CSCI-B565 Data Mining, Fall 2023
Predictive Maintenance

Payal Gore[1] payagore@iu.edu
Alvan Dmello[2] admello@iu.edu
Abhishek Prasanna Walavalkar[3] awalaval@iu.edu
Luddy School of Informatics, Computing and Engineering
Indiana University Bloomington
Bloomington, IN 47405, USA

Abstract
In this study, our primary aim is to develop a predictive maintenance system for transportation assets, specifically trains and trams, using classification-based machine learning models. To achieve this goal, we utilise the Metro PT-3 dataset, which contains data from a metro compressor's Air Production Unit.

Our chosen method involves the use of classification-based machine learning models. These models analyse historical data and classify whether a failure is likely to occur within a specific time frame, enabling proactive maintenance strategies. The success of our project hinges on the system's ability to predict failures before the actual incident date. We measure success based on the lead time between the predicted failure date and the actual failure date.

Keywords: Classification, Predictive Maintenance, Alert Framework.

1. Problem and Data Description

1.1 Problem Statement

The problem at hand involves the development of a predictive maintenance system for transportation assets, specifically trains and trams, utilising classification-based machine learning models. The objective is to analyse historical data from the Metro PT-3 dataset, focusing on a metro compressor's Air Production Unit, to classify and predict potential failures within a specific timeframe. The success of the project is contingent upon the system's ability to forecast failures before the incident date, measured by the lead time between the predicted failure date and the actual occurrence.

1.2 Motivation

The motivation for this project stems from the critical need for businesses to minimise machine downtime, a key factor in sustaining optimal operations and preventing revenue loss. Efficient maintenance strategies, particularly predictive maintenance, are crucial for anticipating and scheduling necessary repairs or equipment replacements. Leveraging predictive maintenance provides valuable insights, allowing companies to proactively manage machinery, ensuring smooth operations, and reducing unplanned disruptions for cost savings.

1.3 Dataset

The dataset in the project was gathered from a metro train's operational context, specifically focusing on readings from various sensors on a compressor's Air Production Unit (APU). These readings include pressure, temperature, motor current, and air intake valve data. The dataset, consisting of 15 attributes, is a multivariate time series with 15,169,480 instances. It was collected between February and August 2020 to support the development of predictive maintenance.The goal is to use machine learning methods to address challenges in the industry related to failure predictions.

1.4 Data Description

The dataset comprises 15,169,480 data points collected at a rate of 1Hz from February to August 2020, describing the operational status of a compressor through 15 features. These features are derived from both analogue (1-7) and digital (8-15) sensors. The key attributes are:

1. TP2 (bar): Pressure on the compressor.
2. TP3 (bar): Pressure at the pneumatic panel.
3. H1 (bar): Pressure from cyclonic separator filter discharge.
4. DV pressure (bar): Pressure drop during air dryer discharge.
5. Reservoirs (bar): Downstream pressure, ideally close to TP3.
6. Motor Current (A): Current indicating different operational states.
7. Oil Temperature (ºC): Temperature of the compressor's oil.
8. COMP: Electrical signal of air intake valve; active in off or offloaded states.
9. DV electric: Electrical signal controlling compressor outlet valve; active under load.
10. TOWERS: Electrical signal defining active tower for air drying.
11. MPG: Electrical signal for starting compressor under load.
12. LPS: Electrical signal detecting pressure drop below 7 bars.
13. Pressure Switch: Electrical signal detecting air-drying tower discharge.
14. Oil Level: Electrical signal detecting low oil level.
15. Caudal Impulse: Electrical signal counting pulse outputs from APU to reservoirs.

These attributes provide detailed information about the compressor's operating conditions and can be utilised for various purposes, including predictive maintenance.

The dataset is unlabeled, but failure reports provided by the company are available. Here is a summary of the failure information:

Nr.	Start Time	End Time	Failure	Severity	Report
#1	4/18/2020 0:00	4/18/2020 23:59	Air Leak	High stress	
#2	5/29/2020 23:30	5/30/2020 6:00	Air Leak	High stress	Maintenance on 30Apr at 12:00
#3	6/5/2020 10:00	6/7/2020 14:30	Air Leak	High stress	Maintenance on 8Jun at 16:00
#4	7/15/2020 14:30	7/15/2020 19:00	Air Leak	High stress	Maintenance on 16Jul at 00:00

Table 1: Failure Report Log

2. Data Preprocessing & Exploratory Data Analysis

2.1 Feature selection and Exploratory Data Analysis

To determine the essential features for our analysis, we initially utilised a heatmap to visualise the correlations among different features. The heatmap served as a valuable tool in gaining insights into the interrelationships and dependencies among the various variables in our dataset. This exploratory step was instrumental in identifying key features that exhibit significant correlations, providing a foundation for further analysis and feature selection.

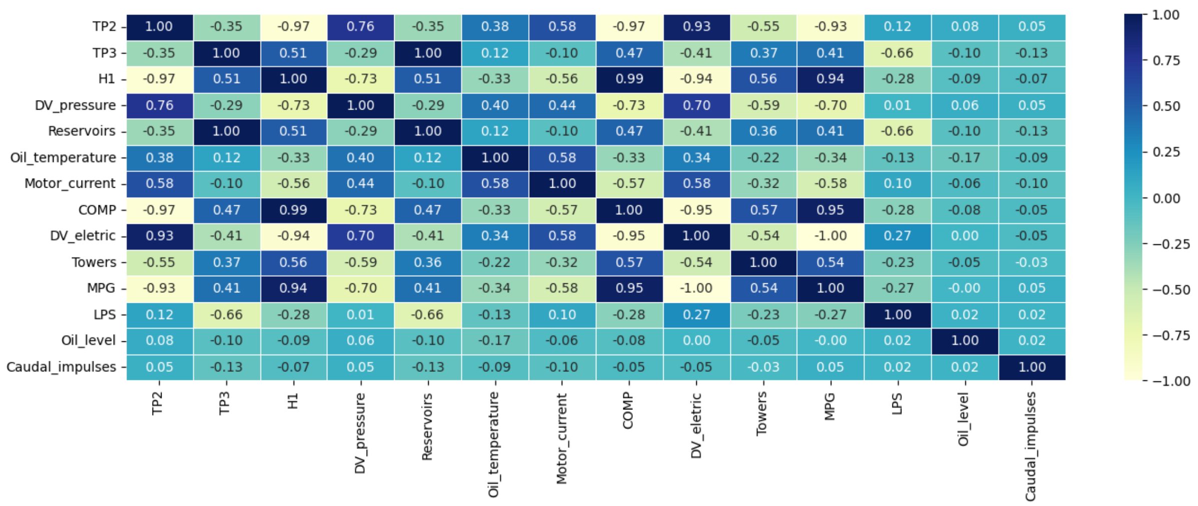


Figure 1: Heatmap of Features

From the above plot we can observe that the following pairs of attributes are highly correlated:

- Reservoirs and TP3
- TP2 and DV_electric
- H1 and COMP
- COMP and MPG
- DV_electric and COMP
- TP2 and COMP
- TP2 and H1
- DV_electric and MPG

Certain features, namely caudal_impulses, oil_level, and pressure_switch, exhibit a static behaviour over time leading up to the failure report date, as depicted in the accompanying graphs.

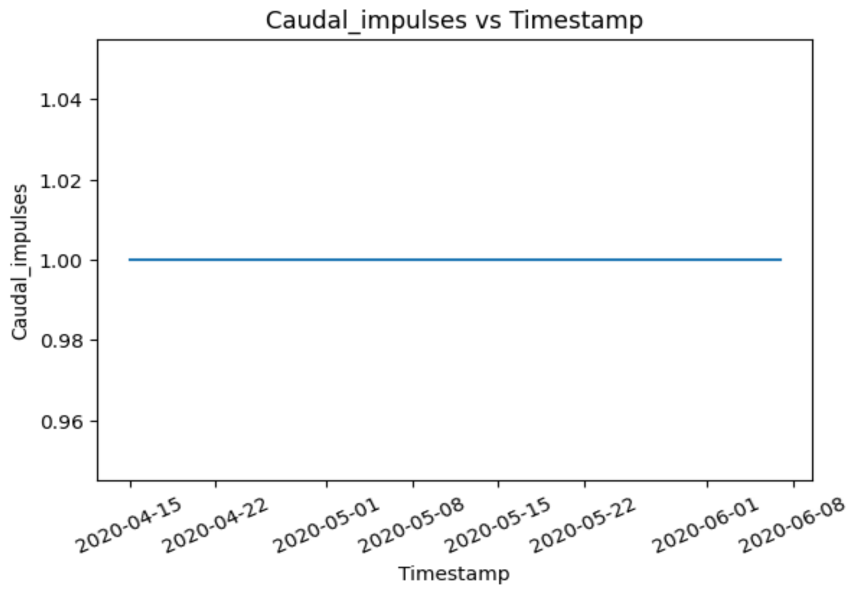


Figure 2: Caudal Impulses vs Timestamp

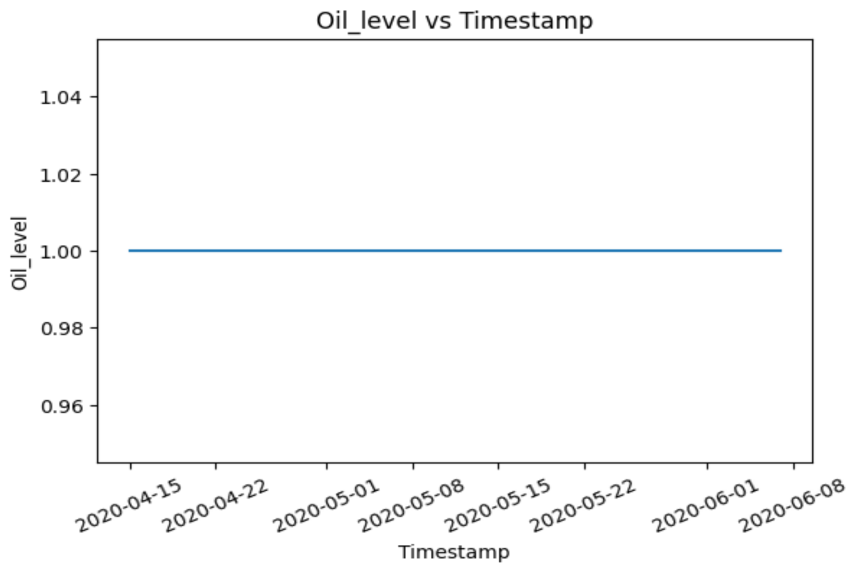


Figure 3: Oil_level vs Timestamp

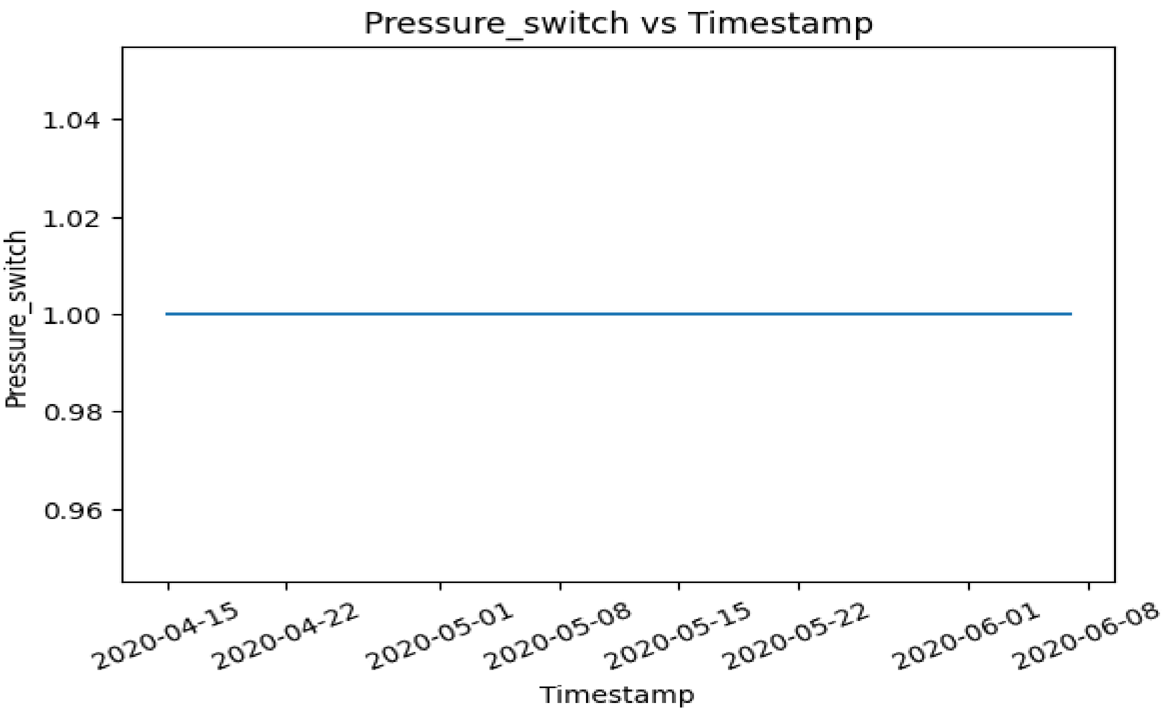


Figure 4: Pressure_switch vs Timestamp

Consequently, due to the absence of discernible changes in their respective time series, we have decided to exclude these features from our analysis. Removal of these features is deemed appropriate as their static nature does not contribute meaningful information to our understanding of the system dynamics and failure patterns.

Certain features exhibit pronounced and sudden fluctuations in their line graphs, especially near reported failure dates. These abrupt changes suggest a potential association with errors or anomalies. The accompanying graphs visually highlight these variations, emphasising their relevance to system failures. Further exploration of these features is warranted, as they may serve as valuable indicators for predictive maintenance strategies.

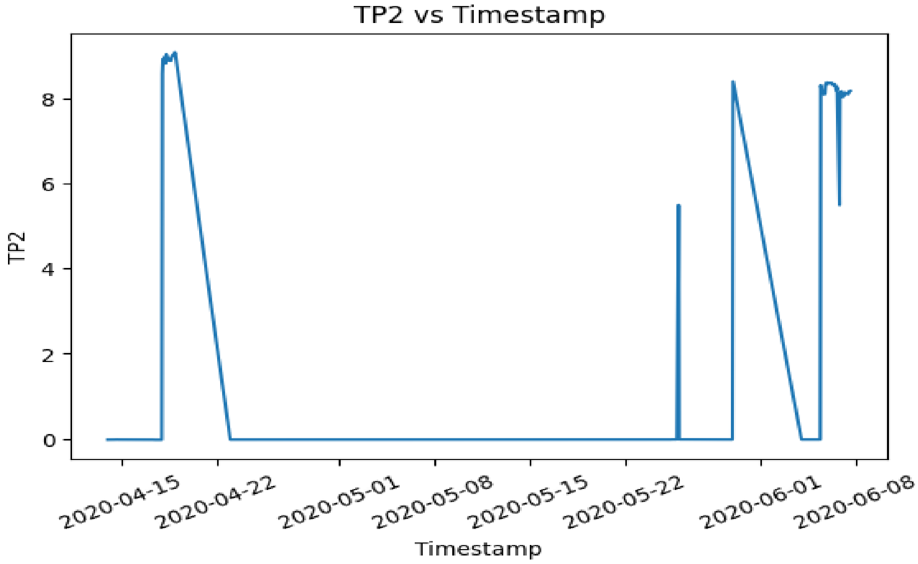


Figure 5: TP2 vs Timestamp

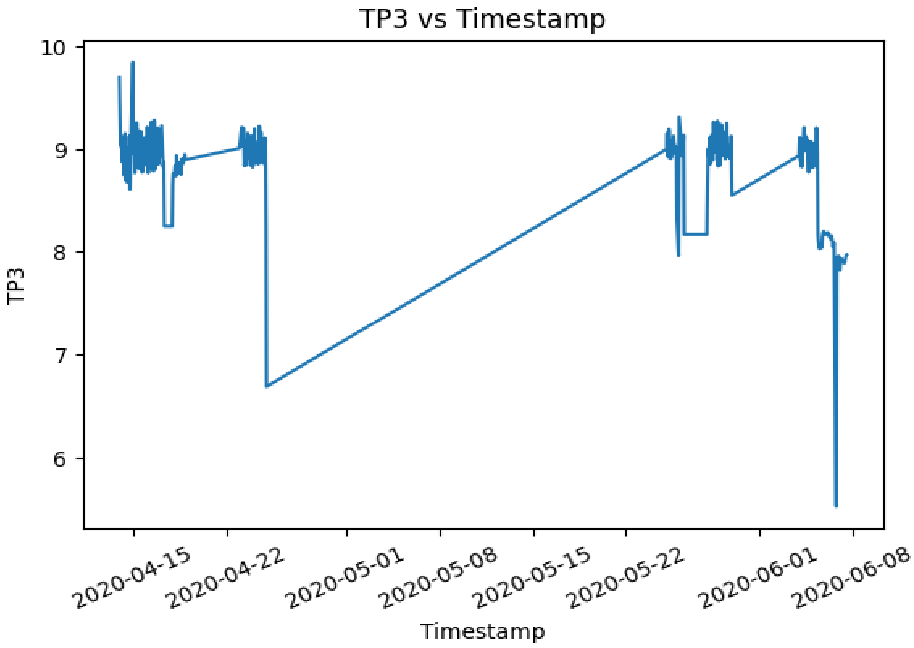


Figure 5: TP3 vs Timestamp

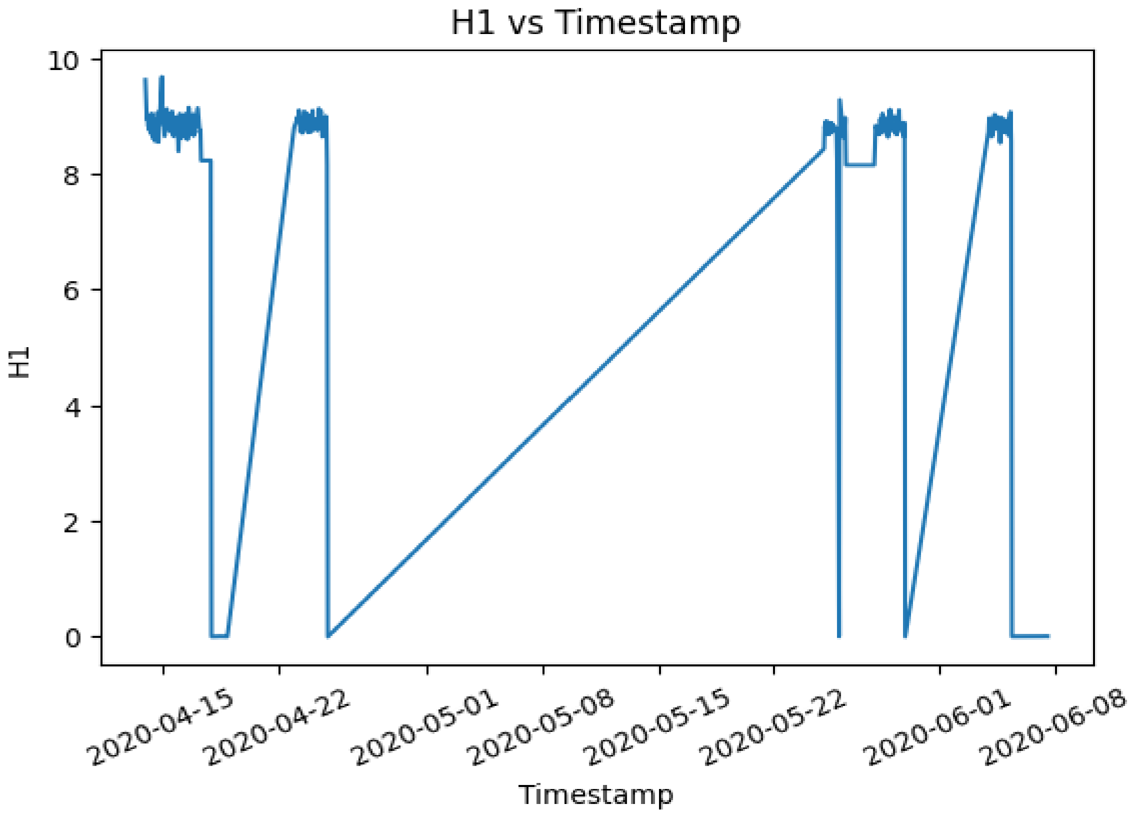


Figure 6: H1 vs Timestamp

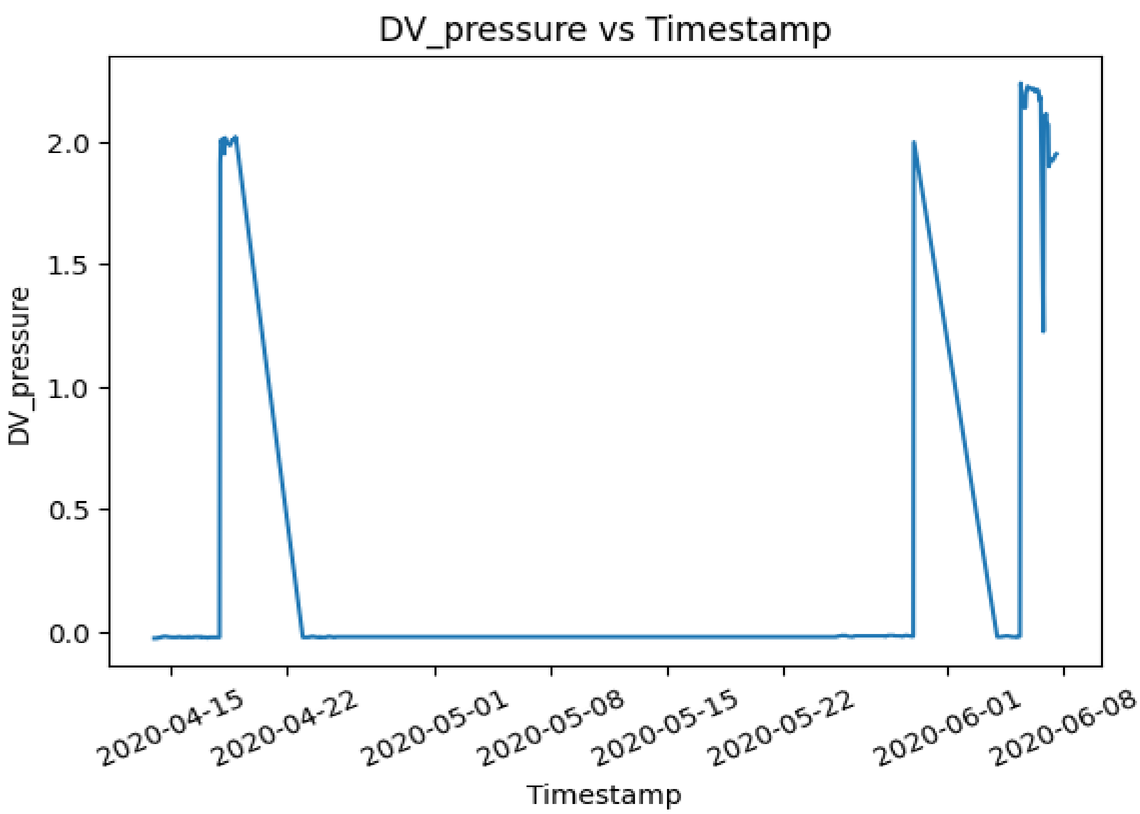


Figure 7: DV_Pressure vs Timestamp

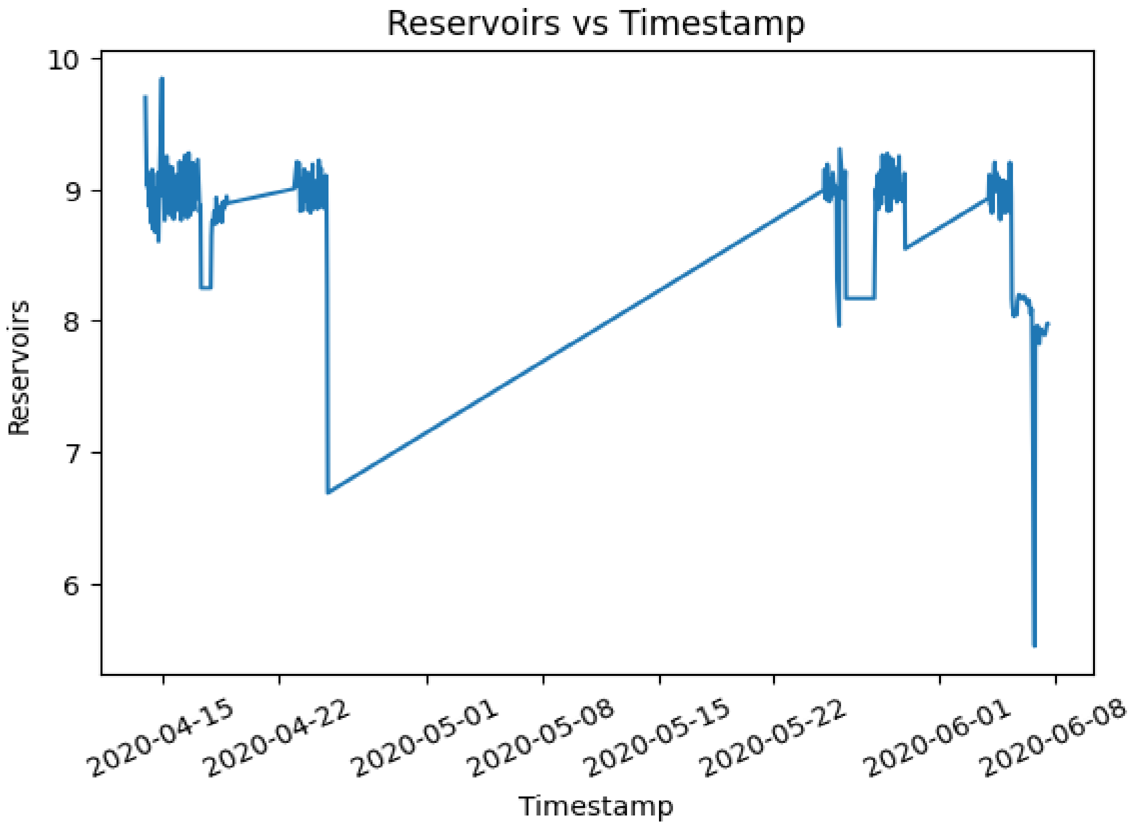


Figure 8: Reservoirs vs Timestamp

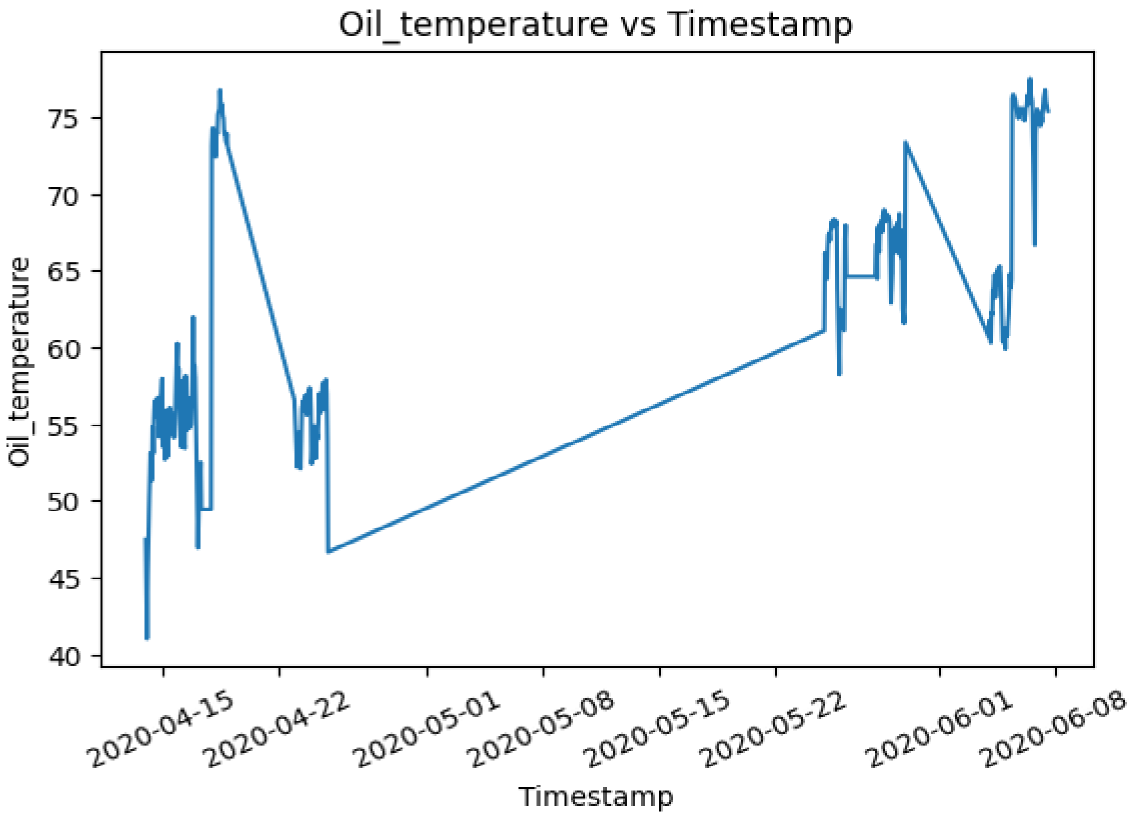


Figure 9: Oil_temperature vs Timestamp

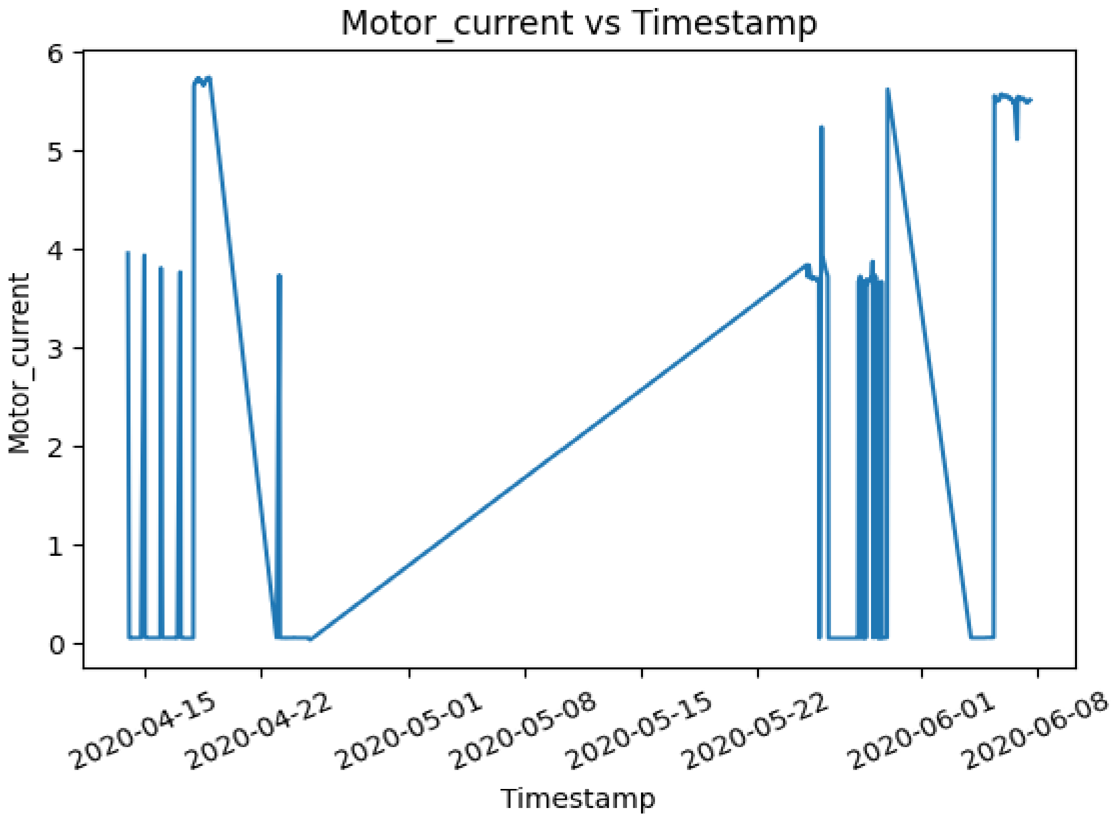


Figure 5: Motor_current vs Timestamp

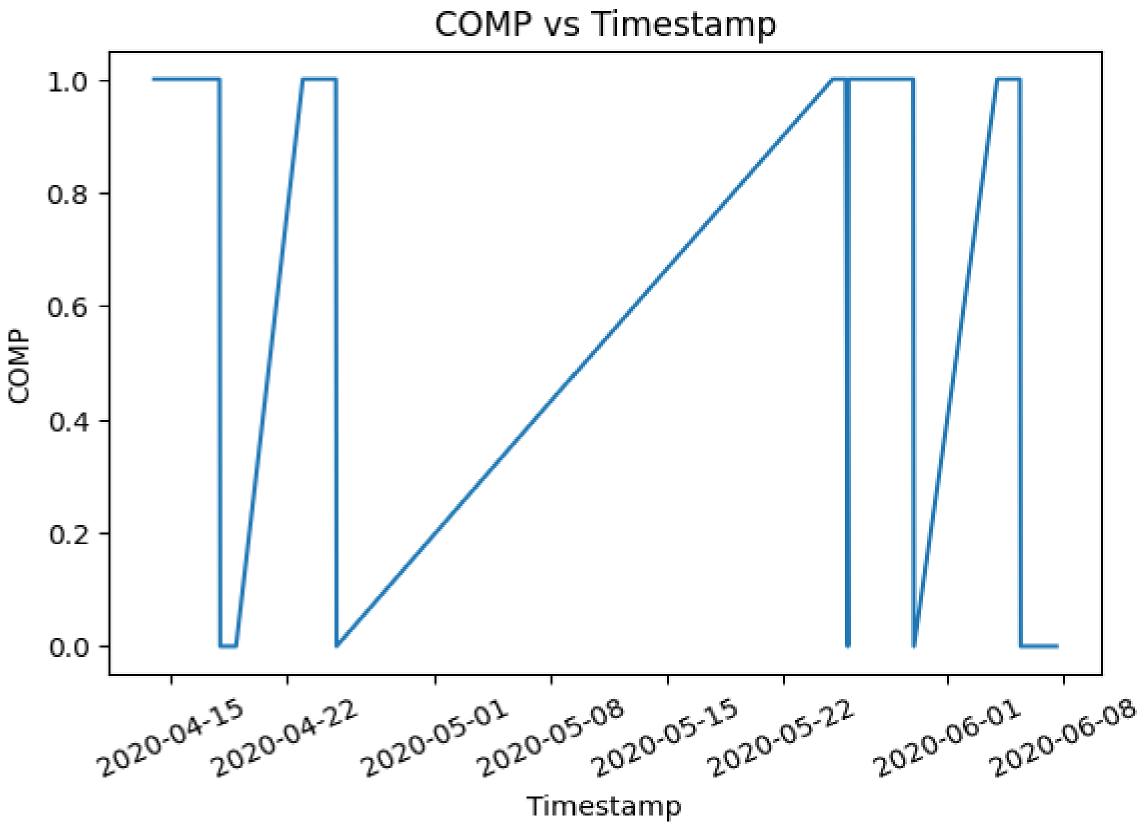


Figure 11: COMP vs Timestamp

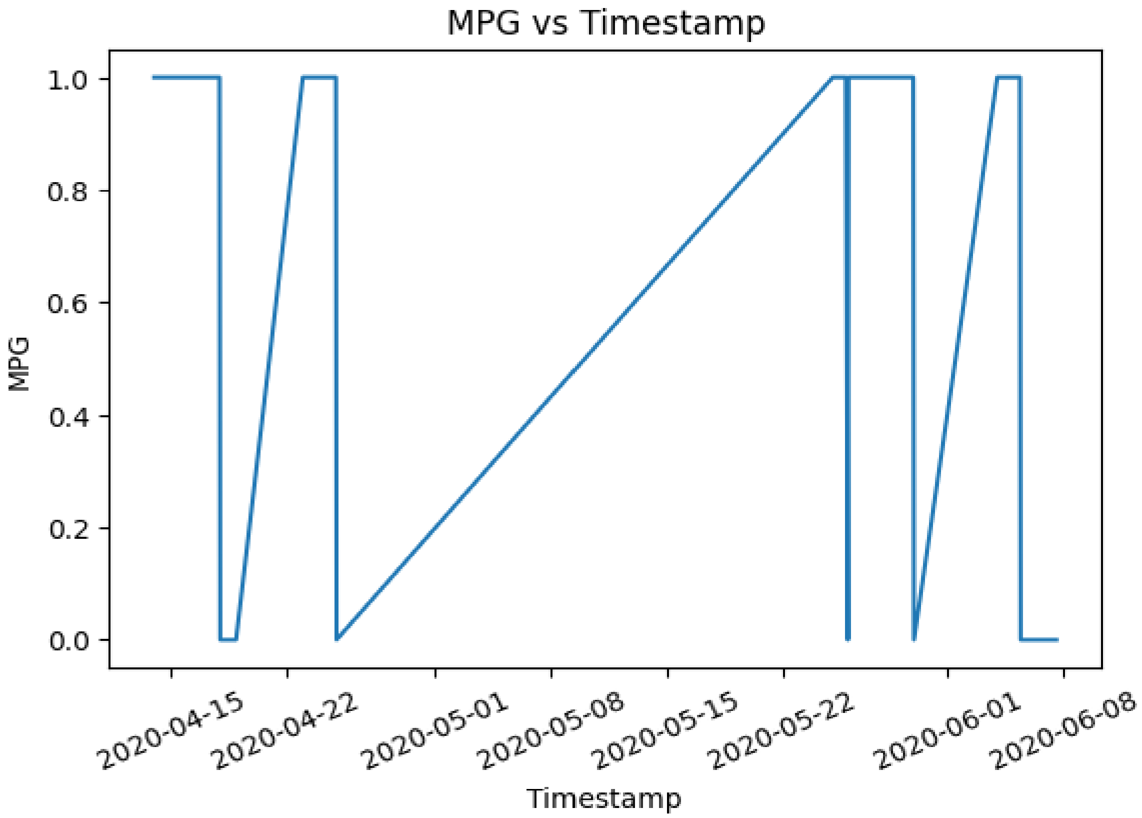


Figure 12: MPG vs Timestamp

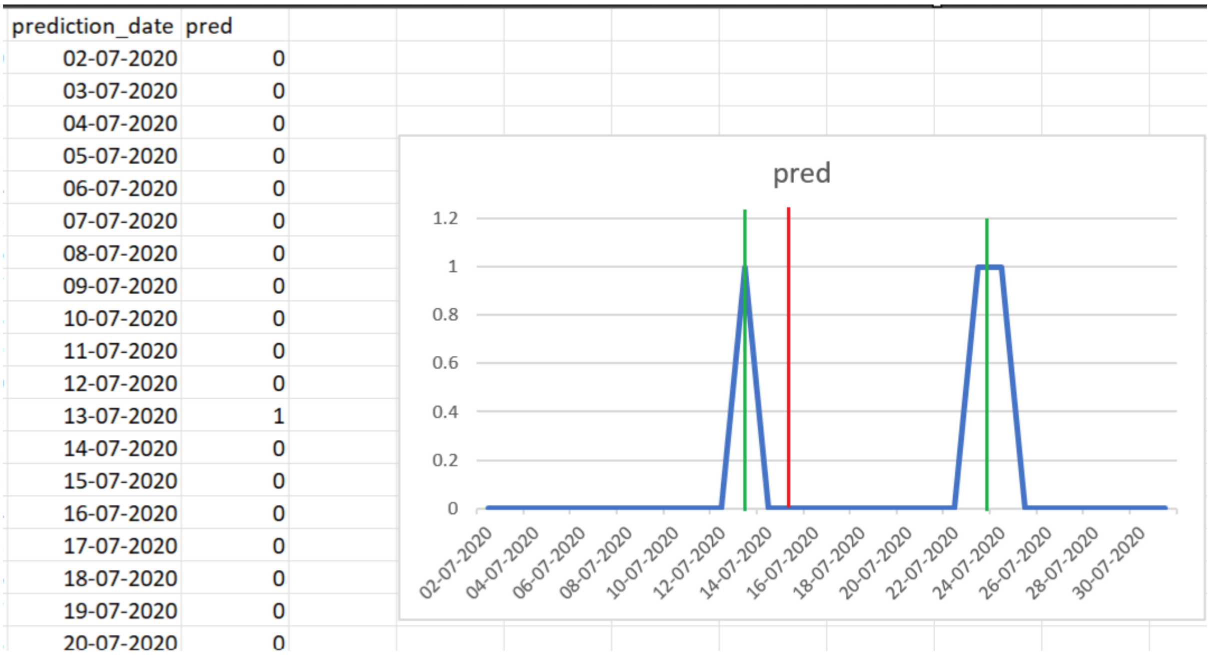
From the analysis above, we are choosing the following features to build the model on:

- TP2
- TP3
- H1
- DV Pressure
- Reservoirs
- Oil Temperature

In the absence of an in-depth understanding of the system, identifying truly influential variables becomes challenging. To achieve efficient and accurate feature selection, the insights and expertise of a Subject Matter Expert are essential.

3. Experiments and Results

For the month of July, we have observed the following
July Month Alerts



prediction_date	pred
02-07-2020	0
03-07-2020	0
04-07-2020	0
05-07-2020	0
06-07-2020	0
07-07-2020	0
08-07-2020	0
09-07-2020	0
10-07-2020	0
11-07-2020	0
12-07-2020	0
13-07-2020	1
14-07-2020	0
15-07-2020	0
16-07-2020	0
17-07-2020	0
18-07-2020	0
19-07-2020	0
20-07-2020	0

Figure 13: Alerts vs Timestamp

In the visual representation, green lines denote alerts generated by the model, while red lines correspond to the actual failure dates.

4. Deployment and Maintenance

Once operational, the model is designed to execute daily on the past 7 days' machine vitals data. To ensure optimal performance, regular retraining is scheduled at predefined intervals. Additional insights into failure reports are to be incorporated into the logs through contributions from both remote monitoring engineers (RME) and field service engineers (FSE).

The alerts generated by the models are subjected to analysis by subject matter experts for a minimum period of 3 months before being considered reliable for practical use. If any performance issues arise during this monitoring period, essential modifications, including adjustments to feature selection, failure report logs, alert logic, etc., should be implemented promptly. This iterative process aims to refine and enhance the model's effectiveness over time, ensuring its robustness in predicting and identifying potential maintenance needs.

5. Discussion

In contemporary settings, machine sensors serve a dual purpose by not only capturing real-time machine vitals data but also acquiring error logs. These error logs typically fall into categories such as fatal, error, and warning. The integration of this error data, alongside the existing machine vitals data, could provide valuable insights. This data could potentially help us recognize the component in which the failure occurs by minimising our range of focus and helping us to take further actions faster. It would further be extremely helpful if we also had error data directly provided to us from different components of the machine.

5. References

Davari, N., Veloso, B., Ribeiro, R.P., Pereira, P.M., Gama, J.: Predictive maintenance based on anomaly detection using deep learning for air production unit in the railway industry. In: 2021 IEEE 8th International Conference on Data Science and Advanced Analytics (DSAA). pp. 1–10. IEEE (2021) (DOI:10.1109/DSAA53316.2021.9564181)

Veloso, B., Ribeiro, R.P., Pereira, P.M., Gama, J.: The MetroPT dataset for predictive maintenance. Scientific Data 9, no. 1 (2022): 764. (DOI: 10.1038/s41597-022-01877-3)

Davari, N.; Veloso, B.; Costa, G.d.A.; Pereira, P.M.; Ribeiro, R.P.; Gama, J. A Survey on Data-Driven Predictive Maintenance for the Railway Industry. Sensors 2021, 21, 5739. https://doi.org/10.3390/s21175739

Zhang, W.; Yang, D.; Wang, H. Data-Driven Methods for Predictive Maintenance of Industrial Equipment: A Survey. IEEE Syst. J. 2019, 13, 2213–2227.

Sohaib, M.; Mushtaq, S.; Uddin, J. Deep Learning for Data-Driven Predictive Maintenance. In Vision, Sensing and Analytics: Integrative Approaches; Springer: Berlin, Germany, 2021; pp. 71–95.

Zhang, W., Yang, D., Wang, H.: Data-driven methods for predictive maintenance of industrial equipment: a survey. IEEE Syst. J. (2019). https://doi.org/10.1109/JSYST.2019.2905565