**Predictive Maintenance for Air Production Unit**

**1. Introduction**

This project focuses mainly on the critical parts of public transport, especially Metro vehicles, and aims to points the challenges of detecting faults during regular operation. Prompt and efficient fault detection can prevent interruptions and minimize damage. Currently, faults in the Air Production Unit (APU), a critical component in Metro vehicles, often go undetected based on predefined thresholds, leading to unexpected maintenance and service disruptions.

The project's primarly focusing to implement predictive maintenance techniques to reduce operational problems, unforeseen stops, and maintenance time. By using the MetroPT dataset, which contains data on anomalies from the company's maintenance reports, the project aims to benchmark various Classification and clustering algorithms for anomaly detection based on continuous sensor data.

The ultimate goal of predictive maintenance is to develop a maintenance plan when a failure is identified. This involves predicting failures, identifying the type of failure and the specific component affected, and estimating the remaining useful life of the components.

**Attribute Information:**

1. TP2 (bar) – The compressor pressure.
2. TP3 (bar) – The pneumatic panel pressure.
3. H1 (bar) – The pressure resulting from the pressure drop.
4. DV pressure (bar) – The pressure drop measurement during air dryer emptying.
5. Reservoirs (bar) – The subsequent pressure in the reservoirs.
6. Motor Current (A) – The current of one phase in the three-phase motor.
7. Oil Temperature (ºC) – The temperature of the compressor oil.
8. COMP - The electrical indication of the air inlet valve on the compressor.
9. DV electric – The electrical signal controlling the compressor outlet valve.
10. TOWERS – The electrical signal determining the air-drying tower responsible.
11. MPG – The electrical signal responsible for starting the loaded compressor by activating the absorption valve.
12. LPS - The electrical signal that identifies and triggers action when the pressure descends below 7 bars.
13. Pressure Switch - The electrical signal detecting the air-drying towers' release.
14. Oil Level – The electrical signal detecting the compressor's oil level.
15. Caudal Impulse – The electrical signal that records the pulse outputs generated by the total air amount.

**2. Data preparation and feature engineering**

For the project, exploration and analysis of the variables were collected from the Air Production Unit (APU). The analogical sensors, dispersion and shape of variable distribution were studied using the Correlation relations, skewness, and histograms.

For the categorical variables, the frequency of each variable activation was analyzed. The goal was to identify their behavior. Some inconsistencies were noted between COMP and MPG sensor values in APU, which could be due to sensor malfunctions.

The correlation between variables is also a good measure to understand how the APU system works. The heat maps that represent the correlations between variables on APU.

Based on the Correlation Heatmap analysis:

* The attributes COMP, MPG, DV Electric, TP2, and H1 exhibit strong correlations with each other.
* The Oil Temperature sensor shows a notable correlation with the TP3 sensor.
* Variables Towers, Oil Level, and Caudal Impulses do not demonstrate strong correlations with other sensors.
* The LPS sensor demonstrates its highest correlation with the TP3 sensor.

A diagram of a temperature

Description automatically generated with medium confidence

**3. Cluster Analysis**

Cluster analysis is a key aspect of this project, specifically applied to the unsupervised learning section. This technique aims to group a set of objects in such a way that objects in the same group (called a cluster) are more similar to each other than to those in other groups (clusters). It's a main task of exploratory data mining, and a common technique for statistical data analysis used in many fields.

In this project, the K-means clustering algorithm is employed. K-means is a centroid-based or partitioning method. It partitions the input data into K clusters where each observation belongs to the cluster with the nearest mean. Elbow method has used to determine the optimal number of clusters for the K-means algorithm:

A graph with a line

Description automatically generated**Elbow Method:** This method plots the variance as a function of the number of clusters. The optimal number of clusters is the point where adding another cluster doesn't improve the total variance significantly. This point resembles an "elbow" in the plot, which is how the method got its name.

After determining the optimal number of clusters, K-means is applied to the standardized data, and the clustering results are visualized using Principal Component Analysis (PCA). PCA reduces the dimensionality of the data while preserving as much variance as possible, which makes it possible to visualize the high-dimensional data in a 2D plot.

In this project, both the Elbow Method suggested the optimal number of clusters to be 3. This means that the air compressor system can be in three distinct working modes.

Once the optimal number of clusters was determined, K-Means was applied to the data, and each cluster was characterized based on the centroid of that cluster. These centroids were used to understand the characteristics of the working modes of the air compressor system.

The cluster analysis helped to understand the operational behavior of the air compressor system and to identify potential anomalies. This knowledge is crucial for predictive maintenance, as it helps to detect malfunctioning states of the system and to predict potential failures.

A chart with dots and numbers

Description automatically generated with medium confidence

**Cluster 1** signifies an operational state of the compressor where it is functioning under a full load. This state is characterized by a motor current near 8A which means peak pressure values on both the compressor (TP2) and the pneumatic panel (TP3)

**Cluster 2** depicts the compressor operating in an off-loaded condition, as suggested by the motor current close to 6A. This state experiences no pressure on the compressor (TP2) but maximum pressure on the pneumatic panel (TP3). It also displays high values of H1 and the greatest Oil Temperature, indicative of the compressor's off-loaded state. The COMP, MPG, and TOWERS are active in this state, whereas DV Electric is not.

**Cluster 3** represents the compressor in an inactive or off state, as inferred from the lowest motor current around 4A. This condition shows the absence of pressure on the compressor (TP2) and minimal pressure on the pneumatic panel (TP3).

**6. Conclusion**

In conclusion, the primary objective of our project was to predict the failure of Auxiliary Power Units (APUs) installed on trains, using their sensor data. Through our comprehensive study, we widely used clustering techniques to identify standard and outlier operation modes of the APU. Our results demonstrated the efficiency of our methodologies. The clustering techniques revealed operational modes for the APUs. These findings underscore the potential of our approach to enhance the predictability of APU failures, providing a foundation for future advancements in this field.

**7. References**

Davari,Narjes, Veloso,Bruno, Ribeiro ,Rita, and Gama,Joao. (2023). MetroPT-3 Dataset. UCI Machine Learning Repository. <https://doi.org/10.24432/C5VW3R>.