industrial rotating machinery
- Commonly used for predictive maintenance research
- Closely matches real IoT sensor behaviour

Each file represents sensor readings collected at a specific time.

3. Data Preparation & Exploration

- Sensor files were loaded in chronological order
- Each file contains thousands of vibration readings
- Raw data was verified to be numeric and consistent
- No missing values were found

4. Feature Engineering

Raw sensor signals are very large and not directly suitable for machine learning.

So, **statistical features** were extracted from each time window:

- **Mean** – average vibration level
- **Standard Deviation (STD)** – vibration variability
- **RMS** – vibration energy
- **Maximum value** – sudden vibration spikes

These features are commonly used in industrial condition monitoring.

All features were **standardized** using z-score scaling so that no single feature dominates the model.

5. Anomaly Detection Models

Two unsupervised models were used.

Model 1: Isolation Forest

Why Isolation Forest?

- No labels required
- Works well for rare abnormal points
- Fast and easy to interpret

How it works:
It isolates data points using random splits.
Points that are isolated quickly are considered anomalies.

Model 2: Autoencoder

Why Autoencoder?

- Learns normal data behaviour
- Suitable for continuous sensor data

How it works:

- Trained to reconstruct normal sensor patterns
- High reconstruction error = anomaly

6. Model Evaluation (Without Labels)

Since no anomaly labels were available:

- Accuracy, precision, and recall could not be calculated

Instead, validation was done using:

- Visual inspection of anomaly plots
- Domain knowledge (increasing vibration = wear)
- Agreement between Isolation Forest and Autoencoder results

Both models detected anomalies mainly toward the later time periods, increasing confidence in the results.

7. Key Findings & Business Insights

- Vibration variability increases over time
- Both models detect abnormal behaviour near the end of the timeline
- Results align with real-world bearing degradation patterns

Business value:

- Early detection of equipment issues
- Reduced downtime
- Supports predictive maintenance strategies

8. Limitations & Future Improvements

Limitations

- No ground-truth anomaly labels
- Threshold selection is heuristic
- Offline analysis only

Future Improvements

- Use LSTM models to capture temporal patterns
- Add real-time anomaly alerts
- Validate using labelled industrial datasets