

# Predictive Maintenance for Railway Air Production Unit (MetroPT-3)

*Simple English. Structured for academic review.*

Student: Abdulaziz Mari

Program: M.Sc. in Artificial Intelligence — Capstone Project

Date: October 29, 2025

## Abstract

Railways must avoid failures in safety-critical systems. This project builds a clear and reproducible predictive-maintenance (PdM) pipeline. We use the open MetroPT-3 dataset from a train Air Production Unit (APU). The data includes tank pressures, differential pressures, motor current, oil temperature, valve states, and flow impulses. We detect anomalies and early signs of failure using time-series features and machine learning. We report transparent metrics and simple explanations to support maintenance decisions.

## Introduction

Modern trains rely on compressed air for brakes, doors, and suspension. Failures in the Air Production Unit lead to delays and safety risks. Public, high-quality datasets for signalling subsystems are rare. MetroPT-3 provides real multivariate time-series from an operating train APU. It is suitable for anomaly detection and early-warning tasks using machine learning.

## Problem Statement & Research Questions

We target early detection of abnormal behaviours in the APU using multivariate time-series.

- RQ1: Can we detect Anomalies and pre-failure patterns reliably on MetroPT-3?
- RQ2: Which features and models are most robust under class imbalance?
- RQ3: Can we provide clear explanations and actionable alert thresholds for operations?

## Objectives

- Build an end-to-end PdM pipeline: ingestion, cleaning, features, models, alerts.

- Compare unsupervised methods (Isolation Forest, Autoencoder) and simple supervised baselines if labels exist.
- Define practical KPIs and early-warning lead time; explain results with SHAP and compact visuals.
- Release reproducible notebooks and a concise report.

## Dataset

Primary dataset: MetroPT-3 (UCI Machine Learning Repository). Signals include: TP2, TP3 (tank pressures), H1 and DV\_pressure (differential pressures), Motor\_current, Oil\_temperature, valve states (e.g., COMP, DV\_electric), tower selection (TOWERS\_1/2), and Caudal\_impulses (flow). Sampling covers normal operation and events such as compressor starts and dryer regeneration.

## Methodology

- Data preparation: time alignment, resampling, outlier handling, missing-value imputation, event windowing.
- Feature engineering: statistical (mean, std, peaks), spectral (FFT/DWT), waveform shape (e.g., DTW distance), domain KPIs (pressure build-up time, peak current).
- Modeling: unsupervised anomaly detection (Isolation Forest, Autoencoder); supervised baselines only if labels exist.
- Evaluation: AUC-PR, Precision/Recall, F1@top-k, early-warning lead time, false alarms per day; time-based splits.
- Explainability: SHAP values and simple plots to show why an alert fired.
- Alerting: threshold calibration with hysteresis to reduce noisy alarms.

## Expected Results & Contributions

- Anomaly and early-warning detection on MetroPT-3 with clear, reproducible metrics.
- Compact dashboard that displays alerts, key signals, and explanations.
- Reproducible codebase and report suitable for future course projects or industry pilots.

## Ethics, Risks, and Limitations

- Decision support only: models do not control equipment.
- Class imbalance and label noise: prefer precision-focused metrics (AUC-PR) and threshold tuning.

- Model drift: recommend periodic re-training and monitoring.
- Generalisation limits: results are specific to APU data but methodology transfers to other train subsystems.

### References (indicative)

- MetroPT-3 Dataset — UCI Machine Learning Repository.
- Recent literature on anomaly detection in industrial time-series (2017–2025).
- Condition-based maintenance practices for rail systems.