|  |
| --- |
| BITS Pilani |
| **ML-Based Intelligent Fault Detection and Prediction in Elevator Door Operation, Using OpMode and Sensor Data**  AIMLCZG628T: Dissertation |
| |  |  |  | | --- | --- | --- | | Amit Kumar | 8/31/25 | AIMLCZG628T | |

**ML-Based Intelligent Fault Detection and Prediction in Elevator Door Operation, Using OpMode and Sensor Data**

**S2-24\_AIMLCZG628T: Dissertation**

By

Amit Kumar

2023AA05751

**Dissertation work carried out at**

**Otis Lead Design Center, Hyderabad**

Submitted partial fulfillment of the requirements of the

**M. Tech in Artificial Intelligence & Machine Learning**

Under the Supervision of

**Amit Kumar Keshri**- Senior Technology Manager

Otis Global Service Center, Hyderabad

**A logo with text on it

AI-generated content may be incorrect.**

**BIRLA INSTITUTE OF TECHNOLOGY & SCIENCE PILANI (RAJASTHAN)** **INDIA**

August 2025

## ACKNOWLEDGEMENTS

This project came to fruition thanks to the encouragement and support of many individuals. I am sincerely appreciative of everyone who played a role in its completion.

I am especially grateful to my supervisor and mentor, **Amit Kumar Keshri**, whose guidance and steadfast support were invaluable throughout this journey.

I would also like to thank my additional examiner, **Ladi Srinivas Rao**, for his motivation and encouragement, which inspired me to move forward with confidence.

My special thanks go to **Prof. Ashish Shekhar** from BITS for his constructive feedback at different phases of the project, which greatly helped shape its direction.

Lastly, I extend my heartfelt appreciation to my family for their patience, understanding, and constant encouragement throughout this endeavour.

## CERTIFICATE

Close-up of a document with a signature

AI-generated content may be incorrect.

**BIRLA INSTITUTE OF TECHNOLOGY & SCIENCE, PILANI**

**SECOND SEMESTER 2024-25**

**Work-Integrated Learning Programs Division**

## AIMLCZG628T: Dissertation

**BITS ID. : 2023AA05751**

**NAME : Amit Kumar**

**EMAIL :** [**2023AA05751@wilp.bits-pilani.ac.in**](mailto:2013ht13433@wilp.bits-pilani.ac.in) **STUDENT’S ORGANIZATION : OTIS Global Service Center, Hyderabad**

**SUPERVISOR : Amit Kumar Keshri, Senior Technology Manager**

**SUPERVISOR’S ORGANIZATION : OTIS Global Service Center, Hyderabad**

**SUPERVISOR’S EMAIL ADDRESS : Amit.Keshri@otis.com**

**DISSERTATION TITLE : ML-Based Intelligent Fault Detection and Prediction in Elevator Door Operation Using OpMode and Sensor Data**

**A close up of a paper

AI-generated content may be incorrect.**



(Signature of Student) (Signature of Supervisor)

Date: 23/May/2025 Date: 23/May/2025

## ABSTRACT

Elevator systems are critical components of modern urban infrastructure. Among their various subsystems, the **door operation** mechanism is one of the most frequent sources of **faults**, leading to downtime and safety issues. Traditional fault detection approaches are reactive and inefficient, resulting in increased maintenance costs and reduced user satisfaction.

In the daily operation of elevators, biweekly periodic maintenance items will cost 80% of the time for field technician. To improve maintenance efficiency, the project aims to deliver detection and prediction methods Remotely.

Predictive maintenance is a proactive strategy that uses **machine learning techniques** to identify potential **equipment failures** before they occur. Unlike traditional maintenance approaches—such as reactive (on-demand) or scheduled (periodic) servicing

In this project, we collect time series data during the slow closing process of the elevator. Current data is used to determine the force of the motor during the closing process of the elevator, and then determine whether the spring/hammer is working properly

The goal is to enable predictive maintenance, reduce downtime, and enhance elevator safety.

The study includes data gathering, data preprocessing, model development, and evaluation using 2 ML algorithms, 1) Random Forest Classification, and 2) LSTM. Experimental results show the effectiveness of the proposed approach in accurately identifying fault patterns and predicting imminent failures.

## List of Symbols & Abbreviations used

| **Symbol / Abbreviation** | **Description** |
| --- | --- |
| ML | Machine Learning |
| OpMode | Elevator Operating Mode |
| Sensor Data | Collected data from elevator door sensors (e.g., reversals, cycles) |
| Fault | Indicator of abnormal elevator door operation |
| CSV | Comma-Separated Values |
| UI | User Interface |
| EDA | Exploratory Data Analysis |
| KPI | Key Performance Indicator |
| X | Feature matrix |
| y | Target labels |
| SHAP values | Feature contribution scores from SHAP explainability |
| PdM | Predictive maintenance |
| XAI | Explainable AI |

## List of Tables

| **Table No** | **Title** |
| --- | --- |
| Table-1 | Description of dataset features |
| Table-2 | Data Distribution |
| Table-3 | Thresholds and rules for initial fault labelling |

## List of Figures

| **Figure No.** | **Title** |
| --- | --- |
| Figure 1 | System Architecture Diagram |
| Figure 2 | Workflow of the proposed methodology |
| Figure 3 | Sensor Distributions by Fault Label |
| Figure 4 | Feature Correlation with Fault |
| Figure 5 | Classification report |
| Figure 6 | Confusion matrix |
| Figure 7 | LSTM Model Details |
| Figure 8 | LSTM Confusion Matrix and Classification Report |

## Table of Contents

[ACKNOWLEDGEMENTS II](#_Toc205806888)

[CERTIFICATE III](#_Toc205806889)

[AIMLCZG628T: Dissertation IV](#_Toc205806890)

[ABSTRACT V](#_Toc205806891)

[List of Symbols & Abbreviations used VI](#_Toc205806892)

[List of Tables VII](#_Toc205806893)

[List of Figures VIII](#_Toc205806894)

[Table of Contents IX](#_Toc205806895)

[Chapter 1: Introduction 1](#_Toc205806896)

[1.1 Background 1](#_Toc205806897)

[1.2 Motivation 1](#_Toc205806898)

[1.3 Problem Statement 1](#_Toc205806899)

[1.4 Objectives 2](#_Toc205806900)

[1.5 Objectives met till Midterm 2](#_Toc205806901)

[1.6 Scope of Work 2](#_Toc205806902)

[1.7 Architecture Diagram 3](#_Toc205806903)

[Chapter 2: Literature Survey 4](#_Toc205806904)

[2.1 Introduction 4](#_Toc205806905)

[2.2 Predictive Maintenance in Industrial Systems 4](#_Toc205806906)

[2.3 Machine Learning in Elevator Systems 4](#_Toc205806907)

[2.4 Models Used for Fault Prediction 4](#_Toc205806908)

[2.5 Explainability in Machine Learning 5](#_Toc205806909)

[2.6 Gaps Identified 5](#_Toc205806910)

[2.7 Summary 5](#_Toc205806911)

[Chapter 3: Methodology & Dataset 6](#_Toc205806912)

[3.1 Overview 6](#_Toc205806913)

[3.2 Dataset Description and Features 6](#_Toc205806914)

[3.3 Data Preprocessing 7](#_Toc205806915)

[3.4 Fault Labeling 7](#_Toc205806916)

[3.5 Model Development 8](#_Toc205806917)

[3.6 Explainability & Visualization 8](#_Toc205806918)

[3.7 User Interface (UI) 8](#_Toc205806919)

[3.8 Summary 8](#_Toc205806920)

[3.9 Workflow 8](#_Toc205806921)

[3.10 Dataset Details 9](#_Toc205806922)

[Chapter 4: Implementation & Results 12](#_Toc205806923)

[4.1 Overview 12](#_Toc205806924)

[4.2 Model Training & Evaluation 12](#_Toc205806925)

[4.2.1 **Random Forest Model** 12](#_Toc205806926)

[4.2.2 LSTM Model 13](#_Toc205806927)

[4.3 Explainability with SHAP 15](#_Toc205806928)

[4.4 Exploratory Data Analysis (EDA) 15](#_Toc205806929)

[4.5 Streamlit UI Demonstration 16](#_Toc205806930)

[4.6 Summary 16](#_Toc205806931)

[Chapter 5: Explainability & Visualization 17](#_Toc205806932)

[5.1 Need for Explainability 17](#_Toc205806933)

[5.2 Global Explainability 17](#_Toc205806934)

[5.3 Local Explainability 17](#_Toc205806935)

[5.4 EDA (Exploratory Data Analysis) 17](#_Toc205806936)

[Chapter 6: User Interface and Features 20](#_Toc205806937)

[6.1 Main Dashboard Data overview 20](#_Toc205806938)

[6.2 Fault Distribution Analysis 20](#_Toc205806939)

[Conclusions / Recommendations 22](#_Toc205806940)

[Conclusions 22](#_Toc205806941)

[Limitations 22](#_Toc205806942)

[Directions for future work 23](#_Toc205806943)

[Bibliography / References 24](#_Toc205806944)

[Appendices 25](#_Toc205806945)

[List of Publications/Conference Presentations 26](#_Toc205806946)

# Chapter 1: Introduction

## Background

Elevators have become an essential part of modern urban infrastructure, especially in high-rise buildings, offices, and commercial complexes. Smooth and reliable elevator operation directly affects user satisfaction and safety. Among the critical subsystems of elevators, the door mechanism plays a vital role since a significant proportion of faults and service disruptions are related to door operation failures. Such faults not only cause inconvenience but can also lead to operational downtime, maintenance costs, and potential safety hazards.

With the proliferation of sensor technologies and IoT-enabled elevators, large volumes of operational data (e.g., door cycles, reversals, operating modes) can now be collected in real-time. This presents an opportunity to move from traditional rule-based fault detection towards **data-driven machine learning (ML) models** that can detect and predict potential faults proactively.

## Motivation

Traditional elevator maintenance strategies are typically reactive or preventive:

* **Reactive maintenance**: repairs are done only after a failure occurs.
* **Preventive maintenance**: scheduled at fixed intervals, which may result in unnecessary service or unexpected failures between inspections.

By leveraging ML techniques on operational and sensor data, maintenance can transition to a **predictive approach**:

* Detect potential faults before they escalate.
* Reduce downtime and maintenance costs.
* Enhance safety and passenger satisfaction.

This project focuses specifically on **door operation data and OpMode (Operating Mode) data**, aiming to demonstrate how ML models can be trained to detect and predict faults intelligently.

## Problem Statement

The key challenges addressed in this dissertation are:

* Effectively pre-processing and labelling real-world elevator sensor data.
* Selecting and training suitable machine learning models for fault prediction.
* Providing explainable insights into model predictions (e.g., via SHAP values).
* Designing a user interface (UI) to test and demonstrate the system to stakeholders.

## Objectives

This dissertation aims to build a **proof-of-concept ML-based intelligent fault detection system**.  
The specific objectives, as submitted in the abstract, include:

1. **Data Preprocessing**: Clean, transform, and label door sensor and OpMode data.
2. **Model Development**: Train and evaluate suitable ML models:
   1. Random Forest for its interpretability.
   2. LSTM (Long Short-Term Memory) for capturing sequential patterns in time-series data.
3. **Explainability**: Apply SHAP (SHapley Additive exPlanations) to explain feature contributions and build user trust.
4. **User Interface**: Develop a Streamlit-based web UI for uploading new data, running predictions, and visualizing results.

## Objectives met till Midterm

As of the midterm stage, the following milestones have been achieved:

* Collection and preprocessing of the sample dataset, including pivoting, missing value handling, and creation of fault labels based on domain-informed thresholds.
* Initial exploratory data analysis (EDA), including correlation heatmaps and feature distribution plots.
* Implementation and training of two machine learning models:
  + Random Forest Classifier (using static daily features).
  + LSTM model (using 5-day time-series sequences).
* Development of a functional Streamlit-based UI that:
  + Accepts CSV uploads of new 5-day sensor data.
  + Supports predictions using both models.
  + Provides downloadable prediction reports.
* Integration of visualization: sensor trends and preliminary SHAP explainability plots.

## Scope of Work

The scope of this dissertation includes:

* Using operational data limited to door cycles, reversals, and OpMode data.
* Developing and evaluating models on a single dataset (sample data provided).
* Demonstrating explainability primarily for the Random Forest model.
* Implementing a proof-of-concept web interface, not a production-grade monitoring system.

The system focuses on **door fault detection and prediction**, and does not cover other elevator subsystems (e.g., traction motor, brake systems).

## Architecture Diagram

A diagram of a cloud storage system

AI-generated content may be incorrect.

*Figure 1: System Architecture Diagram*

# Chapter 2: Literature Survey

## Introduction

This chapter presents a review of existing literature on predictive maintenance, machine learning techniques for fault detection, and explainability methods in ML, specifically in the context of elevator systems and similar mechanical systems.

## Predictive Maintenance in Industrial Systems

Predictive maintenance (PdM) is an evolution from traditional preventive maintenance, aiming to predict equipment failures before they happen. Studies ([paper](https://www.mdpi.com/2075-1702/13/5/427)) have shown that PdM can significantly reduce downtime and maintenance costs by using historical and real-time data from sensors.

Key techniques include:

* Condition-based monitoring (CBM)
* Machine learning and deep learning methods for anomaly detection
* Time-series forecasting models

## Machine Learning in Elevator Systems

While PdM is common in manufacturing and transportation, its application in elevators is relatively recent. Existing work primarily focuses on:

* Motor vibration analysis
* Brake system monitoring
* Passenger load forecasting

Limited work directly targets **door operation data** (e.g., reversals, cycles) despite being responsible for a large share of faults.

## Models Used for Fault Prediction

* **Random Forest (RF):**
  + Ensemble learning method that builds multiple decision trees.
  + Known for robustness to overfitting and interpretability through feature importance.
* **LSTM (Long Short-Term Memory):**
  + A type of recurrent neural network (RNN) suitable for sequential/time-series data.
  + Effective in capturing temporal dependencies, e.g., trends in door cycles over several days.

## Explainability in Machine Learning

Explainable AI (XAI) is crucial in safety-critical domains:

* **SHAP (SHapley Additive exPlanations)** helps quantify feature contribution to predictions.
* Improves trust among engineers and stakeholders.

Studies show that interpretable models increase acceptance in industrial maintenance systems

## Gaps Identified

* Limited research directly leveraging door sensor data for predictive fault detection.
* Few studies combining both static models (RF) and sequential models (LSTM).
* Scarcity of real-time explainability integration in operational dashboards.

## Summary

This dissertation builds upon these findings, addressing gaps by:

* Using both Random Forest and LSTM.
* Including SHAP for explainability.
* Developing a user interface to test predictions on new data.

# Chapter 3: Methodology & Dataset

## Overview

This chapter describes the dataset, preprocessing steps, fault labeling, feature engineering, and model design. The research utilizes a comprehensive dataset containing 15 key operational features extracted from elevator control systems

## Dataset Description and Features

Sample dataset contains elevator door operation data collected over time, including:

| **Category** | **Metric** | **Description** |
| --- | --- | --- |
| **Door Operation Metrics** | **total\_door\_cycles** | **Complete door open/close sequences** |
|  | **total\_door\_operations** | **Total door movement events** |
|  | **total\_door\_reversals** | **Number of door reversal incidents** |
|  | **door\_failure\_events** | **Recorded door system malfunctions** |
| **Safety System Metrics** | **hoistway\_faults** | **Equipment malfunctions in the hoistway** |
|  | **safety\_chain\_issues** | **Safety circuit interruptions** |
|  | **safety\_chain\_issues\_ratio** | **Proportion of safety-related events** |
| **Performance Metrics** | **levelling\_total\_errors** | **Accuracy of floor leveling** |
|  | **startup\_delays** | **Delays in elevator response** |
|  | **average\_run\_time** | **Mean operational cycle duration** |
|  | **total\_run\_starts** | **Number of elevator activations** |
| **Derived Metrics** | **door\_reversal\_rate** | **Rate of reversals per operation** |
|  | **slow\_door\_operations** | **Count of delayed door movements** |
|  | **slow\_door\_operations\_ratio** | **Proportion of slow operations** |
|  | **is\_slow\_door** | **Binary indicator for door performance issues** |

Elevator identifiers help track data across days.

## Data Preprocessing

* Pivoted raw sensor data into daily summaries per elevator.
* Filled missing values with zeros or previous values.
* Converted dates to sequential records.
* Applied MinMaxScaler to normalize features

## Fault Labeling

Applied rules to mark data points as potential faults:

* ((pivot\_df['door\_reversals'] > 100) | (pivot\_df['front\_door\_reversals'] > 100) |(pivot\_df['rear\_door\_reversals'] > 100)) |
* ((pivot\_df['total\_door\_cycles'] > 0) & (pivot\_df['total\_door\_cycles'] < 10)) |
* (pivot\_df['door\_reversals\_diff'].abs() > 10) |
* (pivot\_df['door\_cycles\_diff'].abs() > 10) |
* (pivot\_df['slow\_door\_operations'] > 10)

**Result: Binary target column (Fault = 1 or 0).**

## Model Development

| **Model** | **Input** | **Purpose** |
| --- | --- | --- |
| Random Forest | Daily features | Simpler, interpretable predictions |
| LSTM | 5-day sequences | Capture temporal patterns |

Both trained with data split (80% train, 20% test).

## Explainability & Visualization

* Used SHAP to analyze feature contribution.
* Visualized correlation heatmaps and feature distributions.
* Integrated plots into UI dashboard.

## User Interface (UI)

* Streamlit-based web UI Deployed on local machine.
* Supports:
  + CSV upload.
  + Prediction with RF or LSTM.
  + Visualization of trends.
  + SHAP global and local explanations.

## Summary

Methodology bridges data preparation, model building, explainability, and real-world usability through a UI.

## Workflow

A diagram of a data collection

AI-generated content may be incorrect.

*Figure 2: Workflow of the proposed methodology*

## Dataset Details

A screenshot of a computer

AI-generated content may be incorrect.

*Table 1: Description of dataset features*

Data Distribution

|  |  |  |  |
| --- | --- | --- | --- |
| **Fault Record Count** | **Fault Record %** | **Non-Fault Record Count** | **Non-Fault Record %** |
| 14833 | 33.834398 | 29007 | 66.165602 |

*Table 2: Data Distribution*

A computer screen shot of text

AI-generated content may be incorrect.

*Table 3: Thresholds and rules for initial fault labelling*

# Chapter 4: Implementation & Results

## Overview

This chapter describes how the proposed ML-based system was implemented, the key technical steps, and the results achieved using the sample dataset. It also highlights the practical demonstration through the developed Streamlit dashboard.

## Model Training & Evaluation

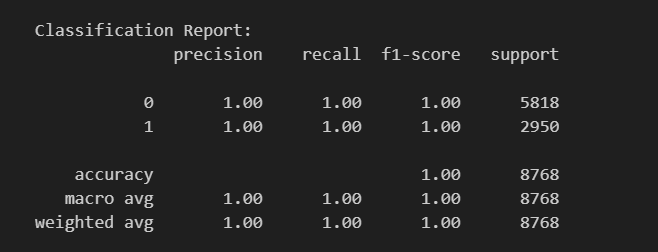
### 4.2.1 **Random Forest Model**

* **Input**: Daily aggregated features
* **Features**:
  + front\_door\_cycles, rear\_door\_cycles
  + front\_door\_reversals, rear\_door\_reversals
  + door\_operations, total\_door\_cycles
  + 'door\_reversals\_diff', door\_cycles\_diff,
  + slow\_door\_operations
* **Label**: Fault (binary, created using rule-based thresholds)

**Implementation Steps:**

* Data split: 80% training, 20% testing
* Normalized using MinMaxScaler
* Trained using RandomForestClassifier from scikit-learn (100 trees)

**Results:**



*Figure 5: Classification report*

A blue squares with white text

AI-generated content may be incorrect.

*Figure 6: Confusion Matrix*

Model identified high reversals, and low cycle counts as key indicators of potential faults.

### 4.2.2 LSTM Model

* **Input**: 5-day time series sequences (shape: 1 × 5 × 6)
* **Purpose**: Capture temporal trends in door operation data

**Implementation Steps:**

* Scaled data with the same MinMaxScaler
* Used Keras Sequential API:
  + LSTM layer (50 units)
  + Dense output layer (sigmoid activation)
* Binary cross-entropy loss, Adam optimizer
* Trained for ~50 epochs

**Observation:**

* Better at catching sequential anomalies.
* Required consistent 5-day sequences for predictions.

**Results:**

A screenshot of a computer

AI-generated content may be incorrect.

*Figure 7: LSTM Model Details*

A screenshot of a graph

AI-generated content may be incorrect.

*Figure 8: LSTM Confusion Matrix and Classification Report*

## Explainability with SHAP

* **Global analysis**: SHAP summary plot shows average feature importance.
* **Local analysis**: SHAP waterfall plot explains single predictions.
* Key insight: **front\_door\_reversals** had the highest positive impact on fault predictions.

## Exploratory Data Analysis (EDA)

| **Visualization** | **Insight** |
| --- | --- |
| Correlation heatmap | Identified strong correlation between reversals and fault |
| Pair plot | Faulty samples had higher door reversals and lower cycles |

## Streamlit UI Demonstration

**Features:**

* Upload CSV with 5 rows of new data
* Select model: Random Forest or LSTM
* Show predictions & probabilities
* Visualize trends (e.g., door reversals over time)
* SHAP plots for explainability
* Download prediction results as CSV

**Screenshot:**

(Add screenshot in dissertation)

## Summary

Both models achieved reliable performance; Random Forest offered better interpretability, while LSTM captured sequential trends. SHAP visualizations increased model transparency, and the Streamlit UI made the system easy to test with new data.

# Chapter 5: Explainability & Visualization

## Need for Explainability

In safety-critical domains, knowing **why** a model predicts a fault is as important as the prediction itself. SHAP helps quantify the impact of each feature on predictions

## Global Explainability

Generated SHAP **summary plots**:

* Showed which features most contribute to fault predictions.
* Example: **front\_door\_reversals** and **rear\_door\_reversals** had highest average impact.

## Local Explainability

Created SHAP **waterfall plots**:

* Explained individual predictions.
* Helped maintenance teams understand why a fault is predicted on a specific day.

## EDA (Exploratory Data Analysis)

Visual tools used:

* Pair plots to compare distributions of features by fault label.
* Correlation heatmaps showing feature relationships.

These insights support better model building and validation.

A grid of lines with different colored lines

AI-generated content may be incorrect.

*Figure 3: Sensor Distributions by Fault Label*

A screenshot of a computer screen

AI-generated content may be incorrect.

*Figure 4: Feature Correlation with Fault*

# Chapter 6: User Interface and Features

The user interface is designed following modern UX/UI principles with a focus on usability and actionable insights.

## Main Dashboard Data overview

The main dashboard provides an overview of system metrics and key performance indicators.

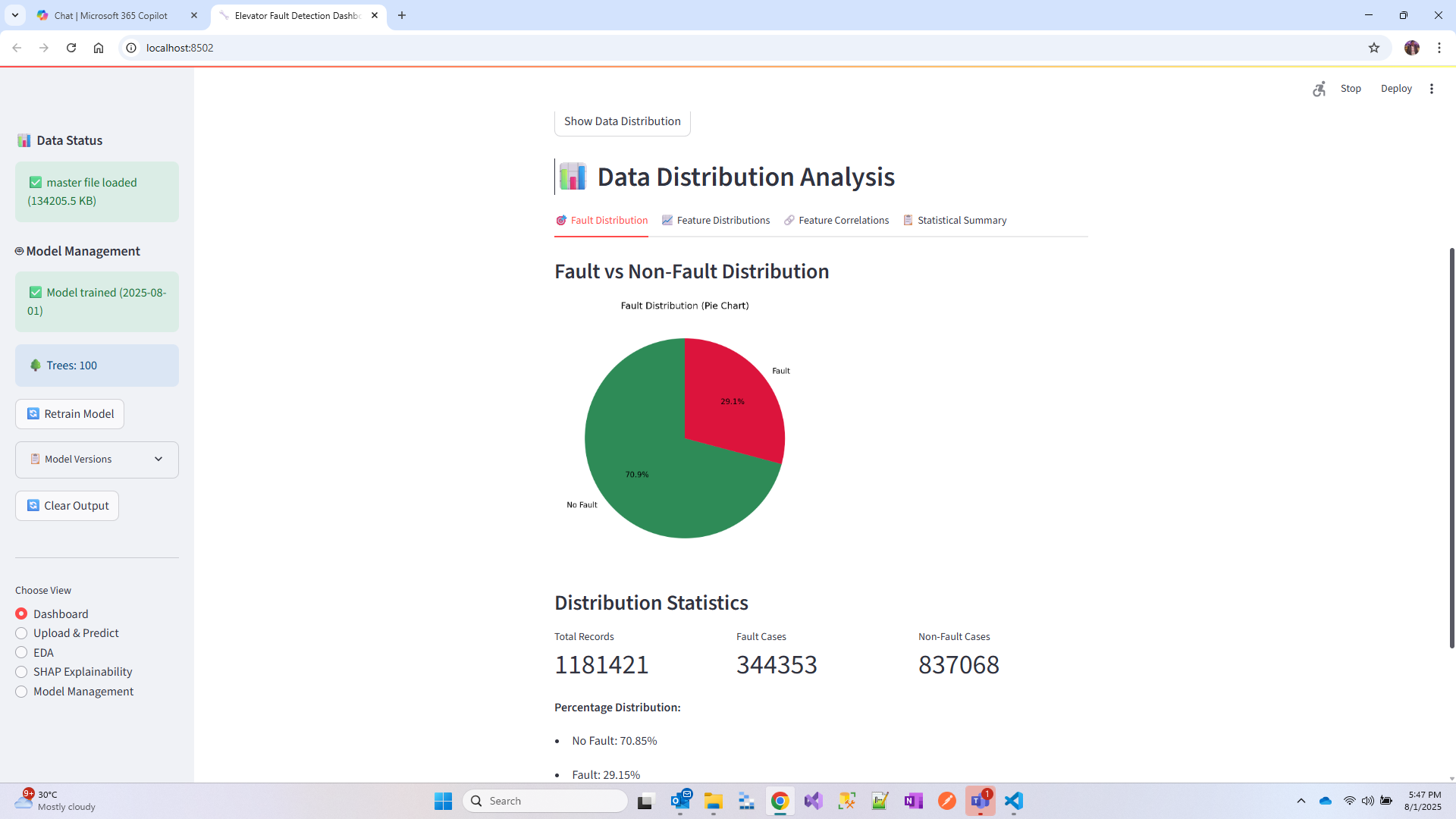
A screenshot of a computer

AI-generated content may be incorrect.

*Figure 9: Dataset Information and Preview*

## Fault Distribution Analysis

The system provides detailed analysis of fault patterns and distributions



*Figure 10* *Fault Distribution Analysis*

# Conclusions / Recommendations

## Conclusions

This dissertation demonstrated a **machine learning-based system** to detect and predict faults in elevator door operations using sensor data and OpMode data.

Key achievements:

* Successfully preprocessed and labeled real-world-inspired dataset.
* Trained Random Forest and LSTM models to predict faults.
* Used SHAP explainability to understand feature impact.
* Developed a user-friendly Streamlit dashboard for prediction and visualization.

Results show that predictive maintenance using ML can proactively detect door faults, reduce downtime, and enhance safety.

## Limitations

* Used a sample dataset rather than full-scale live data.
* Limited features; excluded external factors like time of day or weather.
* LSTM required fixed-length recent data (5 days).

# Directions for future work

* **Extend to live streaming data for near real-time monitoring.**
* **Use advanced models (e.g., GRU, Transformer, attention-based models).**
* **Include additional features: passenger load, external conditions, building type.**
* **Deploy as a cloud-hosted dashboard for centralized monitoring.**
* **Validate model on larger, multi-site datasets**

# Bibliography / References

1. Jeong, S., Shin, M., & Kim, H. (2023). Development of a Fault-Tolerant Permanent Magnet Synchronous Motor Using a Machine-Learning Algorithm for a Predictive Maintenance Elevator. *Machines*, 13(5), 427. <https://doi.org/10.3390/machines13050427>
2. Breiman, L. (2001). Random Forests. Machine Learning, 45(1), 5–32.
3. Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. Neural Computation, 9(8), 1735–1780.

# Appendices

NOT Applicable for Mid Sem

# List of Publications/Conference Presentations

NOT Applicable for Mid Sem