Metro Rail – Predictive Maintenance Based On Anomaly Detection

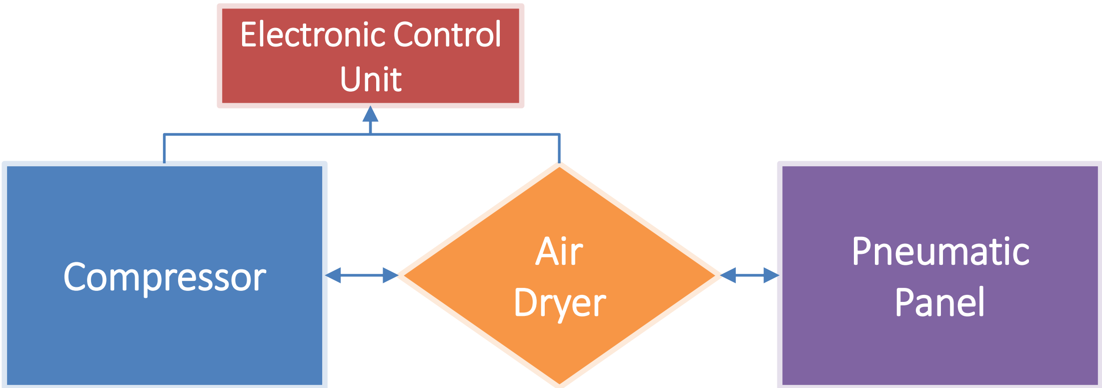
Group 21:

- Vishal Anand Gupta
- Rathil Madihalli
- Sanyam Jain
- Pritesh Singh
- Karthick Vel

CONTENTS

01
Background and Objectives

Background and Objectives



Air Production Unit (APU)

- **Installed on the roof** of metro vehicles
- **Feeds different units** which perform different functions
- **Failure could cause immediate removal** of train for repair

Predictive Maintenance

- **Maintenance can be of 3 types –** preventive, corrective and predictive
- **Preventive Maintenance** leads to waste of resources while **corrective maintenance** waits for failure to take place
- **Predictive maintenance** is the most **optimum approach** which can be used to detect anomalies and perform maintenance before the failure

Project Objectives

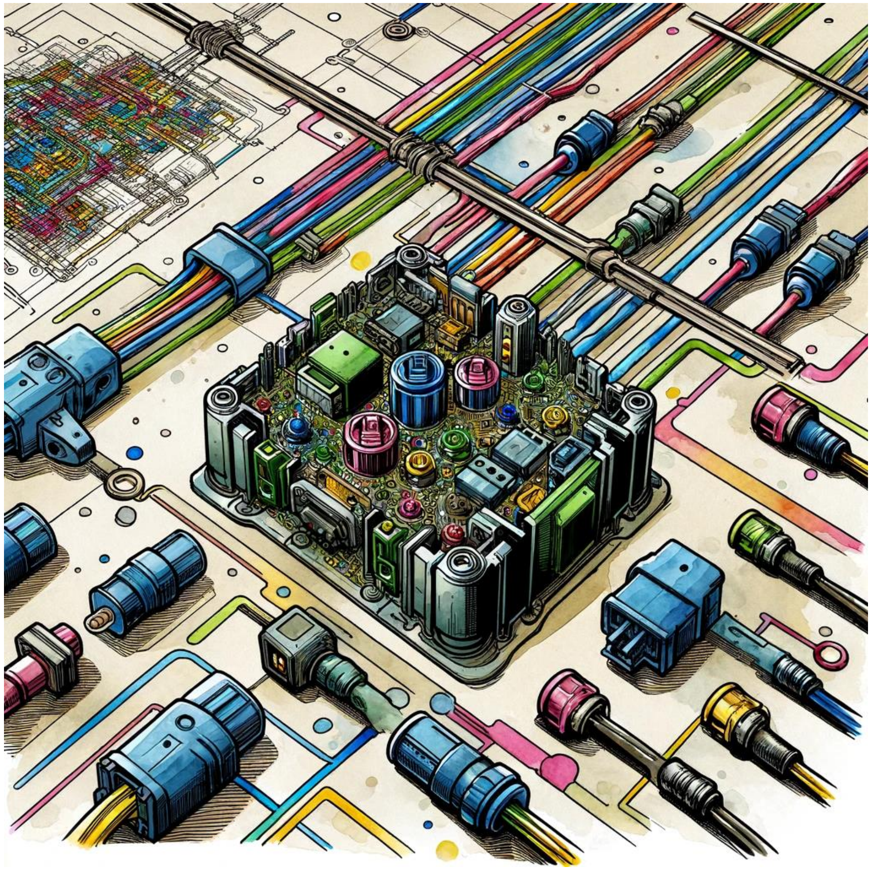
- **Reduce unforeseen maintenance –** thereby reducing the number of stops and stopping time
- **Change maintenance paradigm –** from reactive to predictive by detecting anomalies in the data before the failure actually happens
- **Reduce false alarms –** to make the process of maintenance more efficient and cost-effective

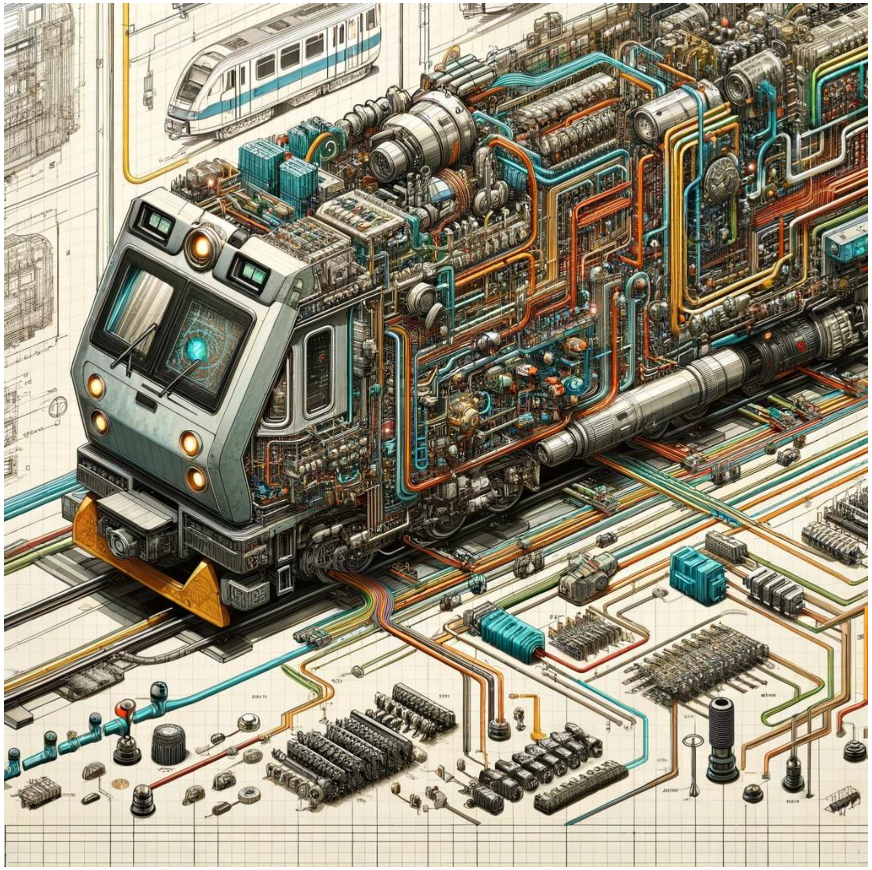
02
Data Description

Data Description

Data Source	Timeline	# Records	Frequency
UCI Machine Learning Repository	Feb’20 – Aug’20	1,516,948	Every 10 seconds

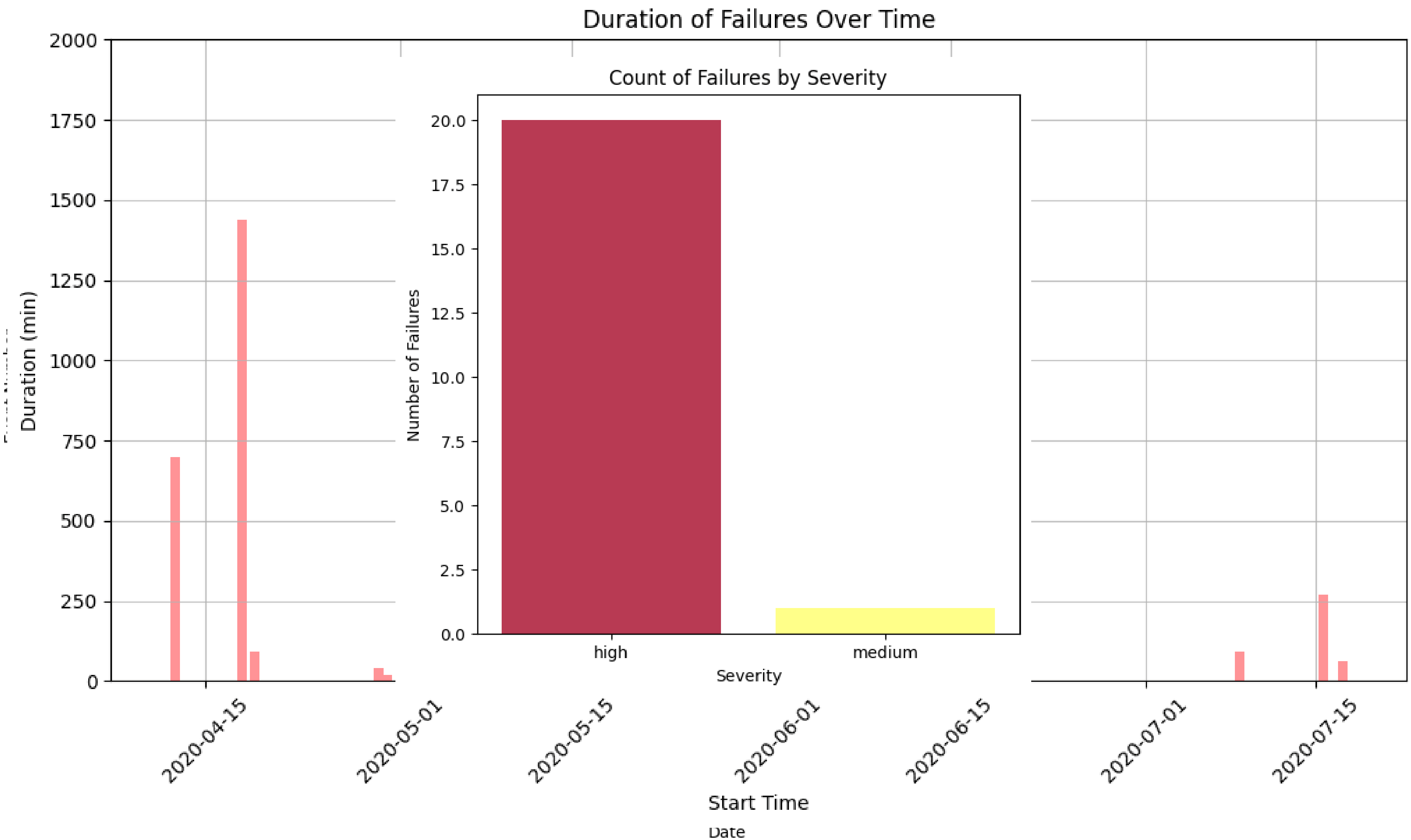
Data corresponds to the **8 digital and 7 analog sensors connected** directly with the **APU.**





Periods of Failure

Failures are infrequent, but usually severe



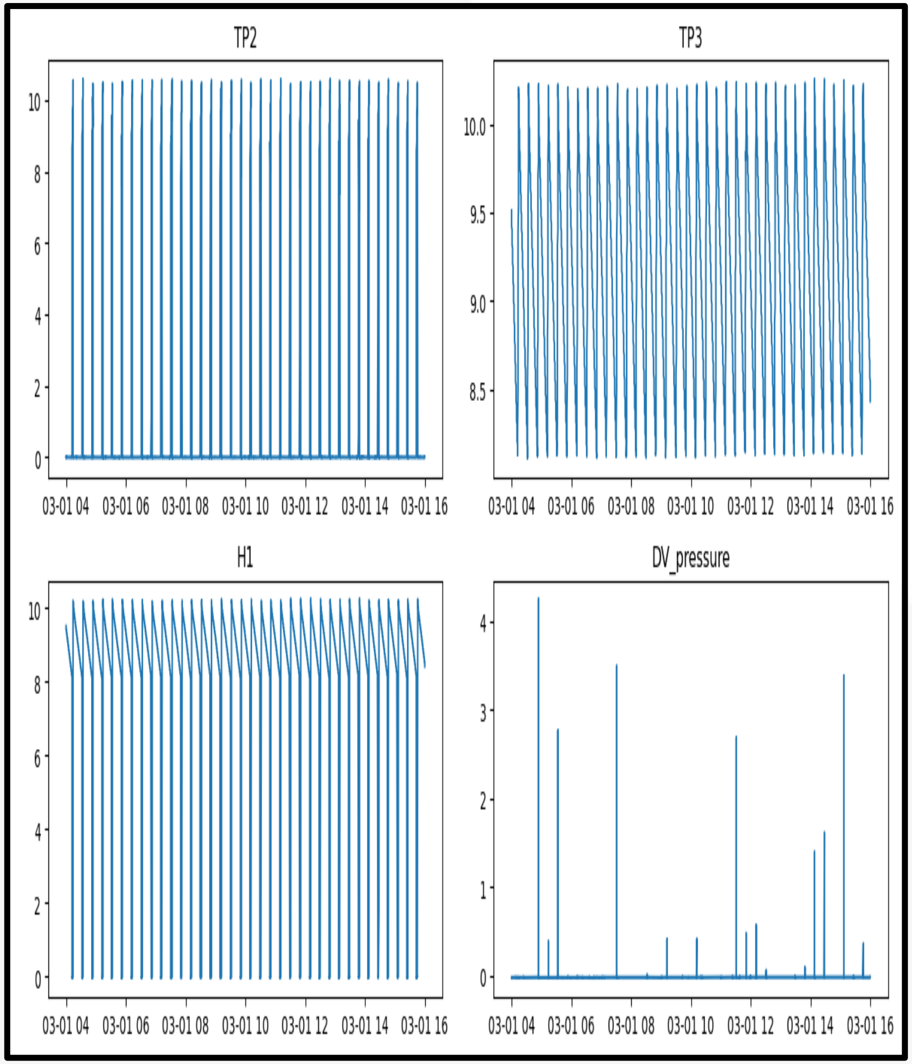
03
Exploratory Data Analysis

1. Normal Day vs Failure Day Sensor Data

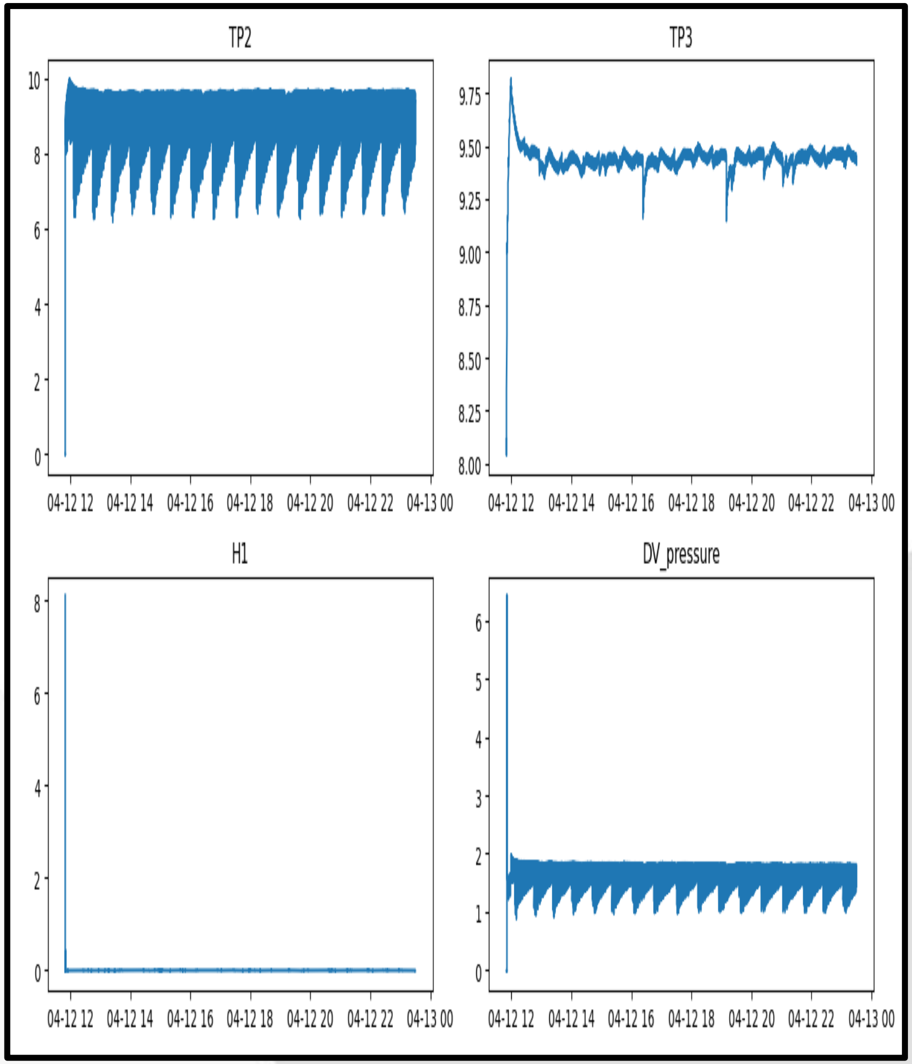
2. Leading to Failure Data

Normal Day v/s Failure Day Analog Signal Trends

Normal Day

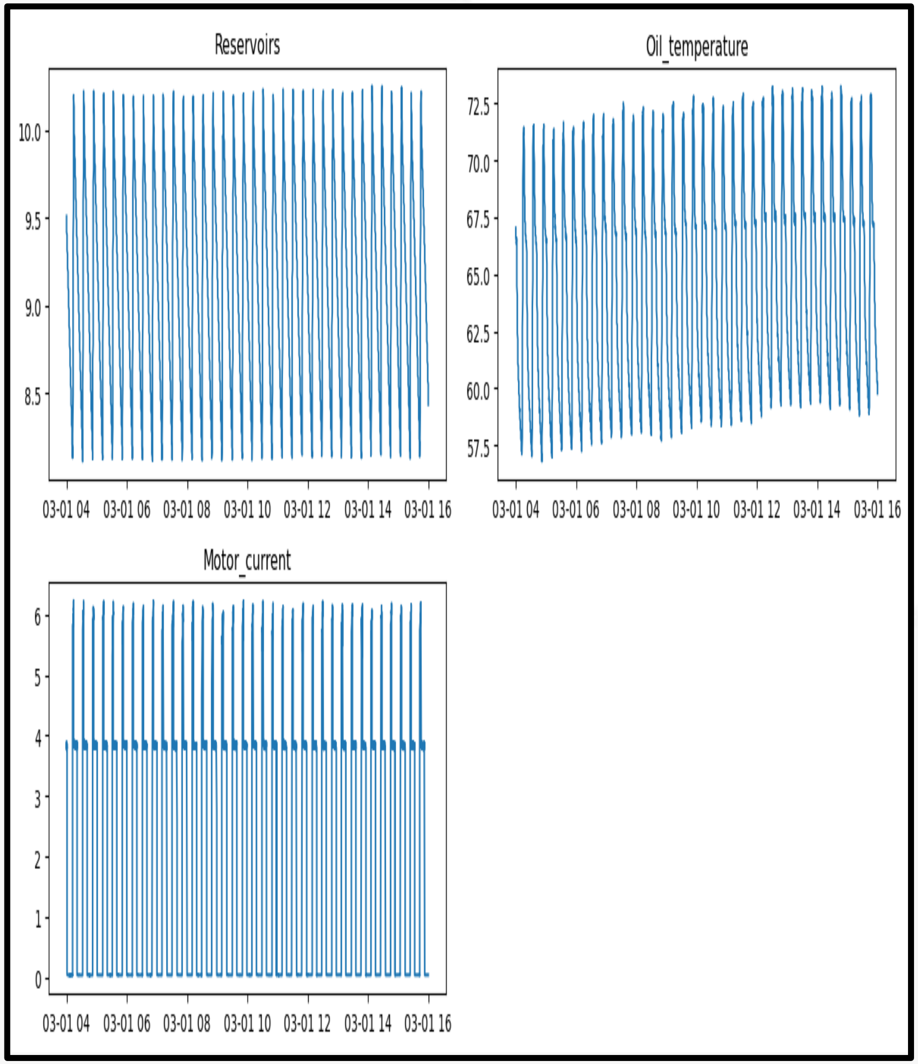


Failure Day

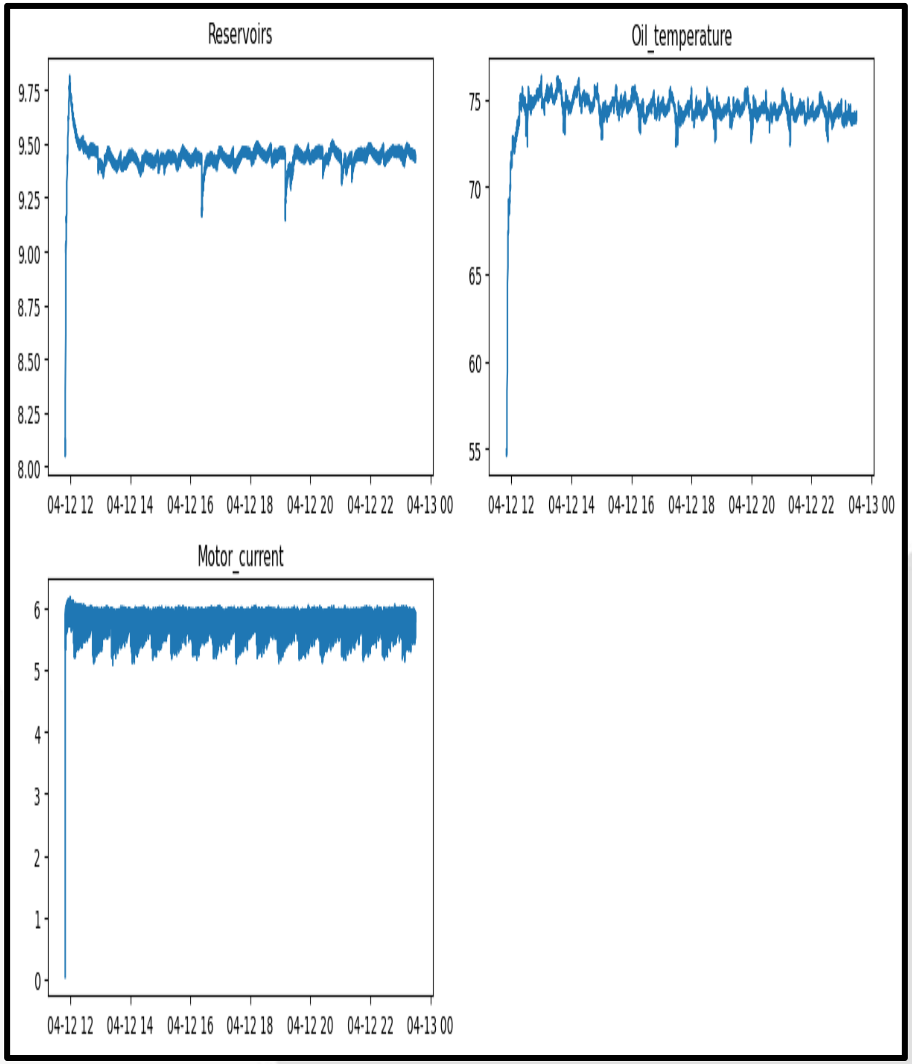


Normal Day v/s Failure Day Analog Signal Trends

Normal Day

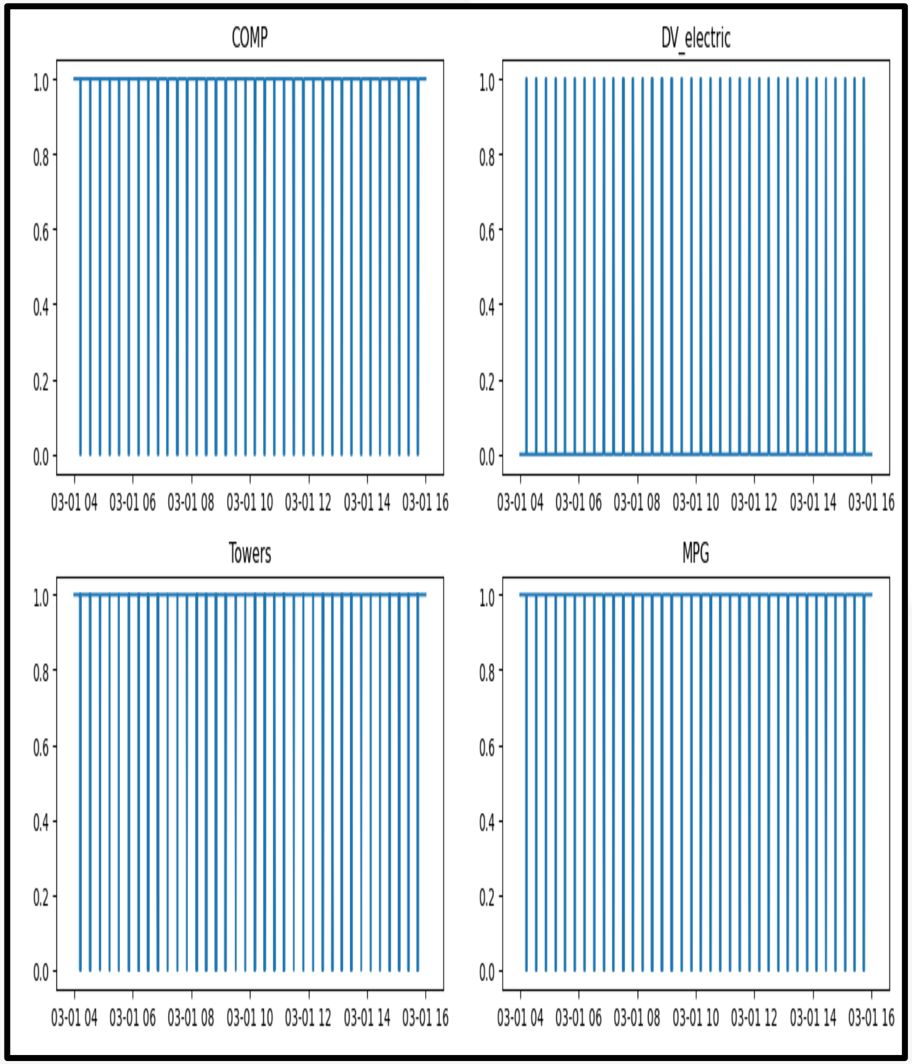


Failure Day

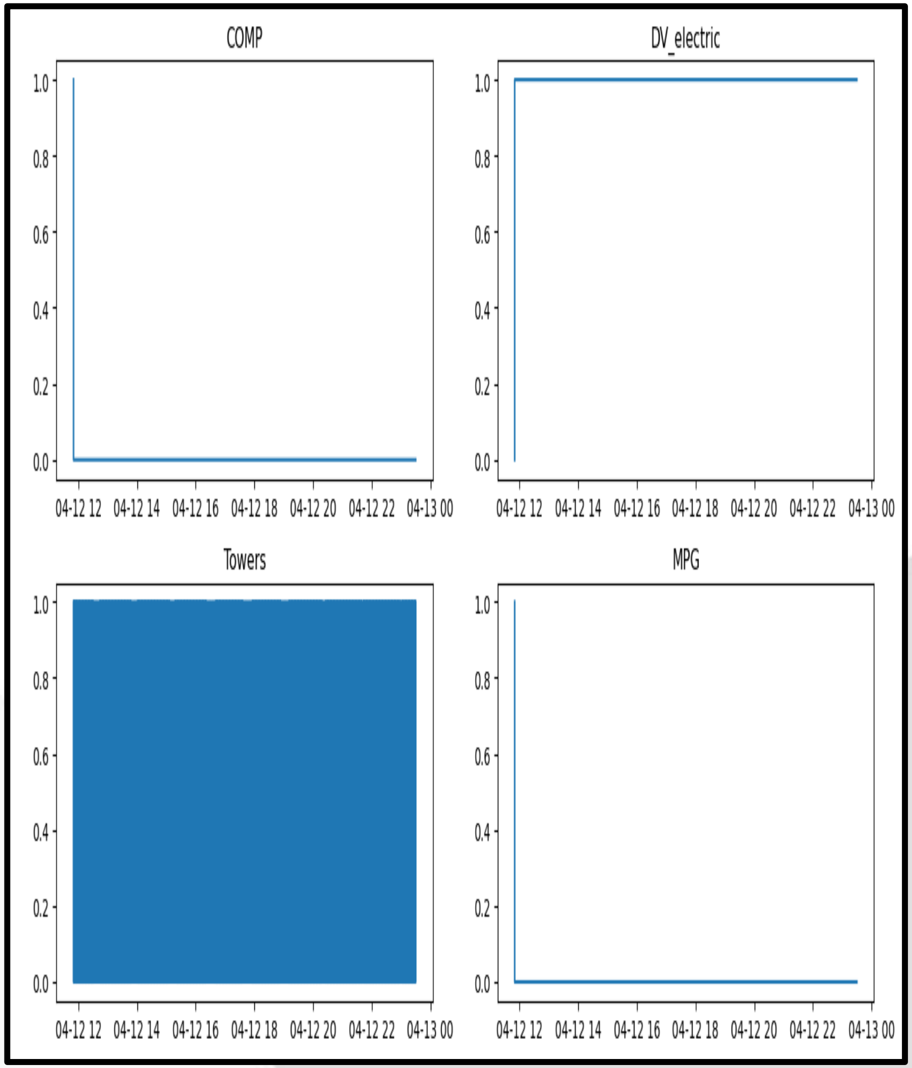


Normal Day v/s Failure Day Digital Signal Trends

Normal Day

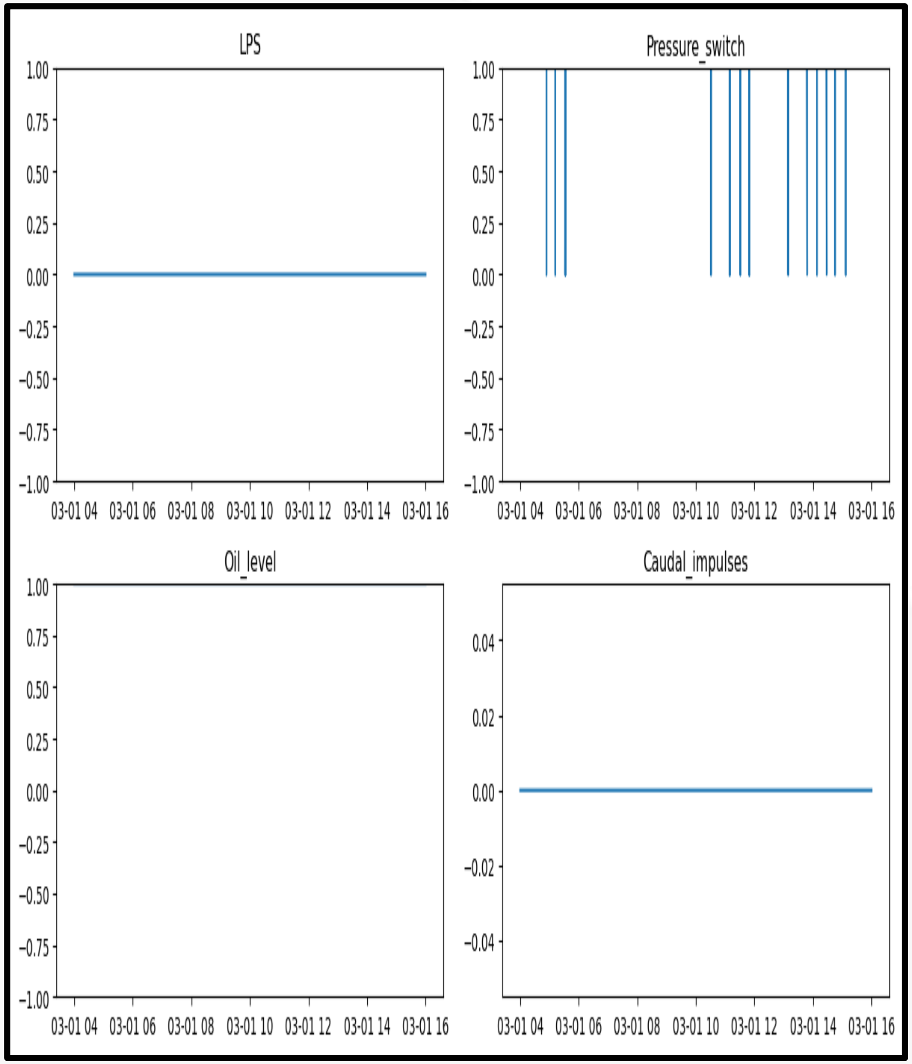


Failure Day

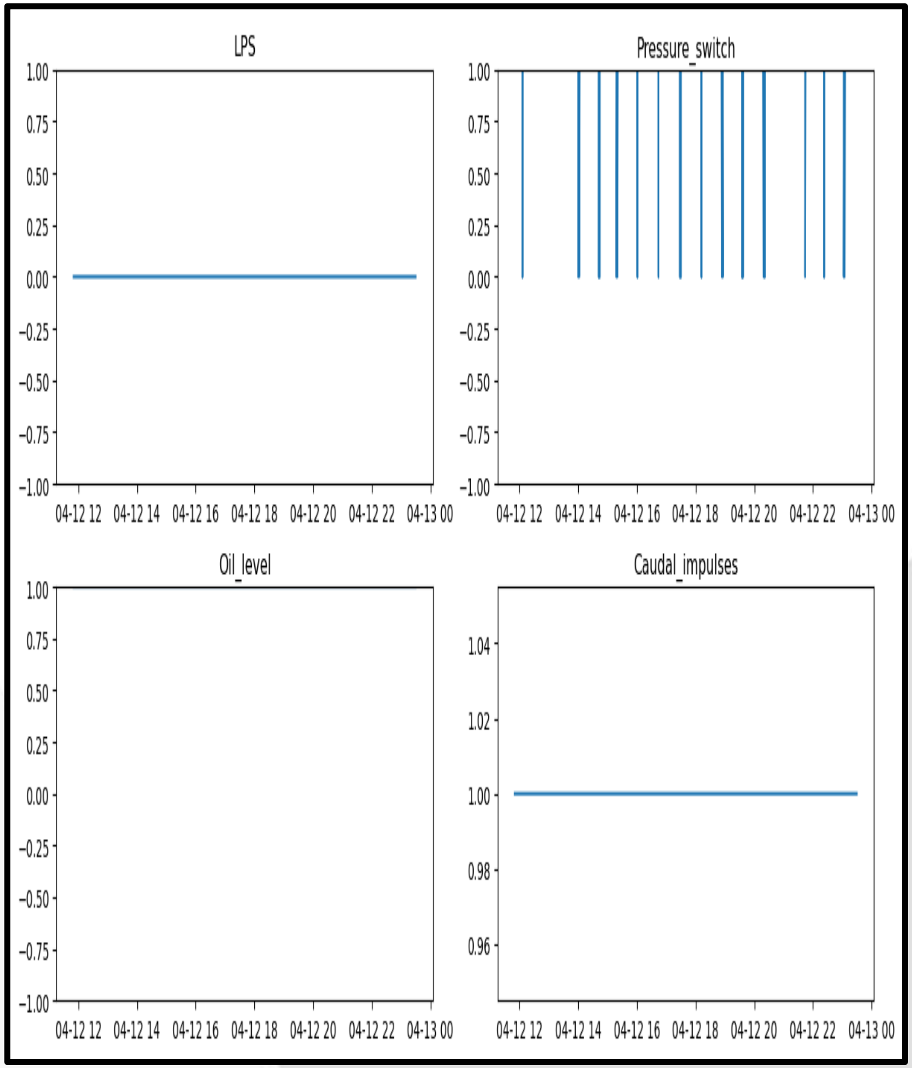


Normal Day v/s Failure Day Digital Signal Trends

Normal Day

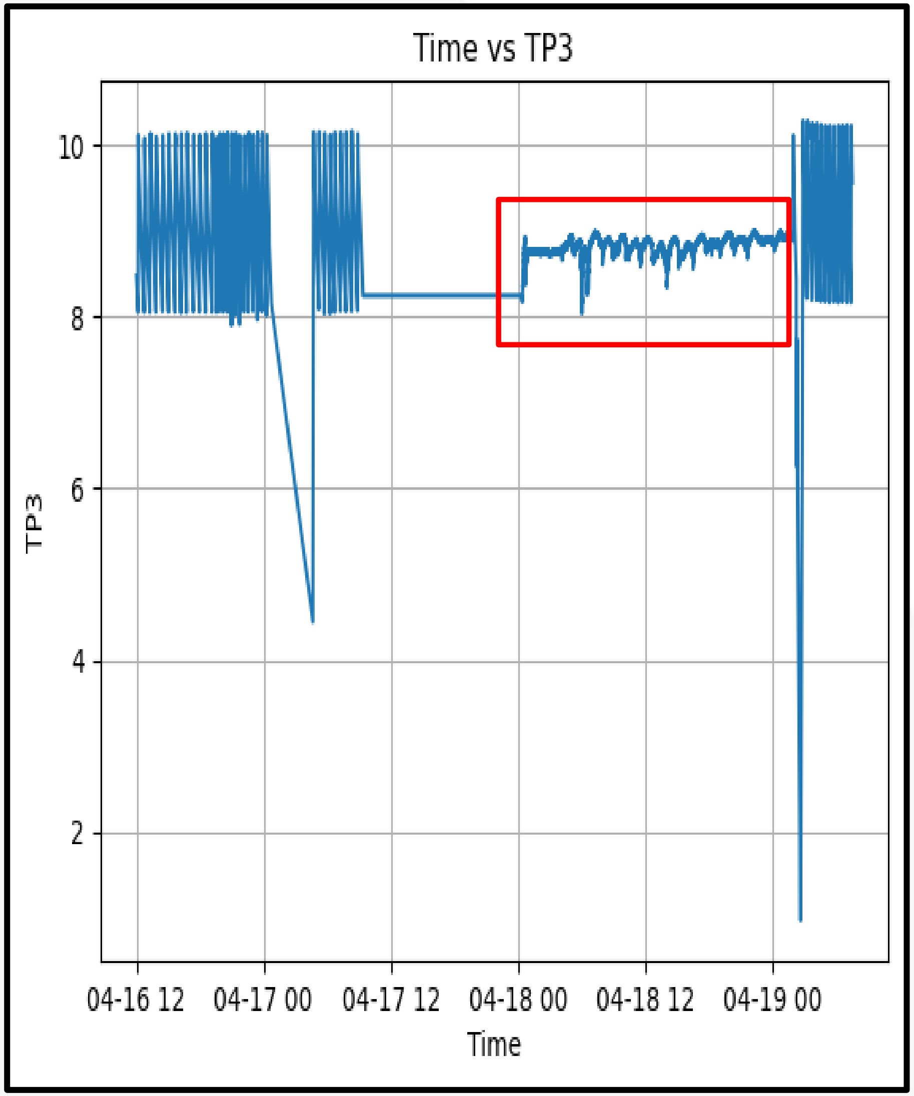


Failure Day

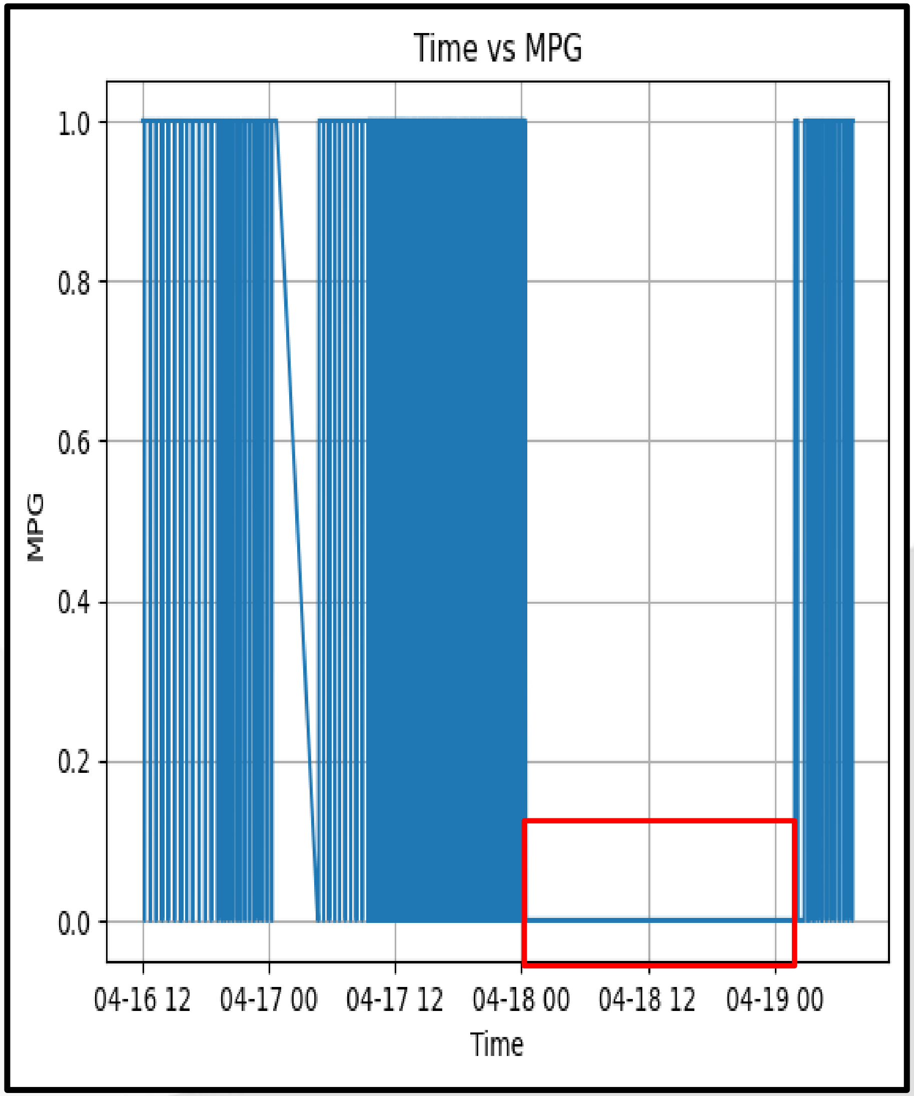


Normal Day v/s Failure Day Sensor Trends

Analog



Digital



04
Preliminary Approach

Classification Approach to Failure Detection

We have labeled data, and we have to predict failure. How difficult can it be? We can use regular classification approaches for this task.

Model	Precision	Recall	F1 Score
Decision Trees	98.2%	97.9%	98.0%
Random Forest	98.5%	98.7%	98.6%
Neural Network	97.4%	91.3%	94.3%

Too good to be true!! What’s the catch?

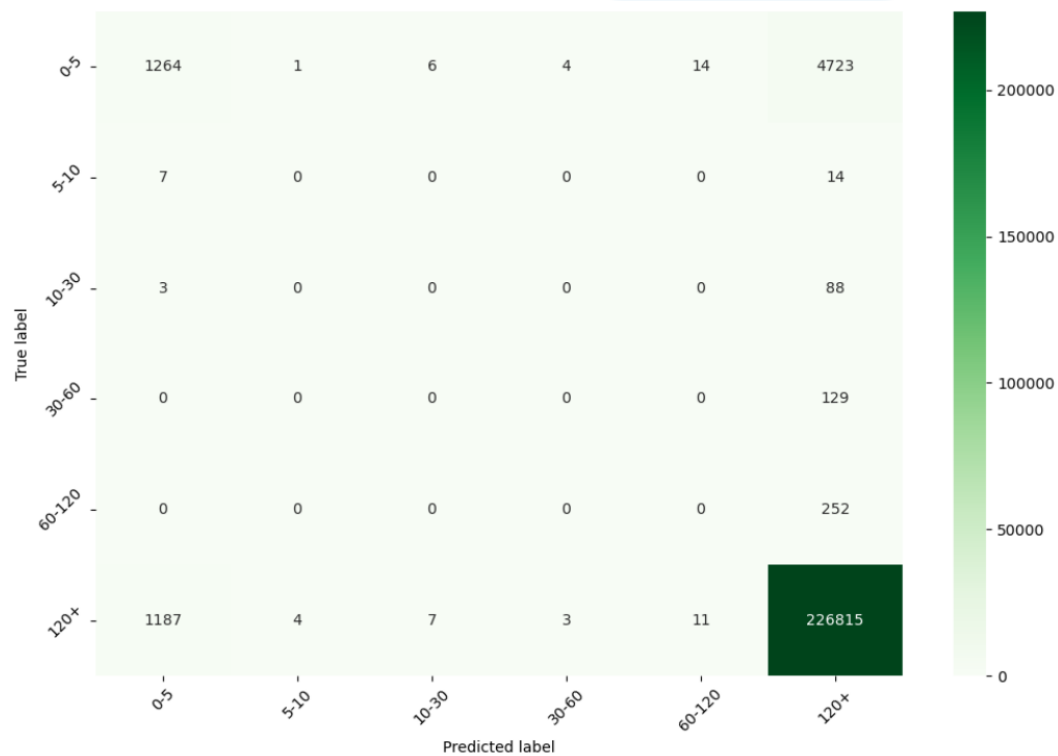
Time to Failure Approach

Process

1. Find aggregate values of features for each row to account history
2. Create Target Variable (time to failure)
3. Split data & run regression model to predict time to failure on test set
4. Convert test set & prediction values to classes & calculate metrics

Results

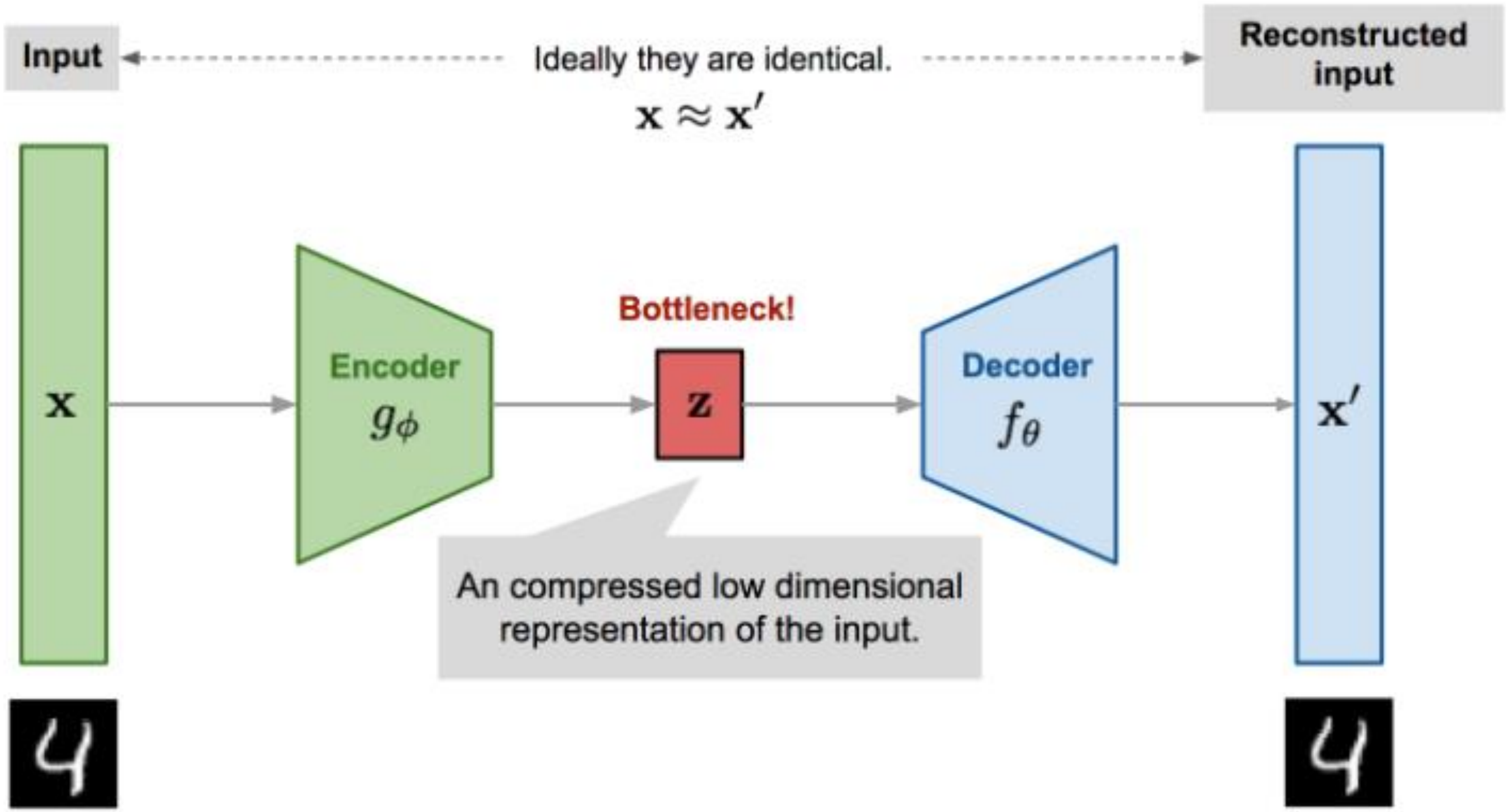
Metric	Value
RMSE(continuous)	21867
R^2(continuous)	0.46
Accuracy	0.97
Precision	0.96
Recall	0.97
F1	0.96



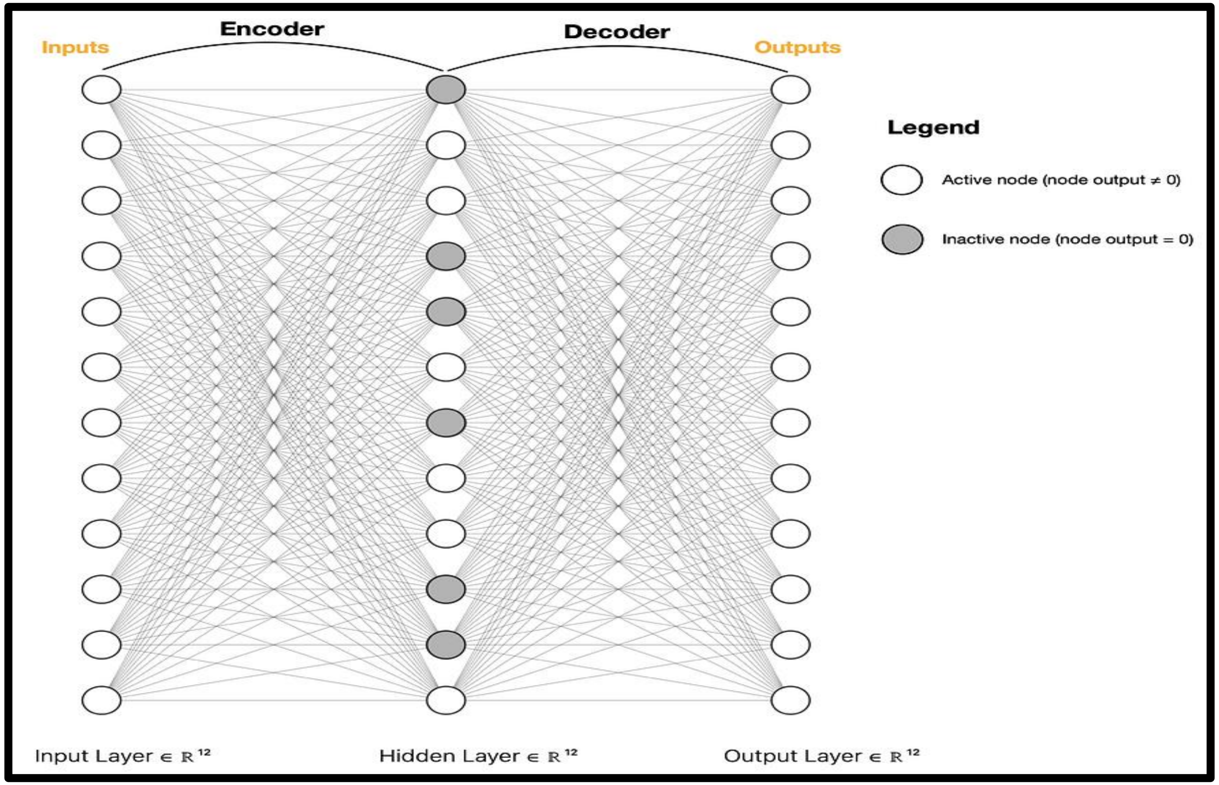
05

Autoencoders for Anomaly Detection

Autoencoder

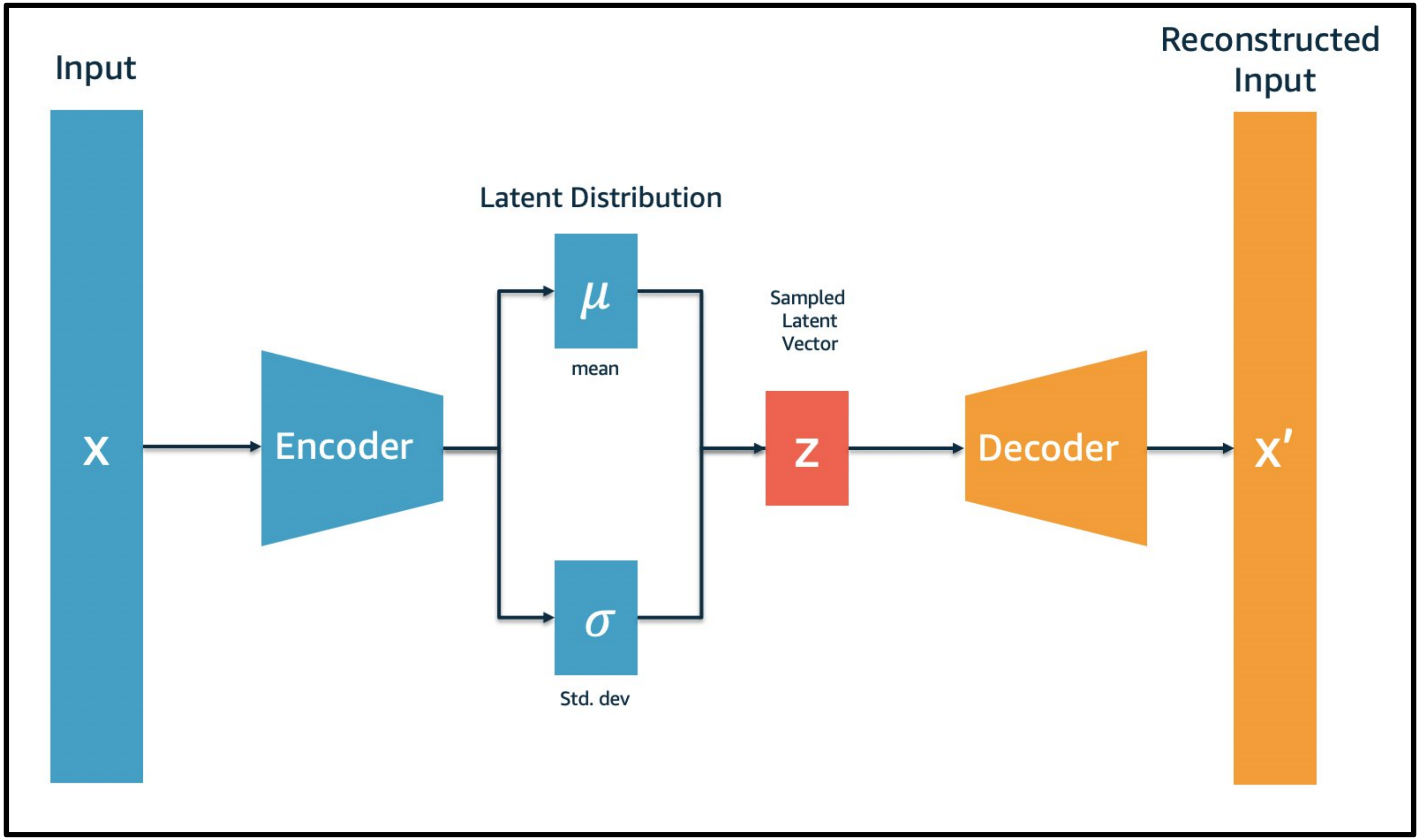


Sparse Autoencoder



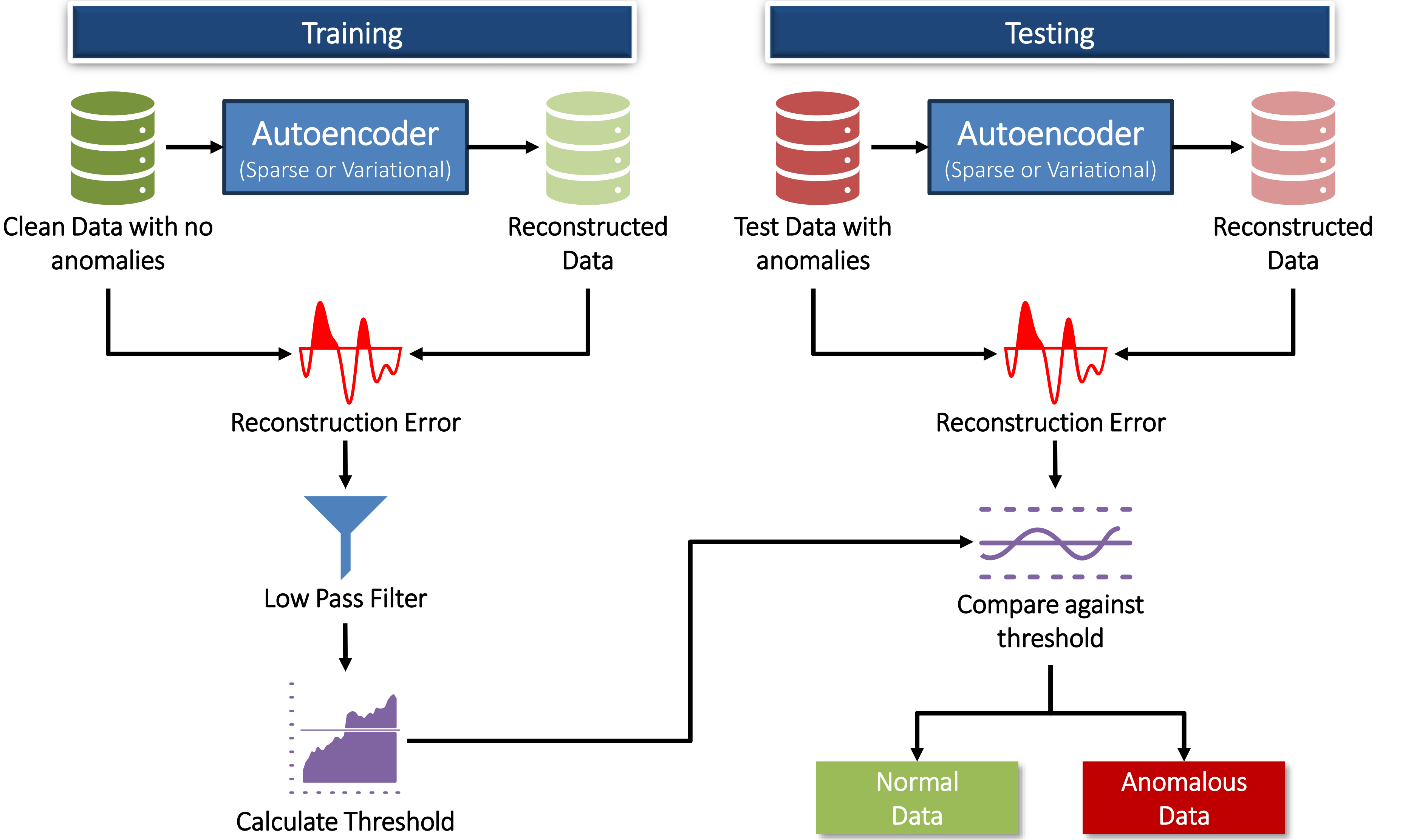
- Regularization item is added to the cost function for penalizing the weights
- Small number of neurons are activated and trained to encode and decode

Variational Autoencoder



- VAE introduces probabilistic elements into the encoding-decoding process

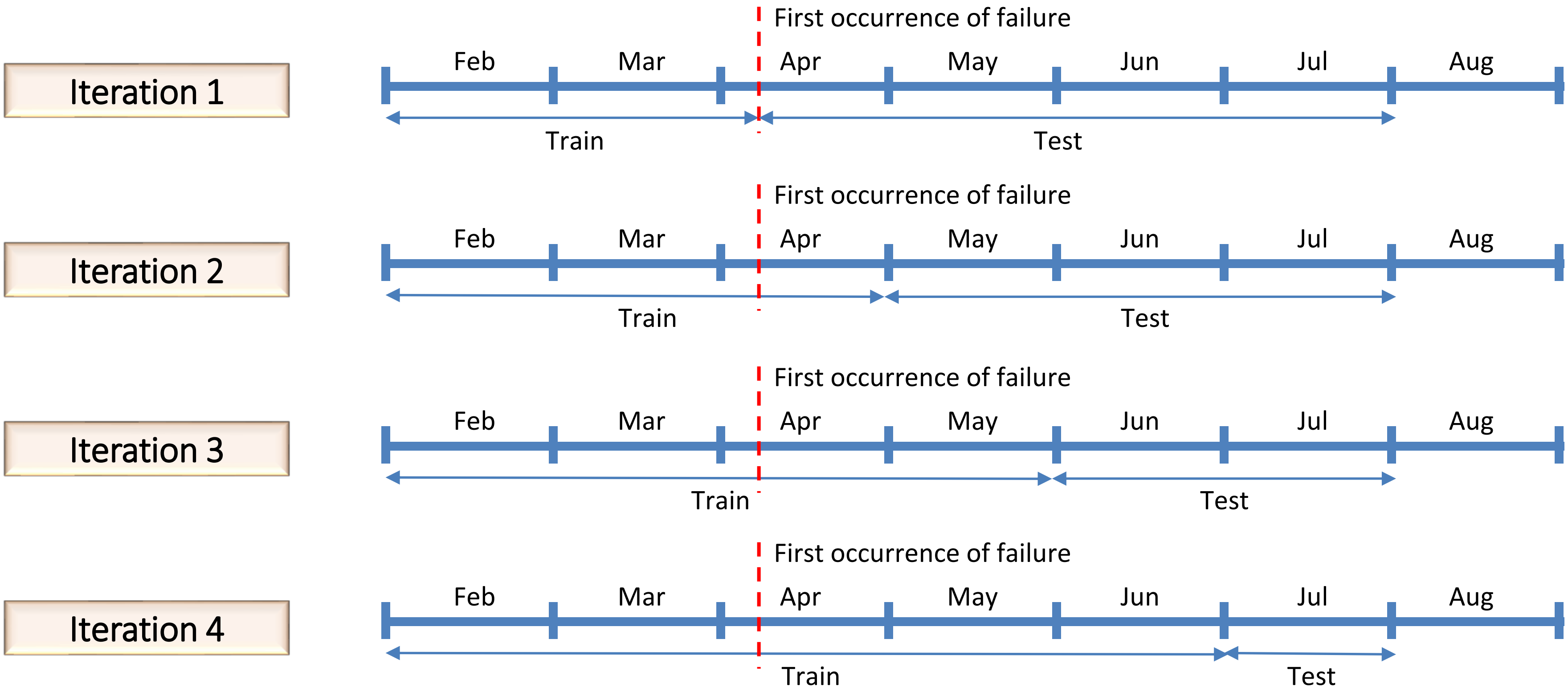
Anomaly Detection using Autoencoder



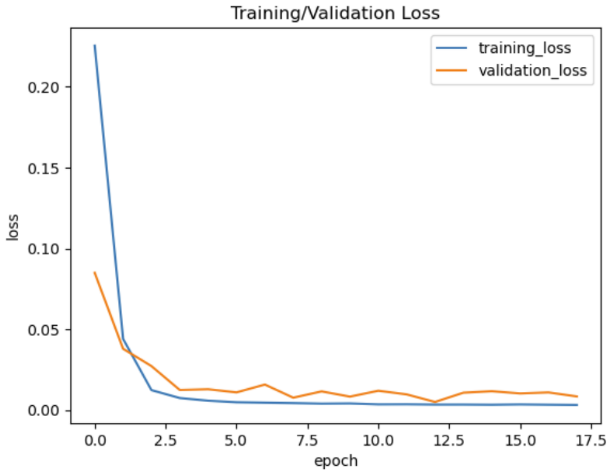
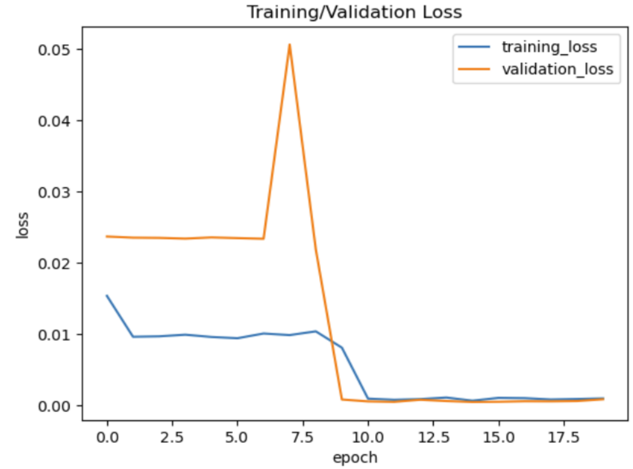
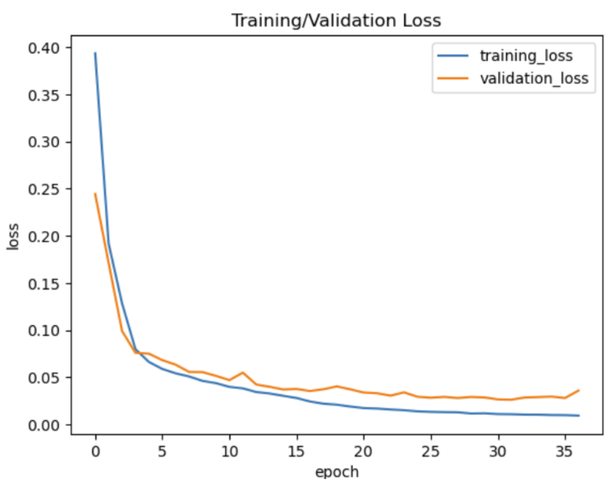
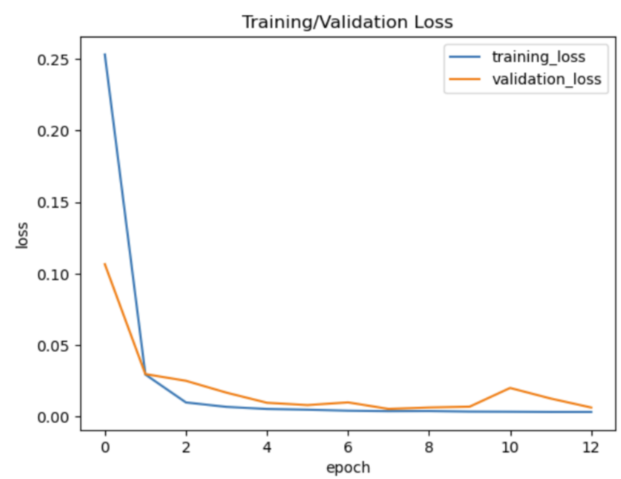
Note : *We flag the failure period as well as a period of 24 hours before the period as anomaly to ensure that we feed absolutely clean (normal) data to the autoencoder during training*

Online Learning

- Given there is a **continuous flow of data**, the model is expected to update itself continuously such that it is trained on all the normal data available till date
- In this particular online learning procedure, the model is **trained in a predetermined time window** and **predicts the anomalies in a subsequent test time window**.



Reconstruction Loss: SAE vs VAE & Analog vs Digital

	SAE	VAE
Digital		
Analog		

Overall, the digital signals, lead to better reconstruction loss. SAE leads to better reconstruction for digital signals while VAE leads to better reconstruction for analog signals

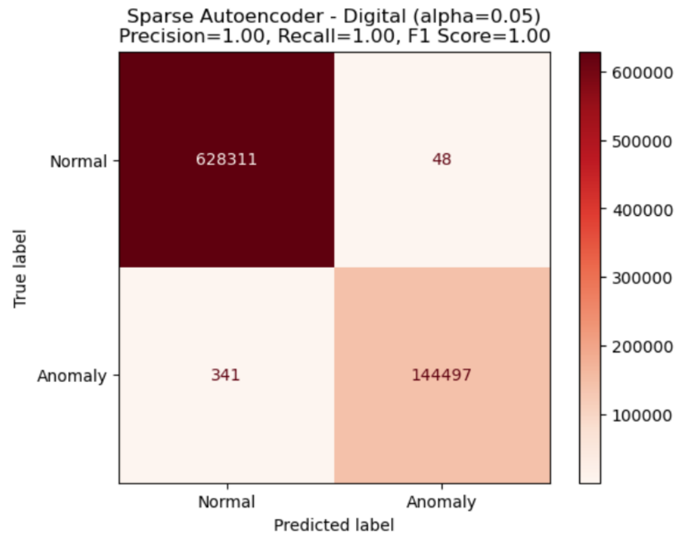
Results – Iteration 1

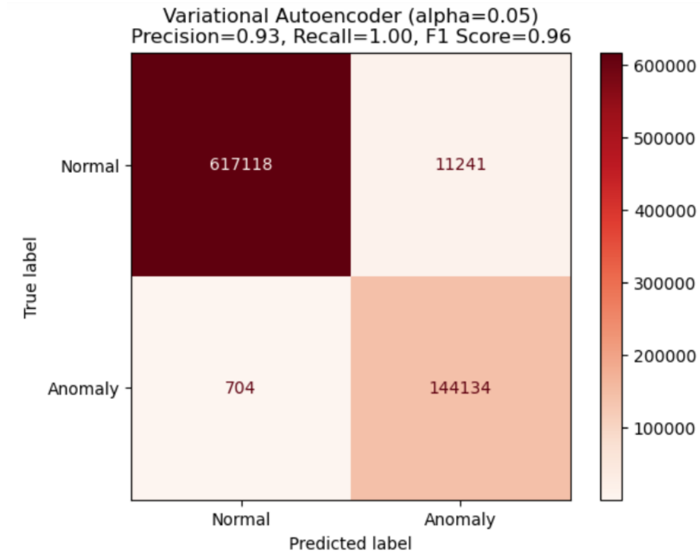
Train – Till the first occurrence of failure | Test – Remaining data till the end of July

SAE

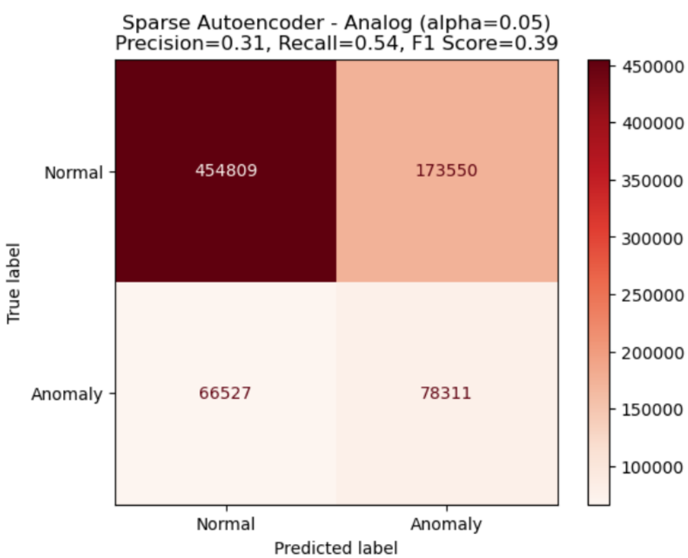
VAE

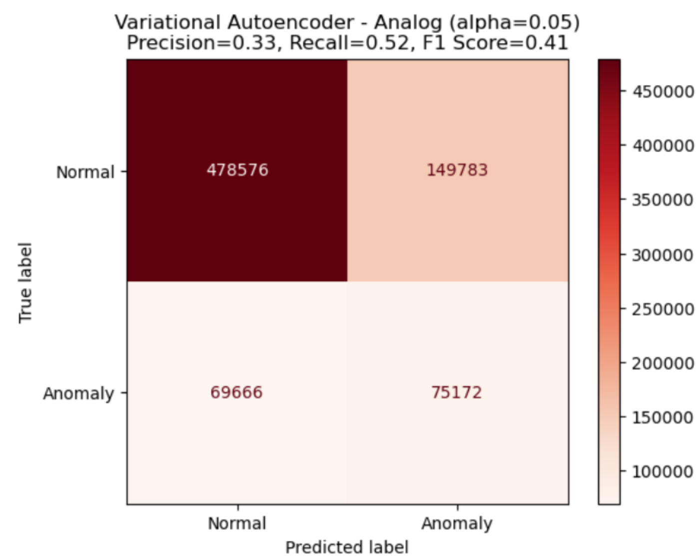
Digital





Analog





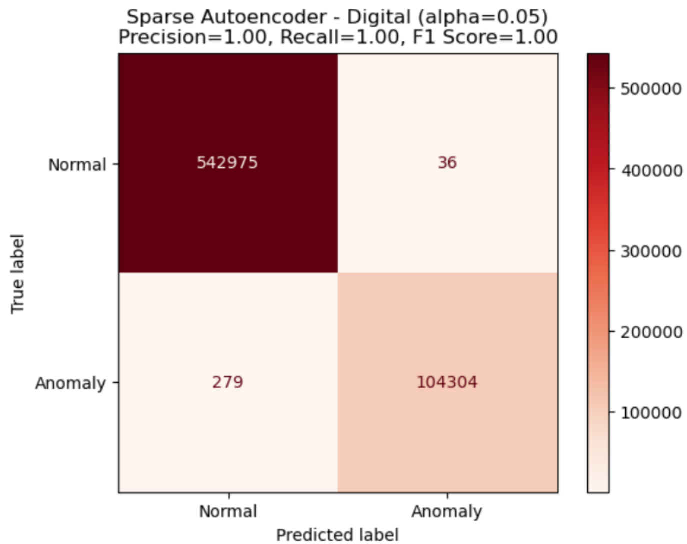
Results – Iteration 2

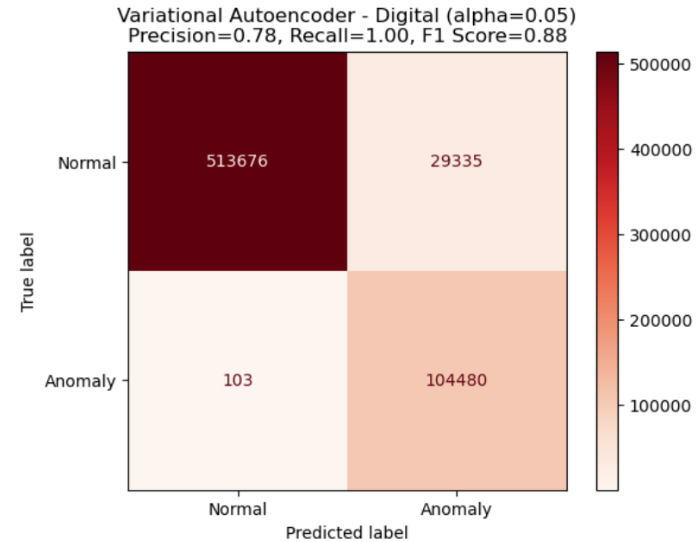
Train – Till April End | Test – Remaining data till the end of July

SAE

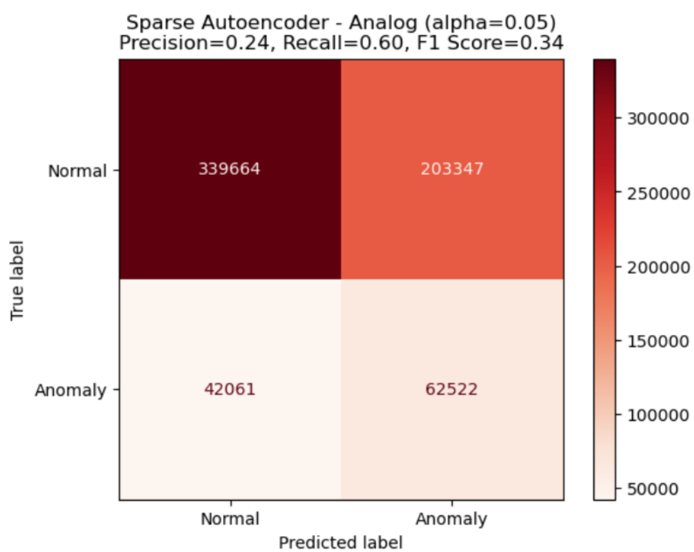
VAE

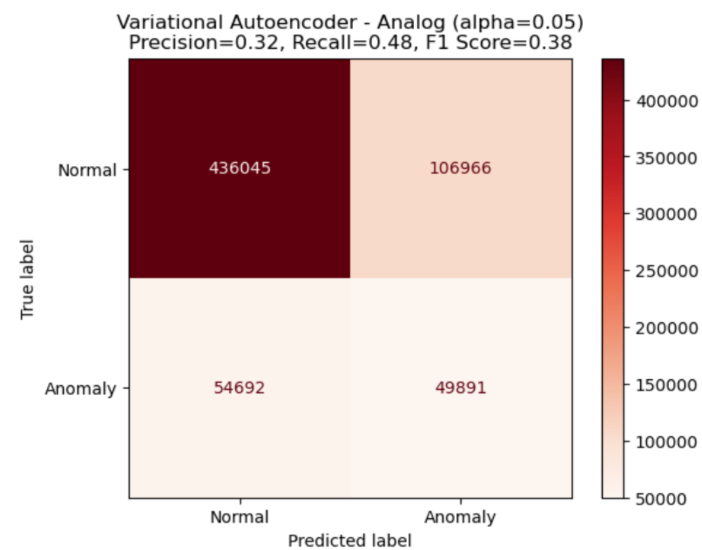
Digital





Analog





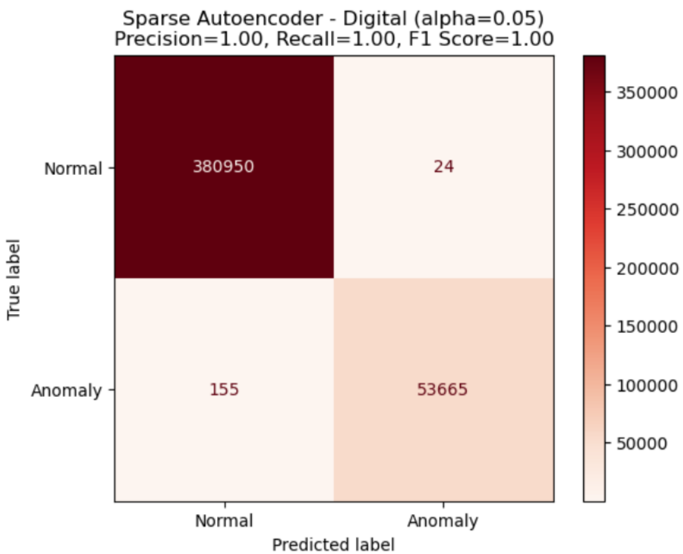
Results – Iteration 3

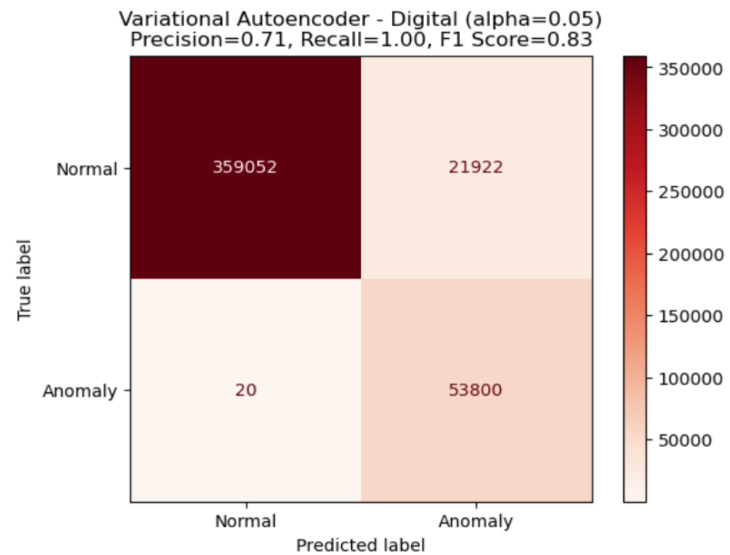
Train – Till May End | Test – Remaining data till the end of July

SAE

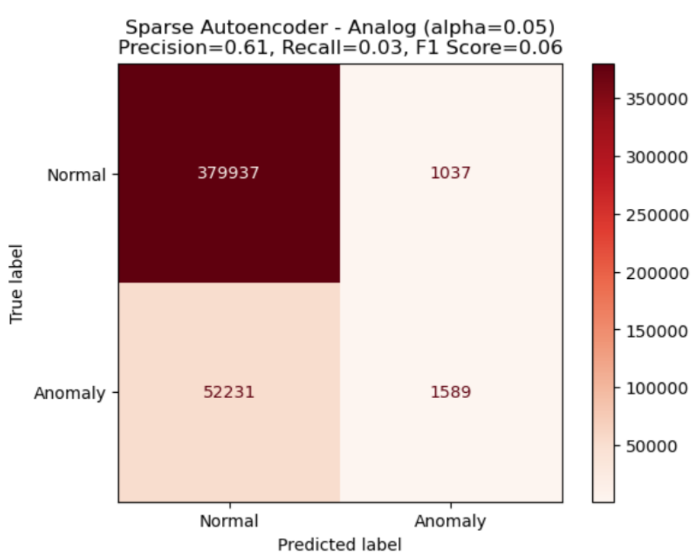
VAE

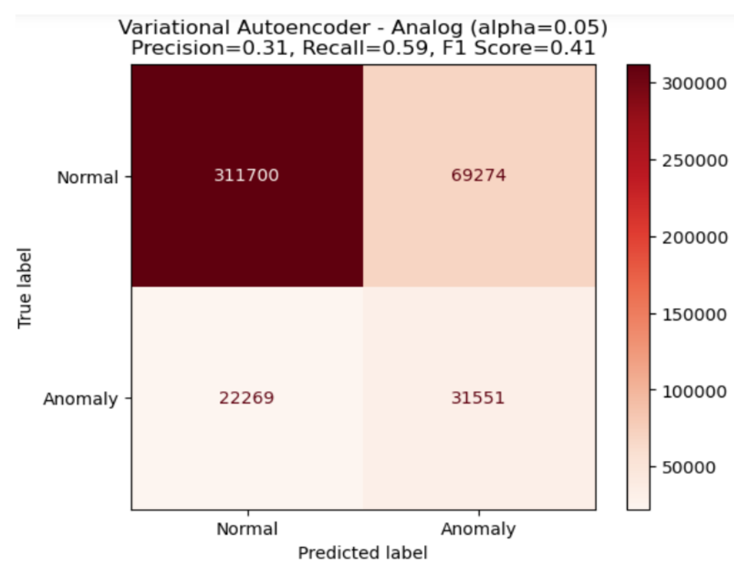
Digital





Analog





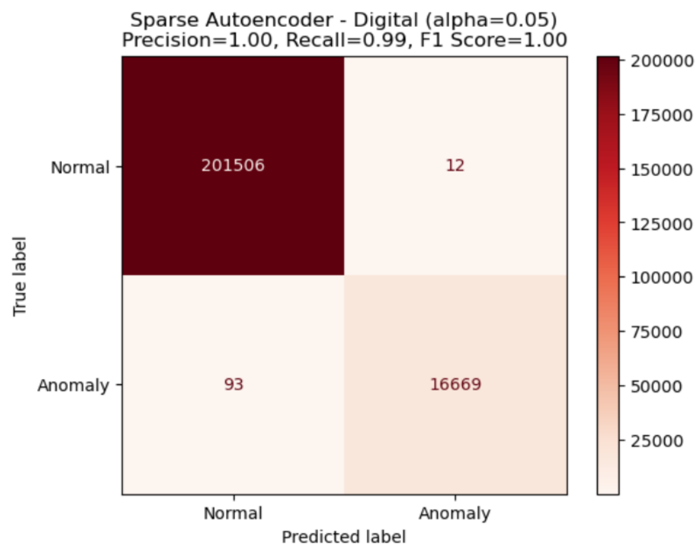
Results – Iteration 4

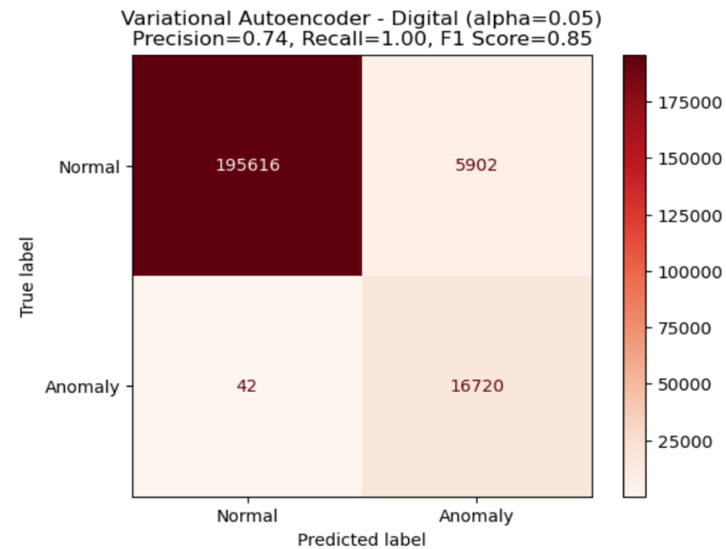
Train – Till June End | Test – Remaining data till the end of July

SAE

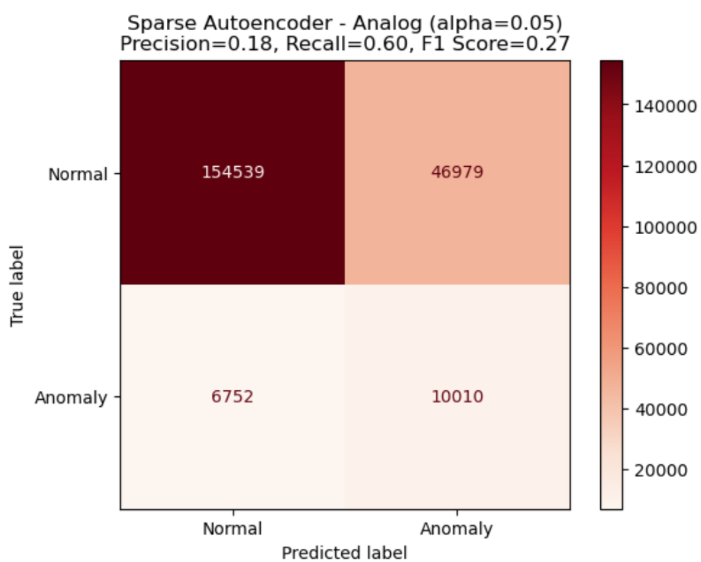
VAE

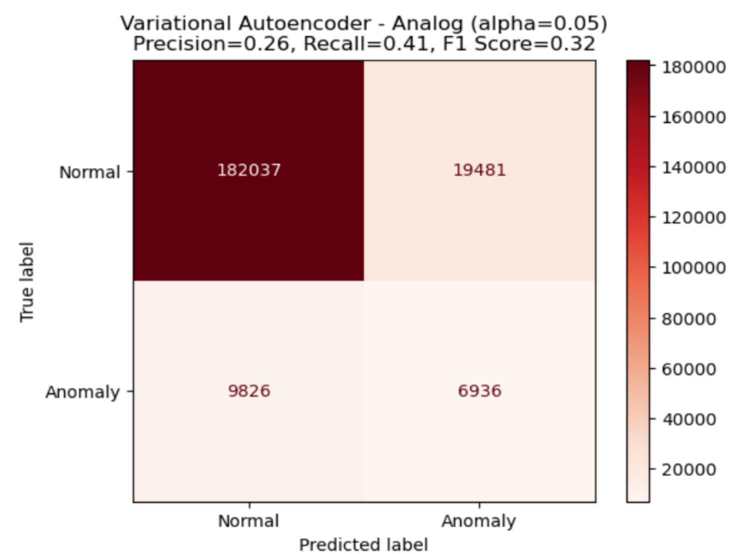
Digital



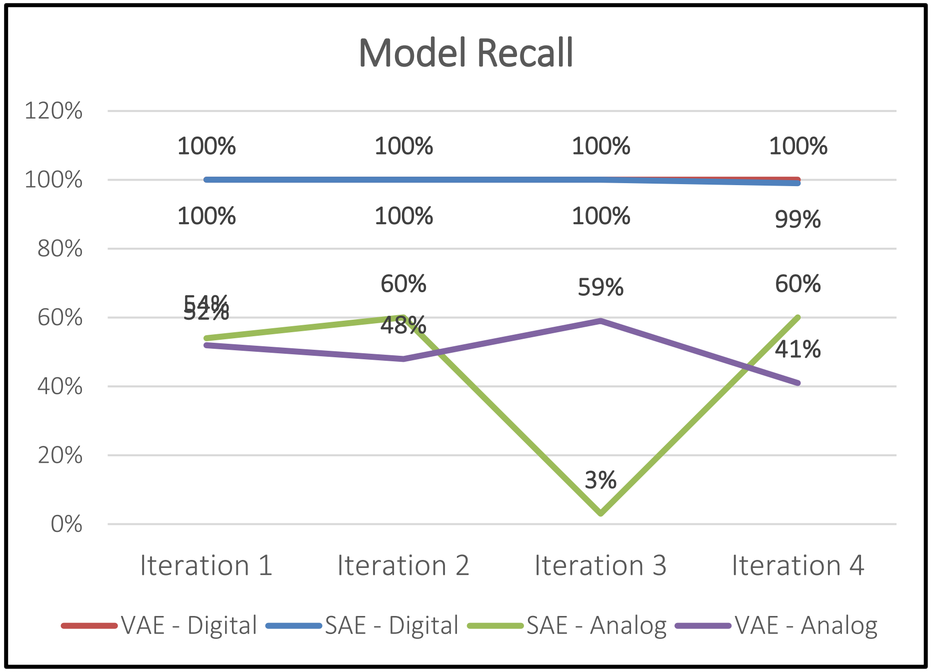


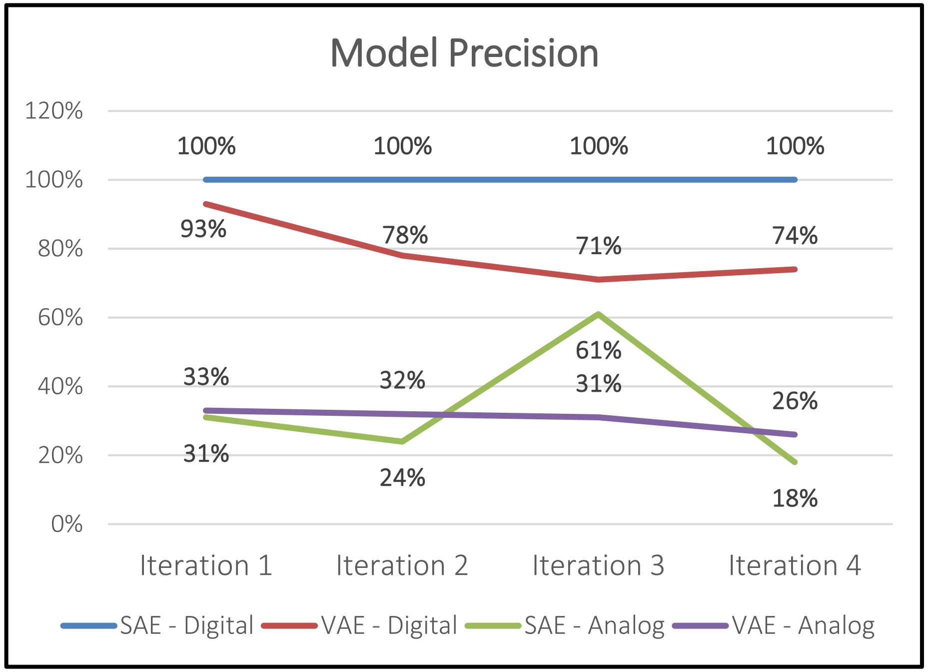
Analog





Results – Summary





- **Sparse Autoencoders work best** on this data as indicated by almost perfect precision and recall period after period
- **Data from digital sensors are more useful** in predicting anomalies in comparison to analog sensors

06
Next Steps

Next Steps

Additional pieces to explore:

Given a signal, how long is the unit going to run without breaking down?

- Possible techniques:
  - LSTM – initial steps
  - Transformers
    - Vanilla – longer term
    - Informer – temporal
    - SpaceTimeFormer – spatial and temporal
    - Autoformer – seasonal decomposition

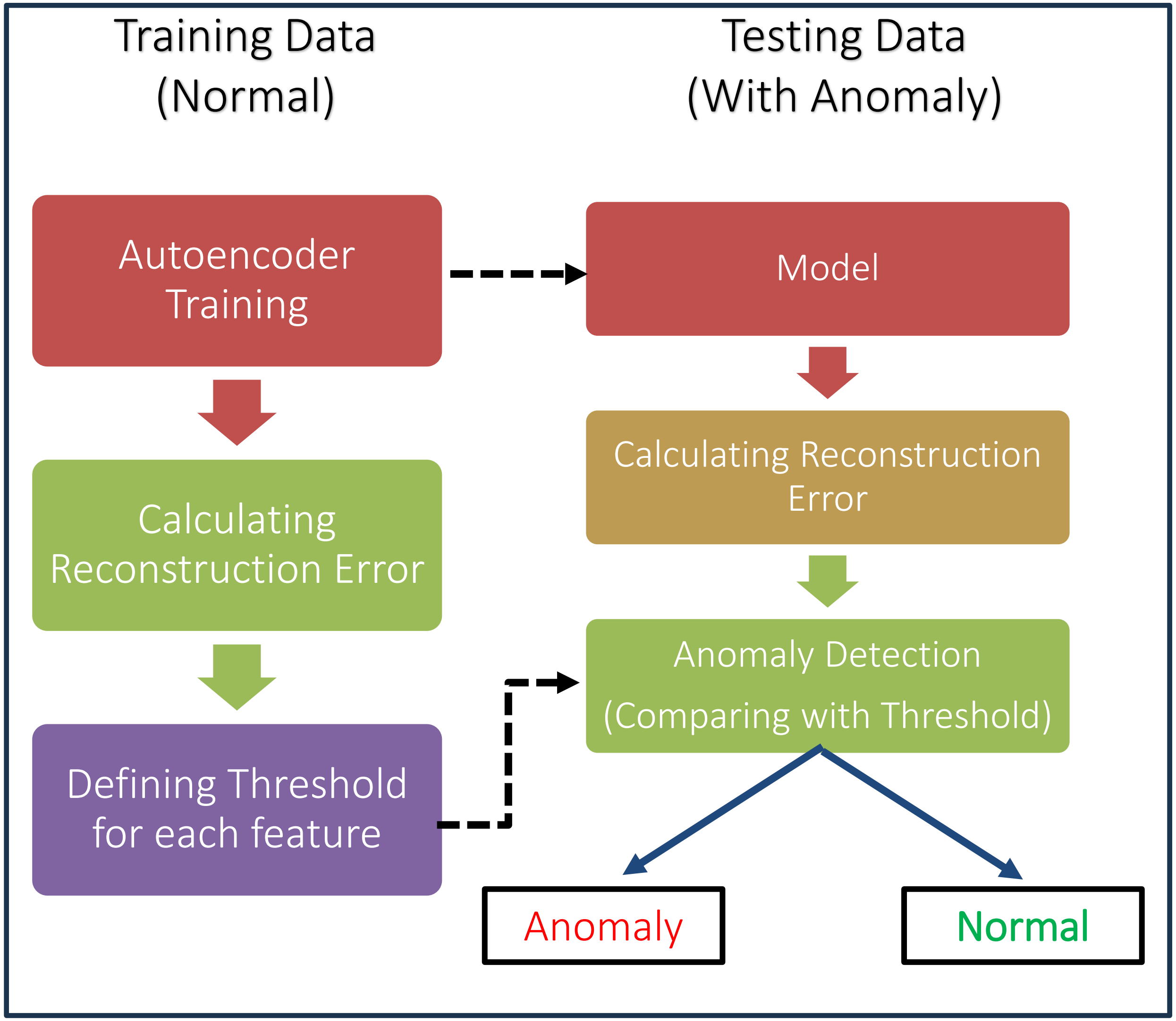
Causal inference between the sensors?

- Causal forest
- Survival Analysis

THANK YOU

Appendix

Anomaly Detection using Autoencoder



- *The model learns to encode and decode normal patterns, capturing the essential features*
- *Once the autoencoder is trained on normal data, it is used with the test data which includes both normal and abnormal data*
- *Instances with reconstruction errors surpassing the threshold are considered anomalies*