Project Report: Predicting Machine Failures Using Sensor Data

---

Submitted by: Mukta Motwani

Submission Date: 30th June 2025

---

1. Abstract

This project aims to predict machine failures using sensor data collected from various machines. Using historical records, we apply machine learning techniques to build a predictive model that helps in identifying potential failures ahead of time. This approach supports preventive maintenance and reduces unexpected machine downtime.

---

2. Introduction

In industrial environments, unplanned machine failures can lead to costly downtimes and safety risks. By analyzing real-time sensor readings such as temperature, air quality, and chemical emissions, it is possible to forecast failures. The goal of this project is to use supervised learning techniques to accurately classify whether a machine will fail based on these sensor inputs.

---

3. Dataset Description

The dataset consists of 944 records with the following 10 columns:

- **Features:**
- `footfall`
- `tempMode`
- `AQ` (Air Quality)
- `USS`
- `CS`
- `VOC` (Volatile Organic Compounds)
- `RP`
- `IP`
- `Temperature`
- **Target:**
- `fail` (0 = No Failure, 1 = Failure)

All values are integers with no missing data.

---

4. Data Preprocessing

- Checked for null values: None found.
- All features are numeric; no encoding needed.
- The dataset is relatively balanced: 551 non-failure and 393 failure records.

---

5. Exploratory Data Analysis (EDA)

- **Class Distribution:** Reasonably balanced.
- **Correlation Analysis:**
- Strong positive correlation between `VOC` and `fail` (0.79)
- Strong negative correlation between `USS` and `fail` (-0.46)

These relationships indicate sensor readings that are key predictors.

---

6. Model Building

- **Algorithm Used:** Random Forest Classifier
- **Data Split:** 80% used for training, 20% used for testing

Performance Metrics:

- **Accuracy:** 87.8%
- **Precision (Class 1 - Failure):** 86%
- **Recall (Class 1):** 87%
- **F1 Score:** 87%
- **ROC-AUC Score:** 87.8%

---

7. Confusion Matrix

```
[[90, 12],
 [11, 76]]
```

This shows a strong balance of true positives and true negatives, indicating good classification performance.

---

8. Conclusion

The model is able to predict machine failures with high accuracy using simple sensor inputs. Features like VOC, AQ, and USS show strong influence on the outcome. This predictive capability can significantly improve preventive maintenance scheduling.

---

9. References

- scikit-learn documentation
- Random Forest classifier articles
- Dataset provided for project