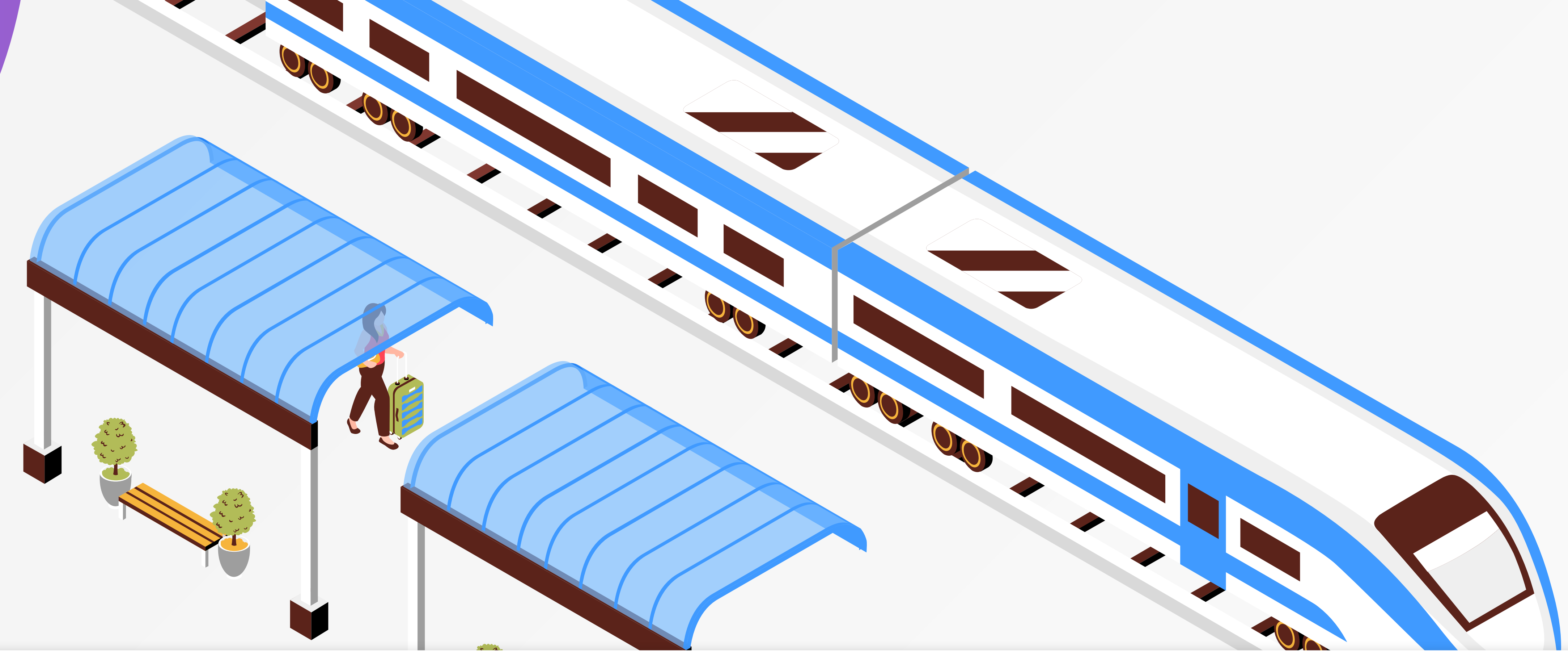
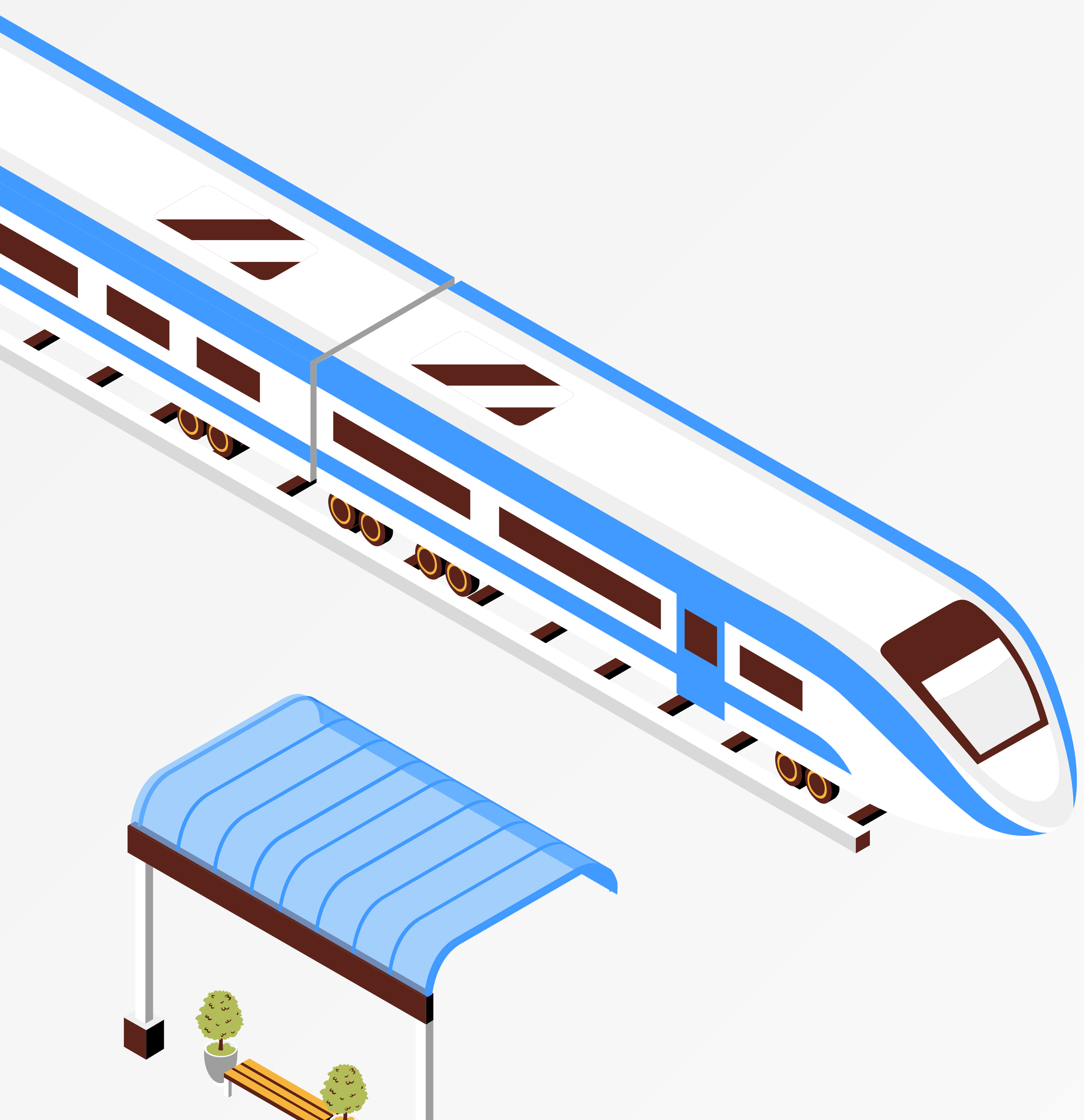


Metro PT3 Predictive Maintenance

Fundamentals of Data mining: Dr Seif Eldawlatly

Malak Gaballa - Masa Tantawy - Moustafa El Mahdy

Table of Contents

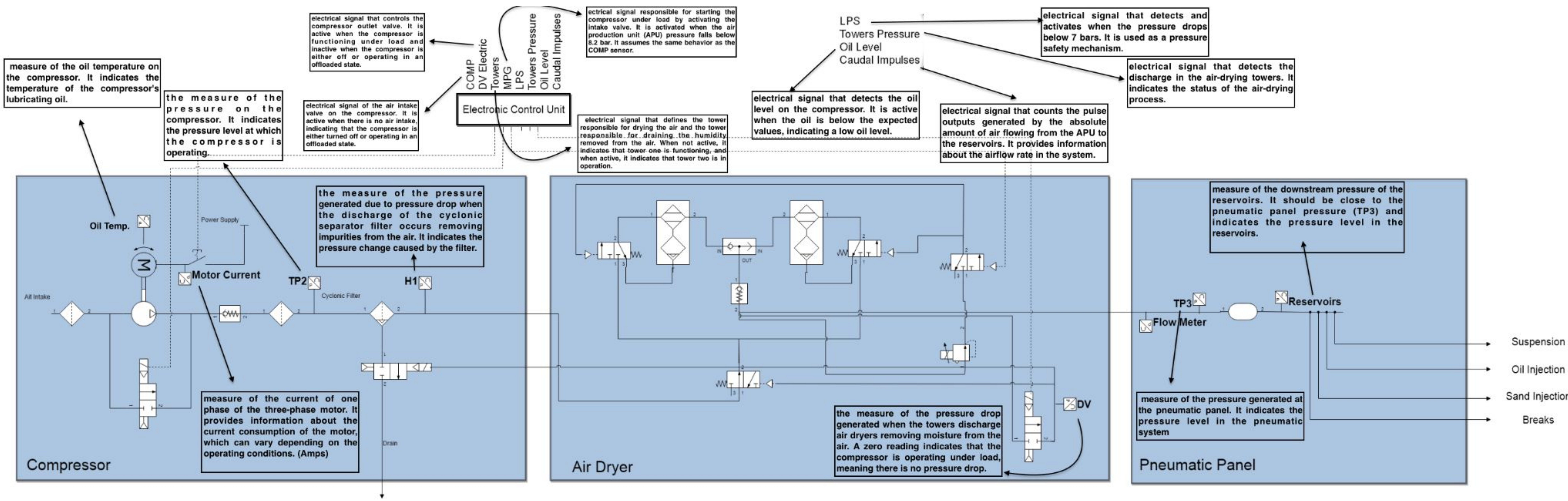


The Dataset

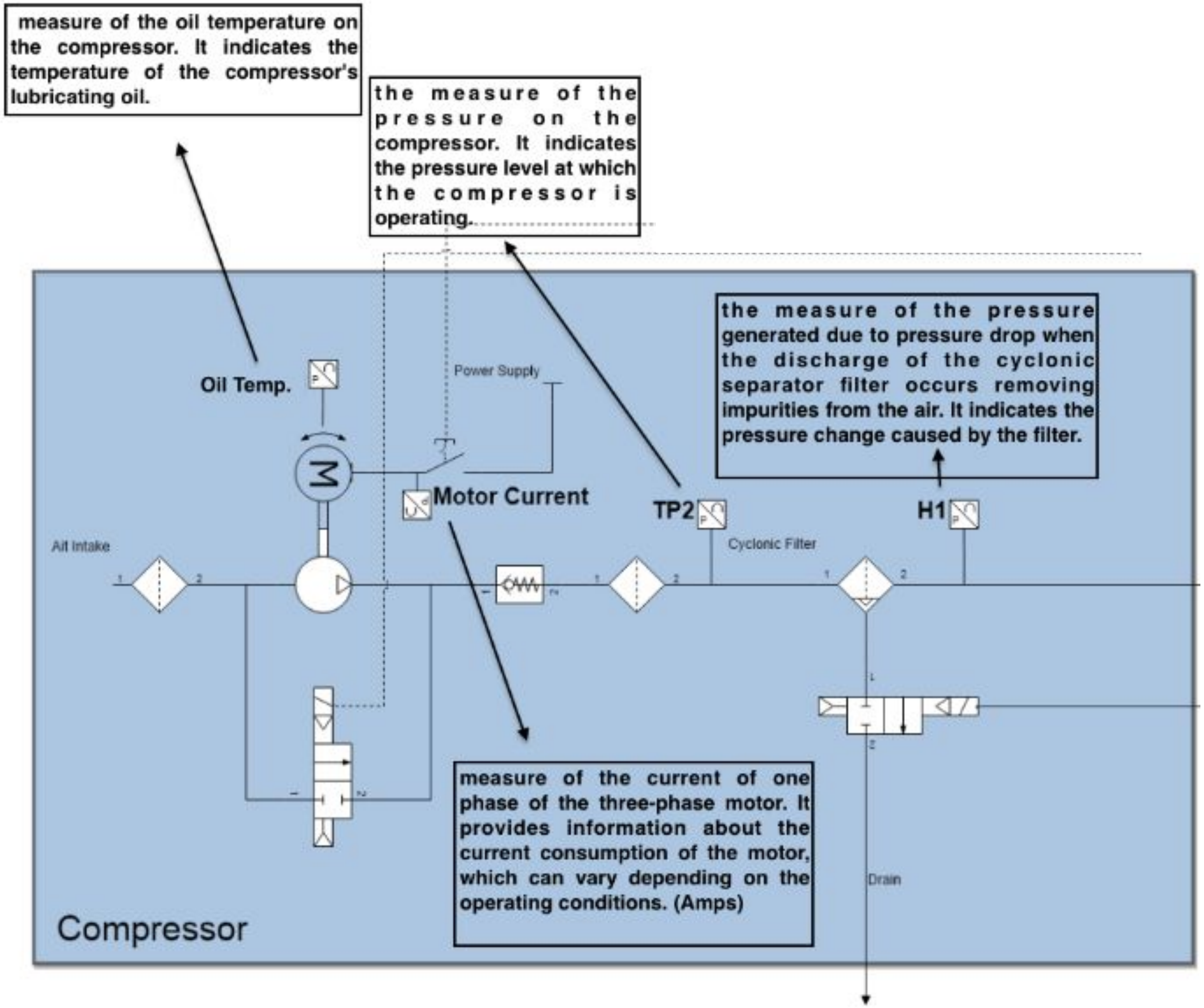
- Metro Train Data -

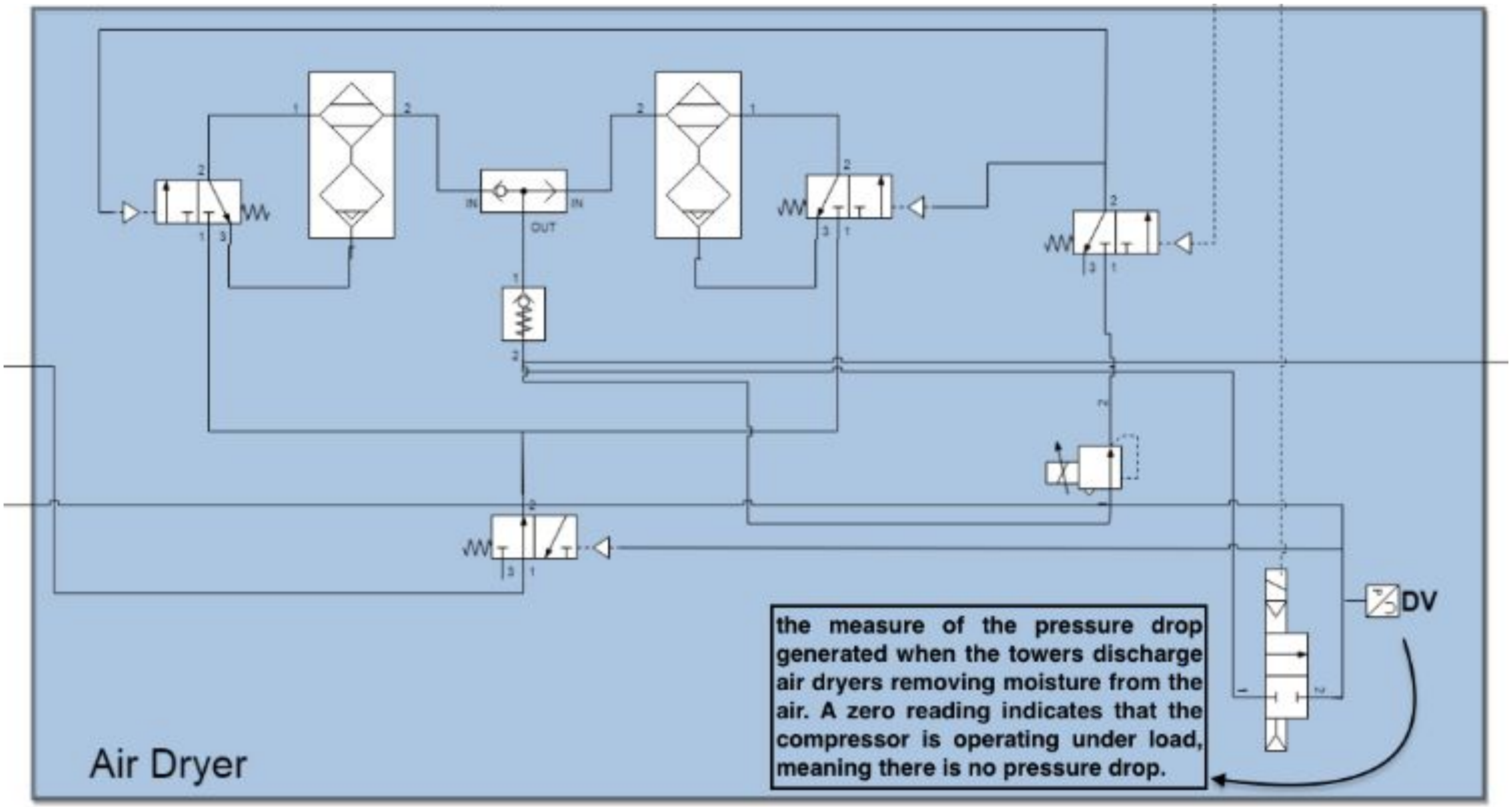
- It consists of Compressor's Air Product Unit(APU) readings.
- Collected at 1Hz from February to August 2020
- Pressure, temperature, motor current consumption, and air intake valves
- 1,516,948 readings or instances and 16 Features
  - Timestamp
  - 7 Analogue Sensors (Continuous)
    - TP2 , TP3, H1 , DV Pressure , Reservoirs , Motor Current , Oil Temperature
  - 8 Digital Sensors (Binary)
    - Comp , DV Electric , Towers , MPG , LPS , Pressure Switch , Oil Level , Caudal Impulses
- Unlabelled data (Initially)

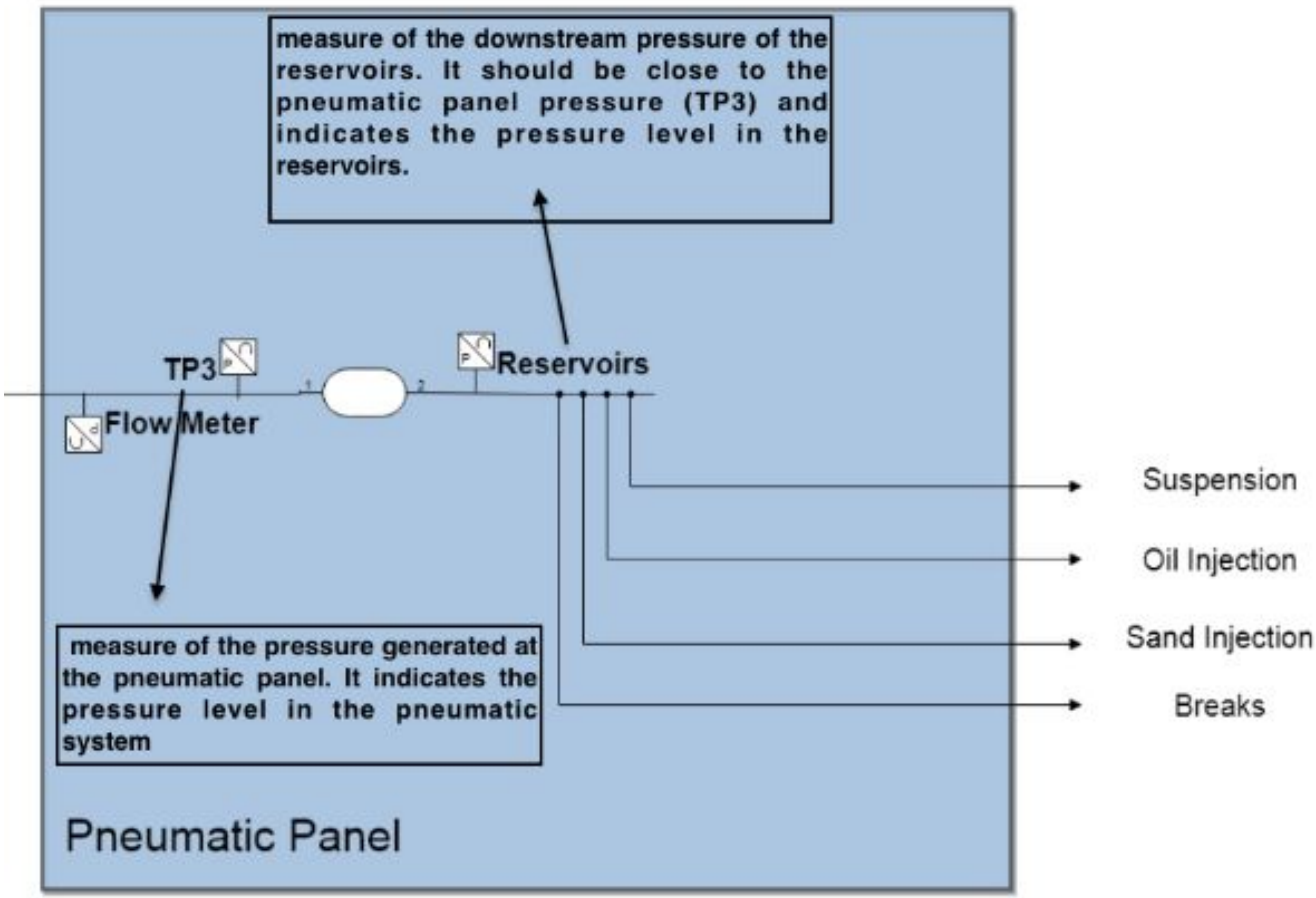
→ Aim: detect observations with APU failure (2.05% of the data)

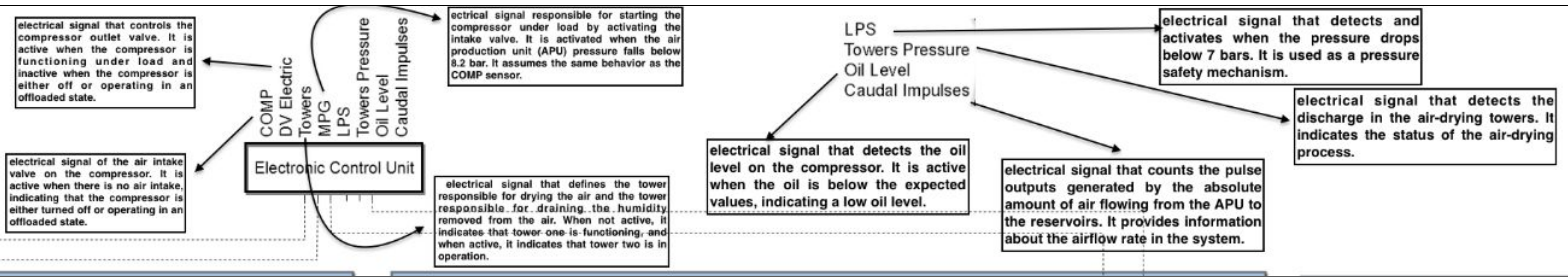


measure of the oil temperature on the compressor. It indicates the temperature of the compressor's lubricating oil.
the measure of the pressure on the compressor. It indicates the pressure level at which the compressor is operating.
electrical signal that controls the compressor outlet valve. It is active when the compressor is functioning under load and inactive when the compressor is either off or operating in an offloaded state.
electrical signal of the air intake valve on the compressor. It is active when there is no air intake, indicating that the compressor is either turned off or operating in an offloaded state.
COMP
DV Electric
Towers
MPG
LPS
Towers Pressure
Oil Level
Caudal Impulses
Electronic Control Unit
ectrical signal responsible for starting the compressor under load by activating the intake valve. It is activated when the air production unit (APU) pressure falls below 8.2 bar. It assumes the same behavior as the COMP sensor.
electrical signal that defines the tower responsible for drying the air and the tower responsible for draining the humidity removed from the air. When not active, it indicates that tower one is functioning, and when active, it indicates that tower two is in operation.
LPS
Towers Pressure
Oil Level
Caudal Impulses
electrical signal that detects and activates when the pressure drops below 7 bars. It is used as a pressure safety mechanism.
electrical signal that detects the discharge in the air-drying towers. It indicates the status of the air-drying process.
electrical signal that detects the oil level on the compressor. It is active when the oil is below the expected values, indicating a low oil level.
electrical signal that counts the pulse outputs generated by the absolute amount of air flowing from the APU to the reservoirs. It provides information about the airflow rate in the system.
Oil Temp.
Power Supply
Motor Current
TP2
H1
Air Intake
Cyclonic Filter
the measure of the pressure generated due to pressure drop when the discharge of the cyclonic separator filter occurs removing impurities from the air. It indicates the pressure change caused by the filter.
measure of the current of one phase of the three-phase motor. It provides information about the current consumption of the motor, which can vary depending on the operating conditions. (Amps)
Drain
Compressor
Air Dryer
IN
OUT
DV
the measure of the pressure drop generated when the towers discharge air dryers removing moisture from the air. A zero reading indicates that the compressor is operating under load, meaning there is no pressure drop.
measure of the downstream pressure of the reservoirs. It should be close to the pneumatic panel pressure (TP3) and indicates the pressure level in the reservoirs.
TP3
Reservoirs
Flow Meter
measure of the pressure generated at the pneumatic panel. It indicates the pressure level in the pneumatic system
Pneumatic Panel
Suspension
Oil Injection
Sand Injection
Breaks


measure of the oil temperature on the compressor. It indicates the temperature of the compressor's lubricating oil.
the measure of the pressure on the compressor. It indicates the pressure level at which the compressor is operating.
the measure of the pressure generated due to pressure drop when the discharge of the cyclonic separator filter occurs removing impurities from the air. It indicates the pressure change caused by the filter.
Oil Temp.
Power Supply
Motor Current
TP2
H1
Air Intake
Cyclonic Filter
measure of the current of one phase of the three-phase motor. It provides information about the current consumption of the motor, which can vary depending on the operating conditions. (Amps)
Drain
Compressor


IN
IN
OUT
DV
the measure of the pressure drop generated when the towers discharge air dryers removing moisture from the air. A zero reading indicates that the compressor is operating under load, meaning there is no pressure drop.
Air Dryer


measure of the downstream pressure of the reservoirs. It should be close to the pneumatic panel pressure (TP3) and indicates the pressure level in the reservoirs.
TP3
Reservoirs
Flow Meter
Suspension
Oil Injection
Sand Injection
Breaks
measure of the pressure generated at the pneumatic panel. It indicates the pressure level in the pneumatic system
Pneumatic Panel



electrical signal that controls the compressor outlet valve. It is active when the compressor is functioning under load and inactive when the compressor is either off or operating in an offloaded state.
electrical signal of the air intake valve on the compressor. It is active when there is no air intake, indicating that the compressor is either turned off or operating in an offloaded state.
COMP
DV Electric
Towers
MPG
LPS
Towers Pressure
Oil Level
Caudal Impulses
Electronic Control Unit
ectrical signal responsible for starting the compressor under load by activating the intake valve. It is activated when the air production unit (APU) pressure falls below 8.2 bar. It assumes the same behavior as the COMP sensor.
electrical signal that defines the tower responsible for drying the air and the tower responsible for draining the humidity removed from the air. When not active, it indicates that tower one is functioning, and when active, it indicates that tower two is in operation.
LPS
Towers Pressure
Oil Level
Caudal Impulses
electrical signal that detects the oil level on the compressor. It is active when the oil is below the expected values, indicating a low oil level.
electrical signal that counts the pulse outputs generated by the absolute amount of air flowing from the APU to the reservoirs. It provides information about the airflow rate in the system.
electrical signal that detects and activates when the pressure drops below 7 bars. It is used as a pressure safety mechanism.
electrical signal that detects the discharge in the air-drying towers. It indicates the status of the air-drying process.

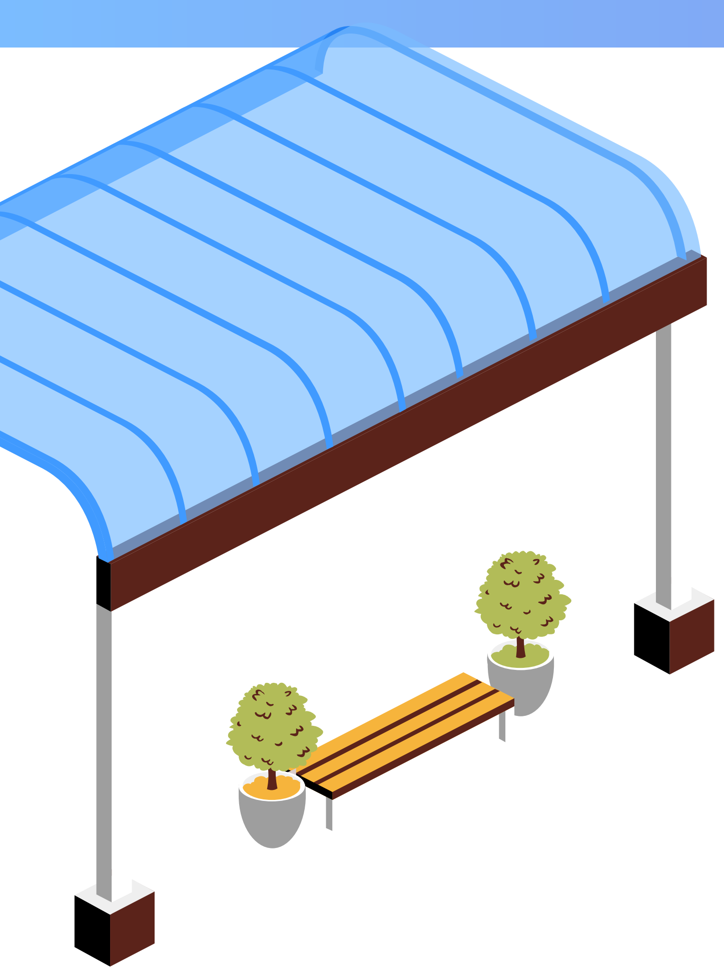
The Dataset

- Data Labelling -

- Aim: detect observations with APU failure
- Data is labelled based on air leaks from maintenance reports
- Label ‘Airleak’: 0 → no airleak, 1 → airleak (flagged)
- 29,865 readings (2.05% of the data)

Nr.	Start Time	End Time	Failure	Severity	Report
#1	4/18/2020 0:00	4/18/2020 23:59	**Air leak**	High stress	
#1	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

Methodology



Approach 1

Full data

(1,459,475 , 15)
2.05% anomalies

Approach 2

PCA on Full data
2 PCA components

1,459,475 Obs.
2.05% anomalies

Approach 3

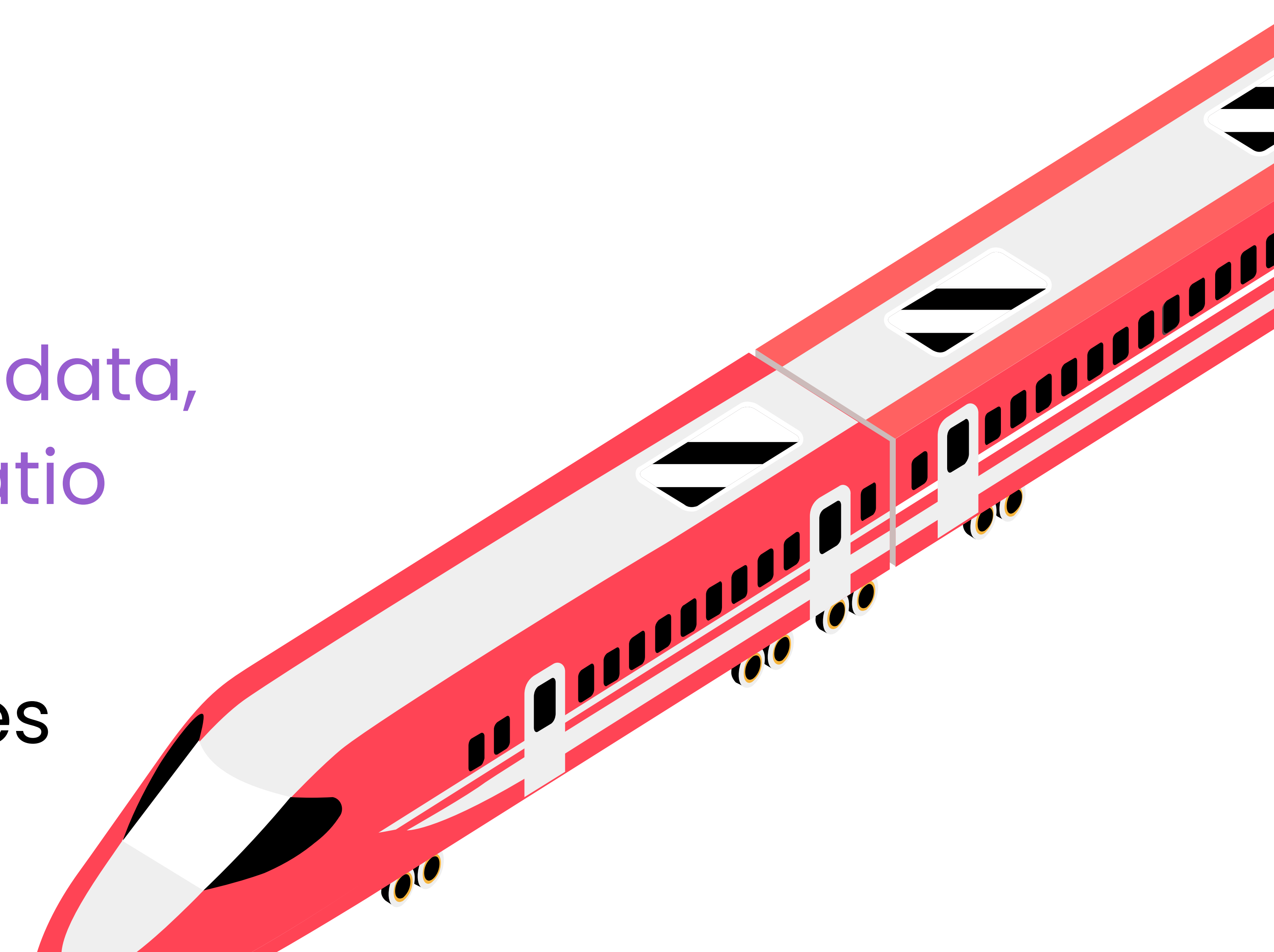
Half the data:
random sample of data, preserving the ratio

(729,738 , 15)
2.07% anomalies

Approach 4

Balanced data:
All minority + RUS majority

(59,730 , 15)
50% anomalies



Methodology

- Anomaly detection -

(Approaches 1,2,3)

<u>1. Clustering-based Approach</u>

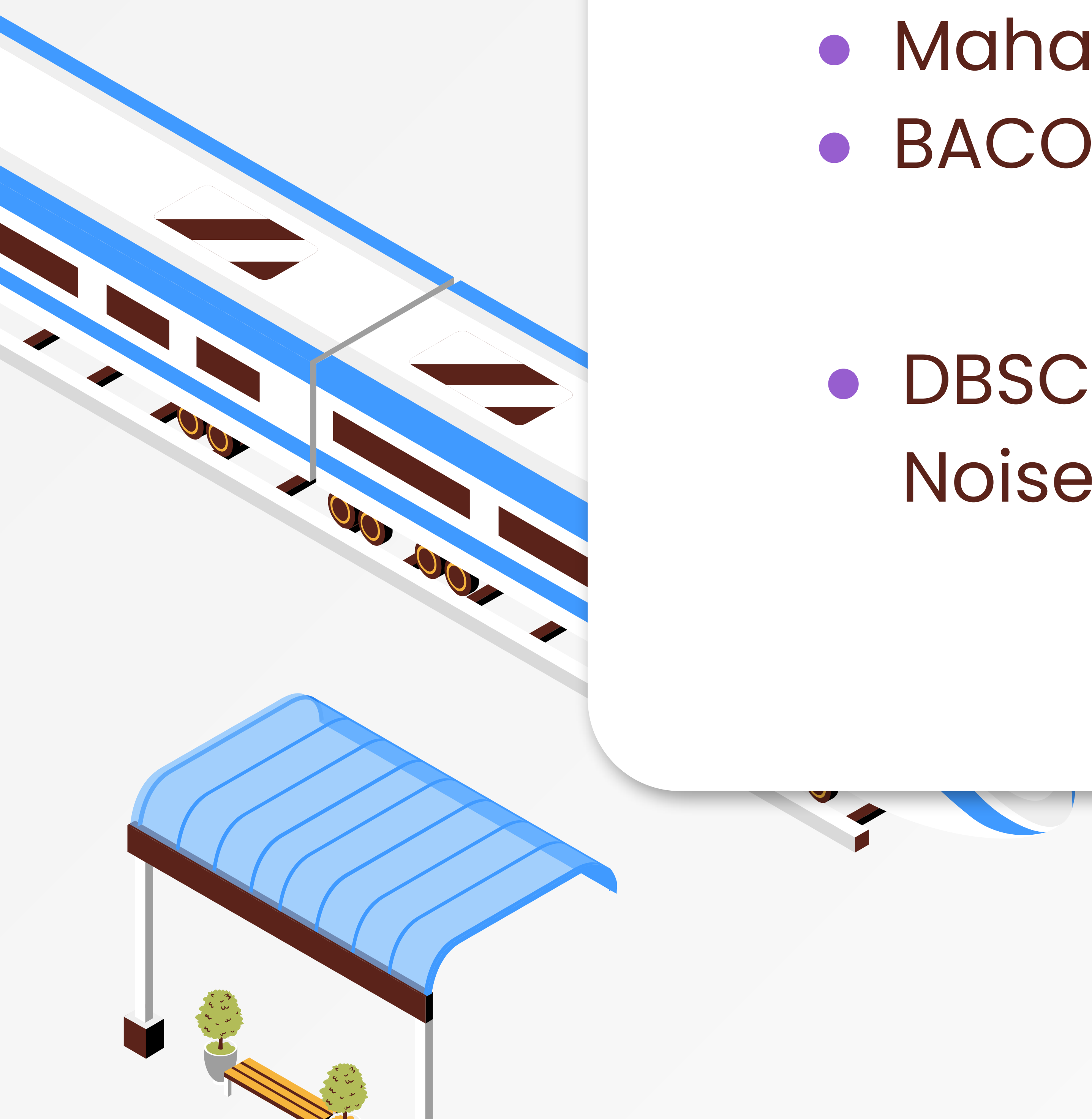
- k-Means clustering (euclidean distance
- Hierarchical Clustering (ward, single, complete, …)
- Spectral Clustering

<u>2. Distance-based Approach</u>

- Fifth Nearest Neighbour (euclidean distance)
- Mahalanobis Distance
- BACON

<u>3. Density-based Approach</u>

- DBSCAN: Density Based Spatial Clustering of Applications with Noise



Methodology

- Classification -

(Approach 4)

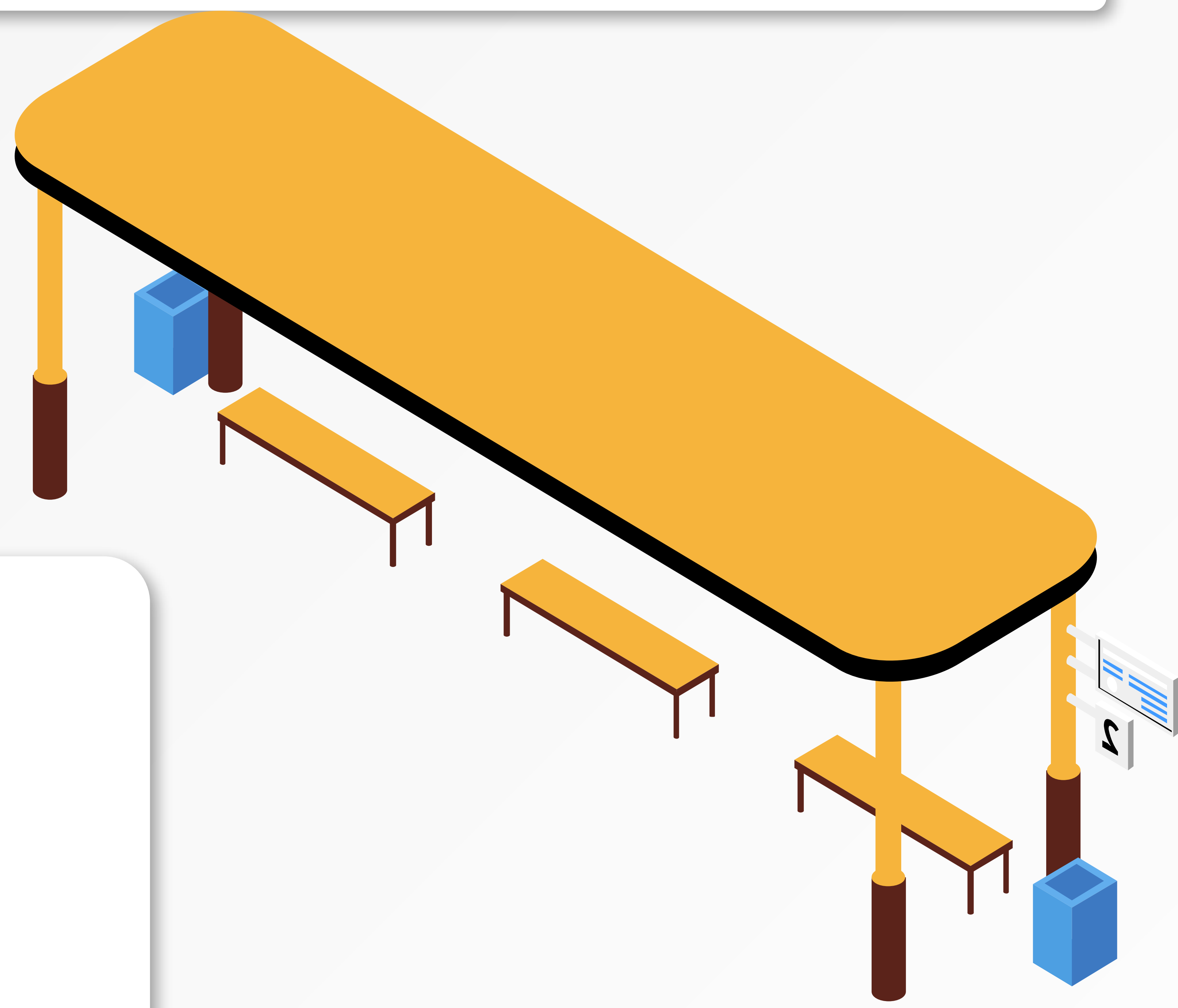
1. Clustering-based Approach

- k-Means clustering (euclidean distance
- Hierarchical Clustering (ward, single, complete, …)
- Spectral Clustering

2. Machine Learning Classifiers

- Random Forest
- K-Nearest Neighbour (K-NN)
- Logistic Regression
- Support Vector Machines (SVM)
- Least squares
- Perceptron

Methodology



<u>Performance Metrics</u>

- Precision

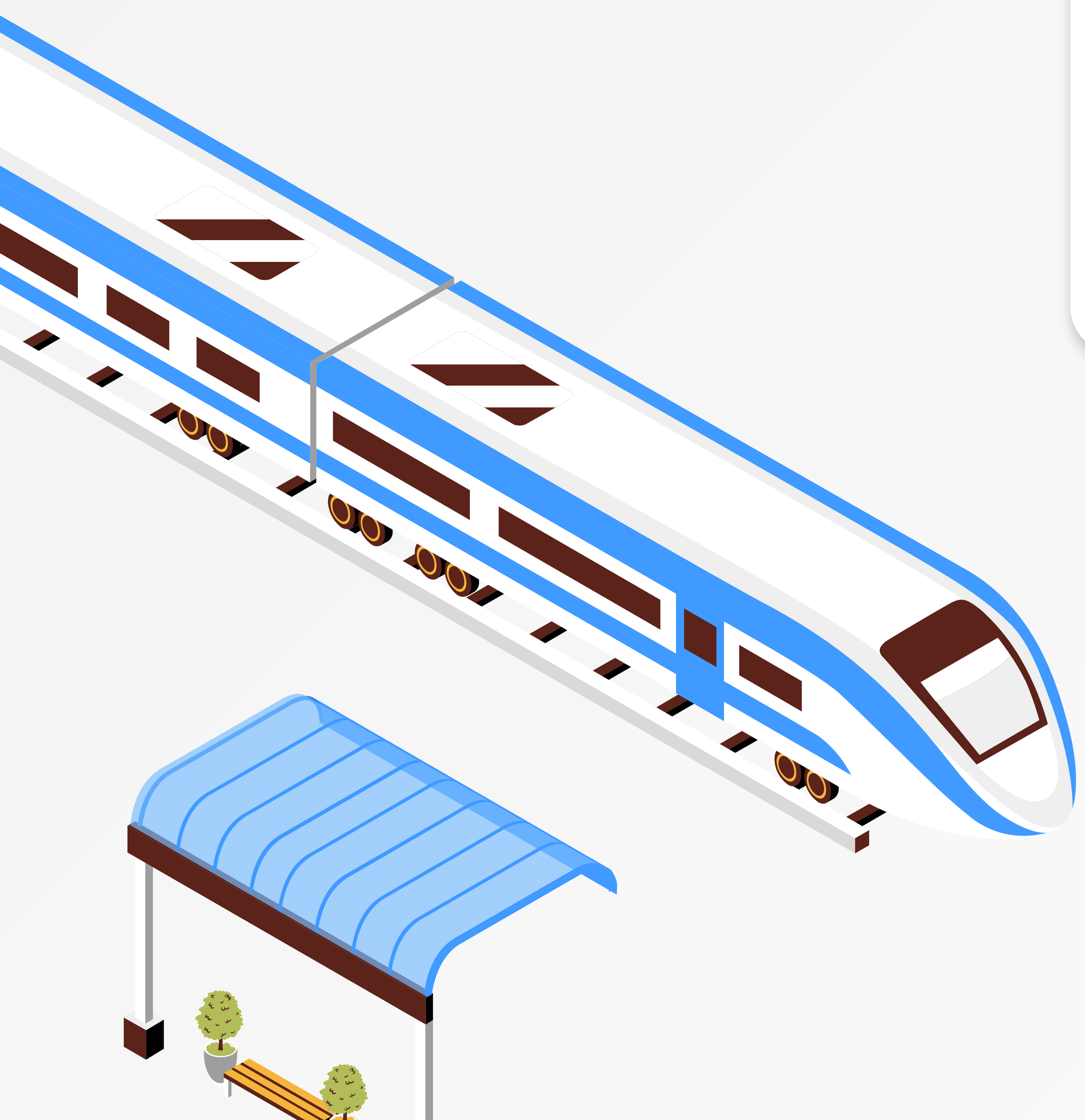
(true positive rate)

- Recall

(false positive rate)

- F1-Score

(weighted average of both)



Approach 1: Full data

1. **Clustering**

	Precision	**Recall**	**F1-score**
K-means	0.05	1.00	0.09
Hierarchical Clustering	Crashed		
Spectral Decomposition	Crashed		

2. **Distance-based Approach for Anomaly Detection**

	Precision	**Recall**	**F1-score**
Fifth Nearest Neighbour	0.09	0.18	0.12
Mahalanobis Distance	0.21	1.0	0.35
BACON	0.07	1.00	0.14

3. **Density-based Approach for Anomaly Detection**

DBSCAN (Density Based Spatial Clustering of Applications with Noise)

	Precision	**Recall**	**F1-score**
eps= 0.2, min_samples= 4	Crashed		
eps= 0.2, min_samples= 3	Crashed		

Approach 2: PCA (2 components) on full data

1. **Clustering**

	Precision	Recall	F1-score
K-means	0.13	1.00	0.24
Hierarchical Clustering	Crashed		
Spectral Decomposition	Crashed		

2. **Distance-based Approach for Anomaly Detection**

	Precision	Recall	F1-score
Fifth Nearest Neighbour	0.13	0.17	0.14
Mahalanobis Distance	0.20	1.00	0.33
BACON	0.06	1.00	0.11

3. **Density-based Approach for Anomaly Detection**

DBSCAN (Density Based Spatial Clustering of Applications with Noise)

	Precision	Recall	F1-score
eps= 0.2, min_samples= 4	Crashed		

Approach 3: Random Sample (1/2 the data, same ratio)

1. **Clustering**

	Precision	Recall	F1-score
K-means	0.05	1.00	0.09
Hierarchical Clustering	Crashed		
Spectral Decomposition	Crashed		

2. **Distance-based Approach for Anomaly Detection**

	Precision	Recall	F1-score
Fifth Nearest Neighbour	Crashed		
Mahalanobis Distance	0.22	1.00	0.35
BACON	0.07	1.00	0.13

3. **Density-based Approach for Anomaly Detection**

DBSCAN (Density Based Spatial Clustering of Applications with Noise)

	Precision	Recall	F1-score
eps= 0.2, min_samples= 4	Crashed		
eps= 0.2, min_samples= 3	0.04	0.05	0.05

Approach 4: Balanced Data

1. **Clustering**

	Precision	Recall	F1-score
K-means	0.92	1.00	0.96
Hierarchical Clustering	Crashed		
Spectral Decomposition	Crashed		

2. **Machine Learning Classifiers**

Results are based on cross-validation on 5-fold

	Precision	Recall	F1-score
Random Forest	1.00	0.93	0.96
K-NN (3)	0.99	0.93	0.96
Logistic Regression	0.98	0.94	0.96
SVM	0.98	1.00	0.99
Least Squares	0.98	0.94	0.96
Perceptron	0.98	0.98	0.98

Conclusion

Finally, it can be concluded that **machine learning classifiers performed the best**.

For future recommendations, it would be valuable to attempt techniques that crashed in our application, or attempt dimensionality reduction (PCA) on the sampled datasets.