Finanziato dall'Unione europea
NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA
ICSC
Centro Nazionale di Ricerca in HPC, Big Data and Quantum Computing

ICSC
Centro Nazionale di Ricerca in HPC,
Big Data and Quantum Computing

Harnessing the Power of Artificial Intelligence for Predictive Maintenance of Industrial Plants

INAF ISTITUTO NAZIONALE DI ASTROFISICA
ALMA MATER STUDIORUM UNIVERSITÀ DI BOLOGNA
INFN Istituto Nazionale di Fisica Nucleare
eni

Costa,A. Mastriani,E.
INAF OACT

Spoke 2 meeting 26.06.2024

Harnessing the Power of Artificial Intelligence for Predictive Maintenance of Industrial Plants

- Goals:
  - Demonstrate the usability of AI-based techniques for predictive maintenance.
  - Model interdependencies in complex industrial apparatus.
- Partners:
  - Eni, UniBO, INAF, INFN
  - Eni production sites, including sensor readings and apparatus data.

Work Packages (WP)

WP1 – Eni

Objective: Identify and prioritize equipment with complex behaviors.

Tasks: Data selection, delivery, and interpretation with benchmarking.

WP2 – UniBO

Objective: Compare performance of autoencoders (AE) and graph neural networks (GNN).

Tasks: Data exploration, prototype development, validation, and performance tuning.

WP3 – INAF

Objective: Identify best unsupervised learning approaches for anomaly detection.

Tasks: Data acquisition, anomaly detection demonstrator, model validation.

WP4 – INFN

Objective: Manage data streaming and analysis using various techniques.

Tasks: Data ingestion platform POC (Proof Of Concept), comparison of approaches.

Research Plan and Reporting

Reporting Period: M8 – March 2024 – June 2024

WP1 – Eni

- Target: Interpretation and benchmarking of results.
- KPI: Benchmarking report.

WP2 – UniBO

- Target: Prototype of AE-based approach.
- KPI: Prototype description report.

WP3 – INAF

- Target: Data acquisition campaign & Prototype of AE-based approach.
- KPI: Report with main results.

WP4 – INFN

- Target: Selection of analysis approaches and data ingestion platform POC.
- KPI: Report on selected approaches and POC description.

INAF ISTITUTO NAZIONALE DI ASTROFISICA
ALMA MATER STUDIORUM UNIVERSITÀ DI BOLOGNA
INFN Istituto Nazionale di Fisica Nucleare
eni

Agenda

- The object of our study
- Goal and main questions
- The adopted pipeline
- Main results
- Next preliminary steps

The object of our study: high-pressure compressor

Attribute	Description
Type	High-pressure compressor
Sensors	Equipped with 75 sensors
Current Status	Currently in production
Primary Use	Increases the pressure of gas before introduction into the SNAM methane pipeline
Design	Made up of a piston that slides inside a cylinder and two suction/delivery valves
Application	Designed for high-pressure air compression applications typical of petrochemical and refining plants
Stages	2-stage design
Delivery Pressure	Can reach up to 40 bar
Critical Asset Reliability	Extensive set of integrated sensors
Vibration Sensors	Monitors the condition of bearings and shafts
Temperature Sensors	Monitors engine, cylinders, and lubricating oil
Pressure Sensors	Monitors suction, delivery, and oil
Flow Sensors	Monitors compressor output

If it breaks, what's the consequence?

FAIL

Consequence	Description
Disruption of Gas Supply	Interrupts the continuous supply of gas, leading to reduced flow or complete halt in gas delivery.
Pressure Imbalance	Causes pressure imbalance in the pipeline system, creating safety concerns.
Equipment Damage	Leads to further mechanical damage, necessitating expensive repairs or replacement.
Safety Hazards	Poses risks of leaks, explosions, or other dangerous scenarios endangering personnel and infrastructure.
Operational Downtime	Results in halted operations, affecting productivity and potentially leading to financial losses.
Environmental Impact	Potential gas leaks could release methane into the atmosphere, contributing to greenhouse gas emissions.
Financial Losses	Includes direct costs (repair, replacement) and indirect costs (lost production, fines, reputational damage).

Goal and questions

The Goal: Create a model to accurately classify events that indicate potential compressor failure

The questions: Using unsupervised clustering algorithms as additional features for clustering models:

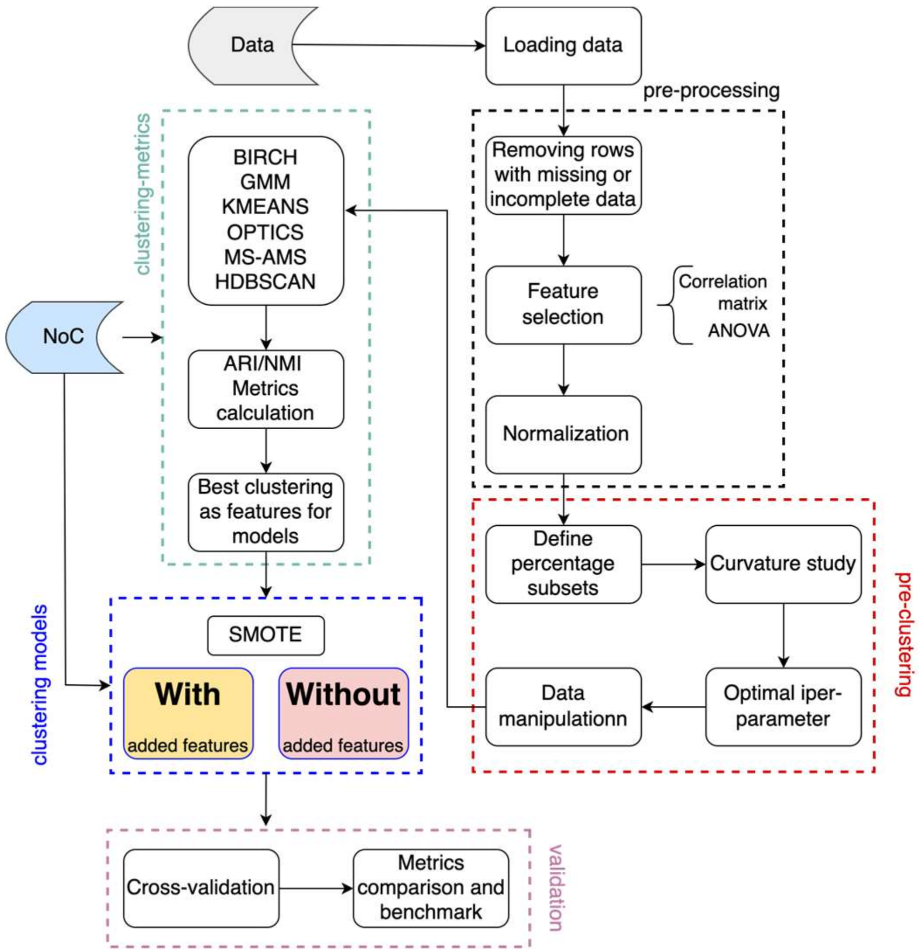
Does it improve predictions?

If yes, how much and when?

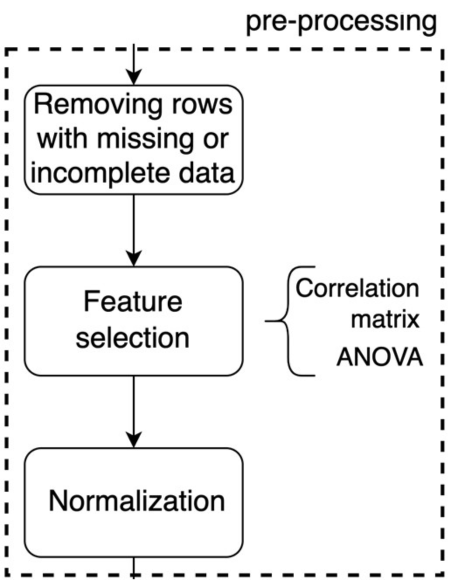
Do training times change using clustering features?

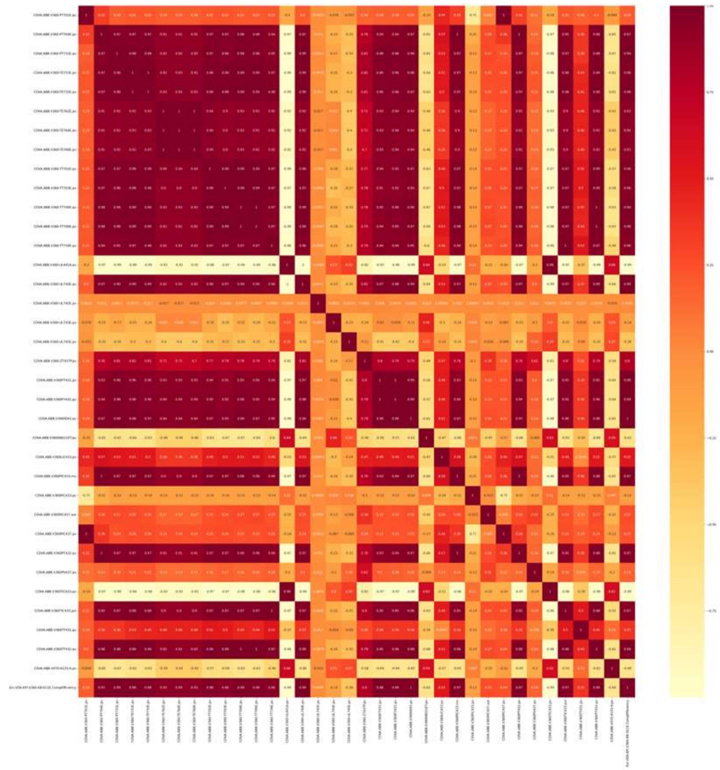
The main process

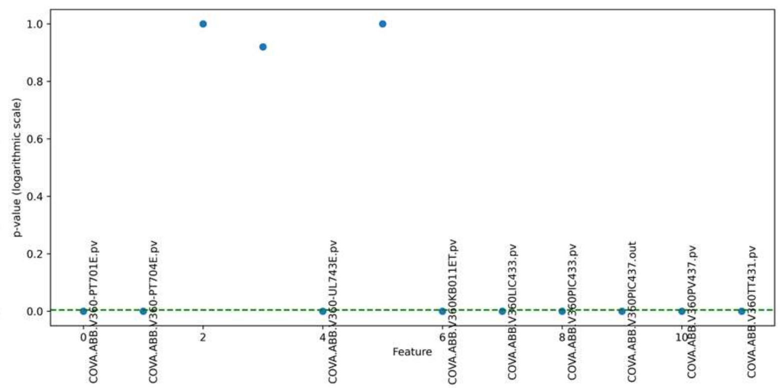
- Pre-processing
- Pre-clustering
- Clustering and Metrics
- Clustering models
- Validation



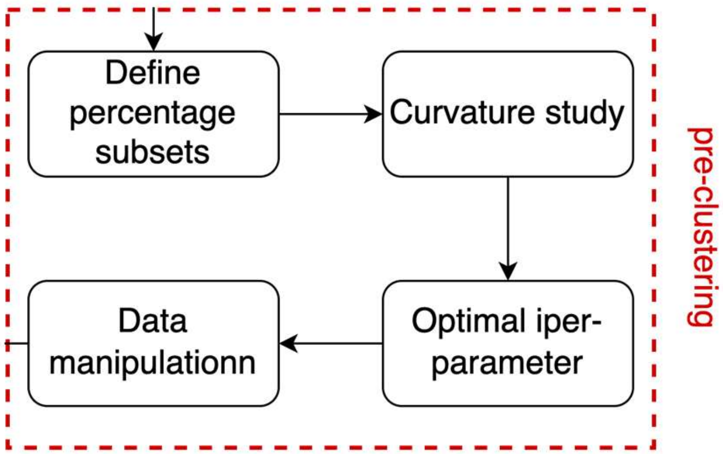
Pre-processing

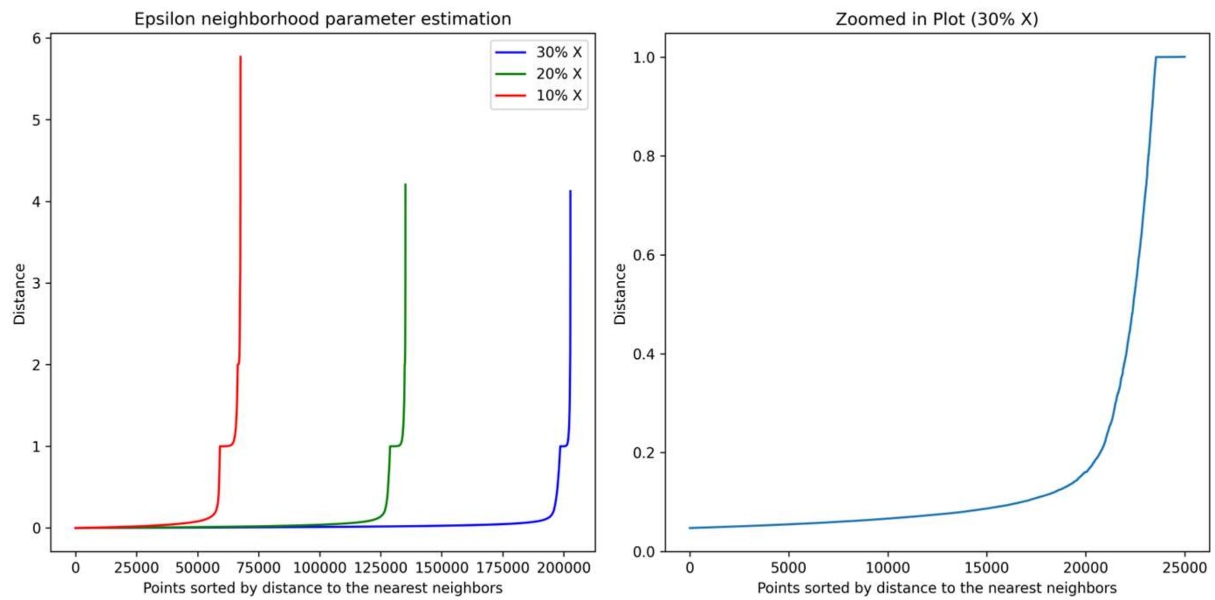




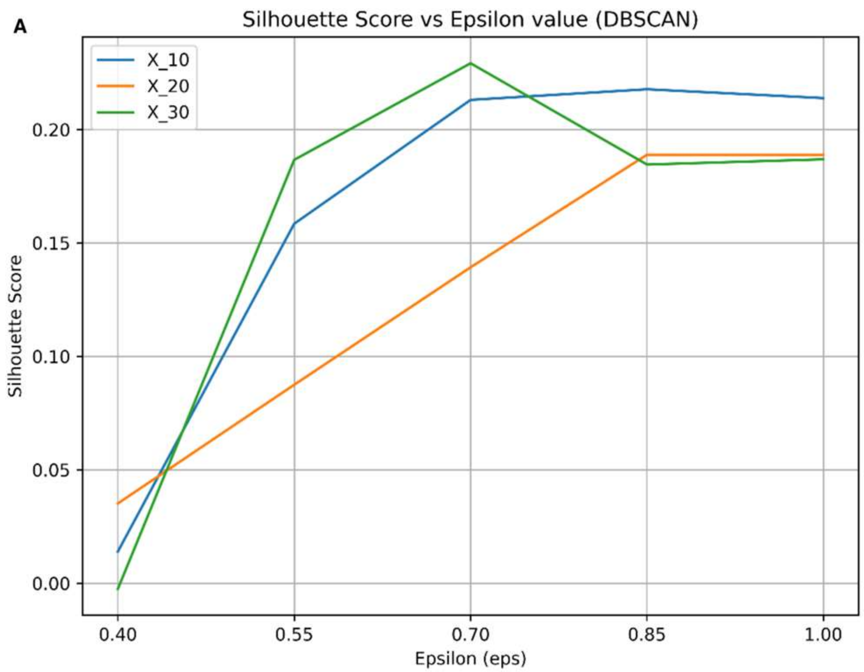


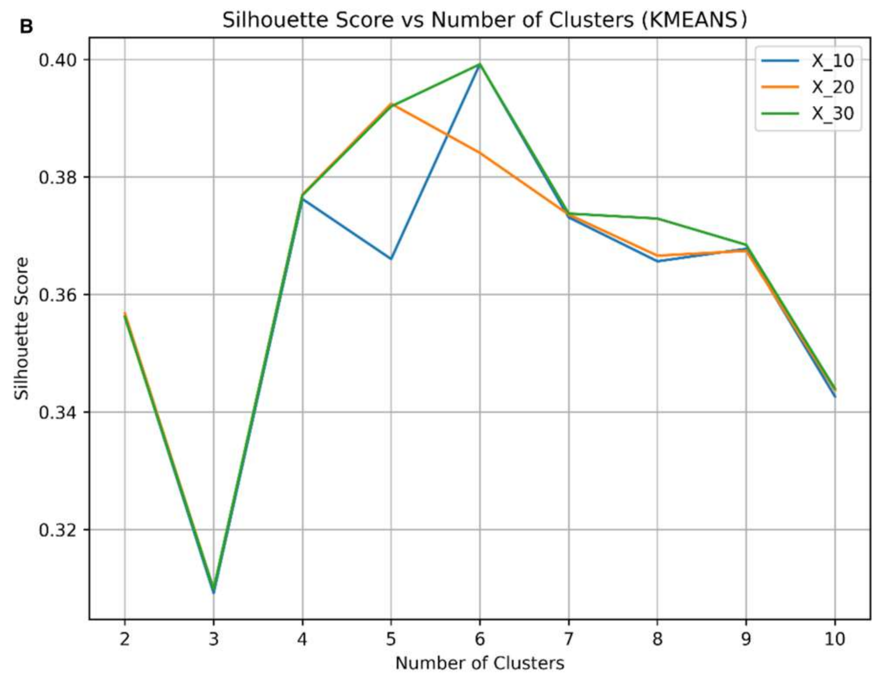
Pre-clustering (1)



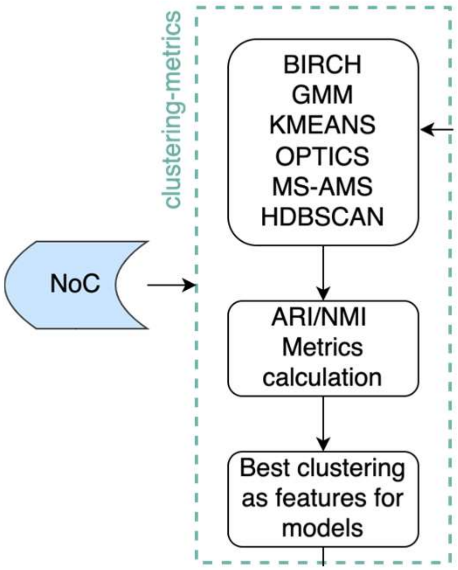


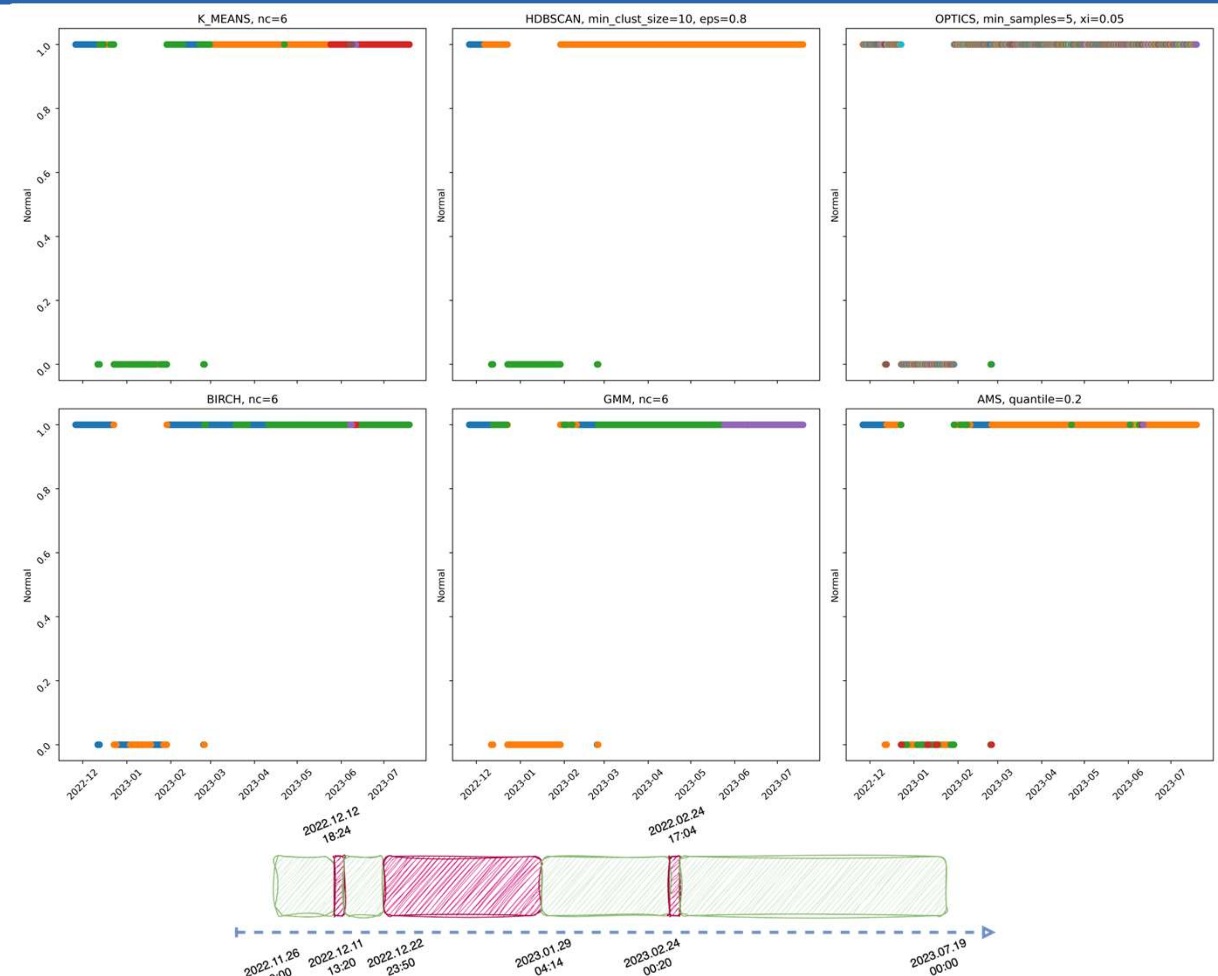
Pre-clustering (2)





Clustering and …



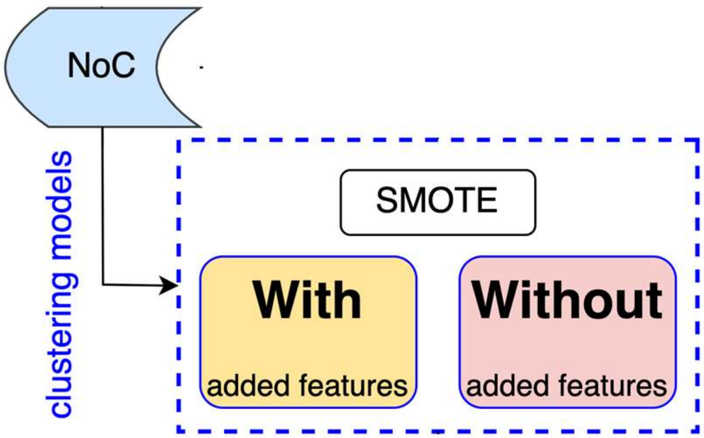


...Metrics

Clustering Algorithm	ARI	NMI
KMEANS	0.30	**0.42***
HDBSCAN	**0.55***	**0.64***
OPTIC	-0.02	0.16
BIRCH	0.03	0.25
GMM	**0.45***	**0.56***
MS-AMS	**0.35***	0.36

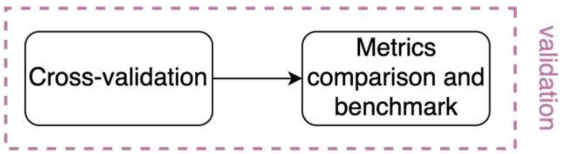
Clustering models

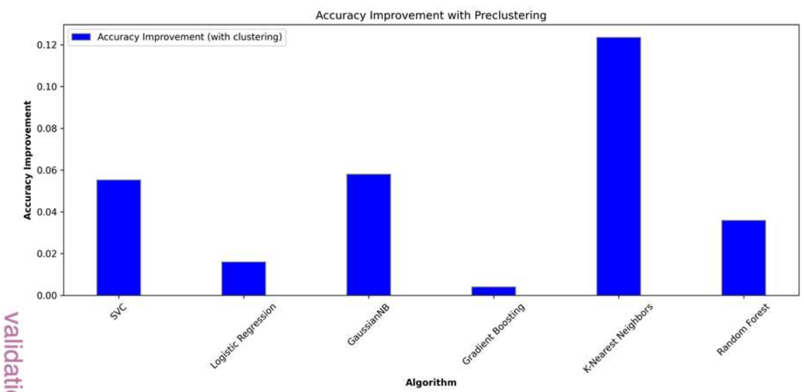
data imbalance: SMOTE (Synthetic Minority Over-sampling Technique)

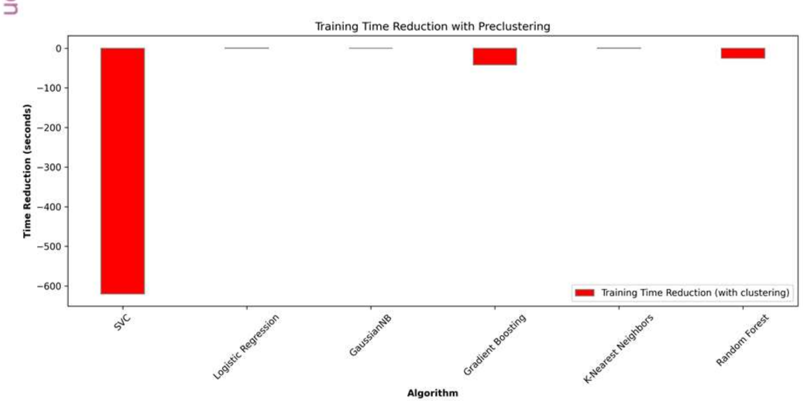


- Linear Algorithms: Logistic Regression
- Kernel-based algorithms: SVC
- Probabilistic Algorithms: GaussianNB
- Gradient Algorithms: GradientBoostingClassifier
- Instance-based algorithms: KNeighborsClassifier
- Decision Tree Algorithms: RandomForestClassifier

Cross-validation and results







Algorithm	Mean Accuracy (train)	Mean Accuracy (test)	Mean Execution Time (train)
SVC	1.0000	0.9786	112.28 s
	0.9953	0.9234	732.22 s
Logistic Regression	0.9837	0.8037	1.28 s
	0.9019	0.7877	0.74 s
GaussianNB	0.9899	0.9924	0.14 s
	0.8853	0.9344	0.11 s
Gradient Boosting	1.0000	0.9281	96.20 s
	1.0000	0.9240	138.25 s
K-Nearest Neighbors	1.0000	0.9819	1.17 s
	0.9999	0.8585	0.86 s
Random Forest	1.0000	0.9601	32.69 s
	1.0000	0.9242	57.80 s

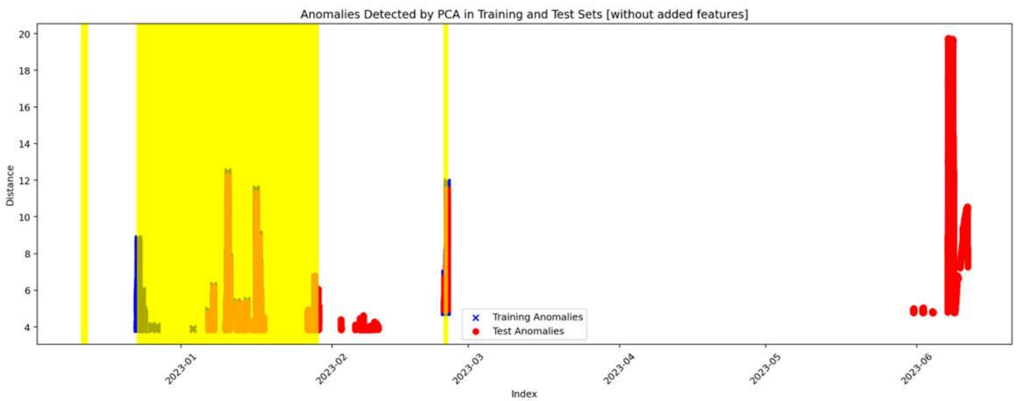
Results

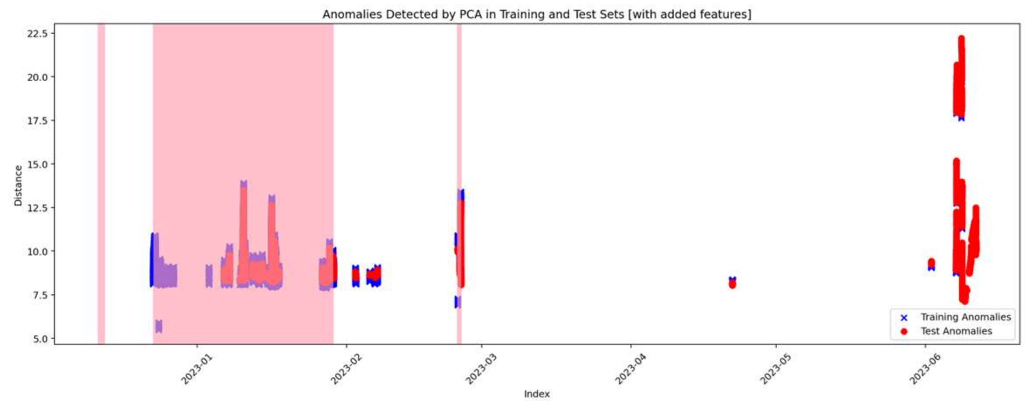
Utilizing the best features identified through unsupervised pre-clustering enhances the accuracy of the test set in identifying potential faults by an average of 4.87%.

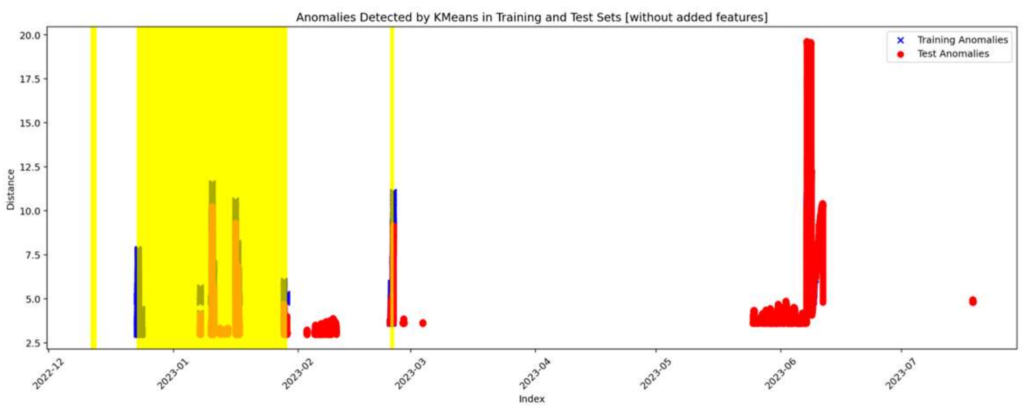
Additionally, the training time for clustering models is reduced by an average of 22.96%.

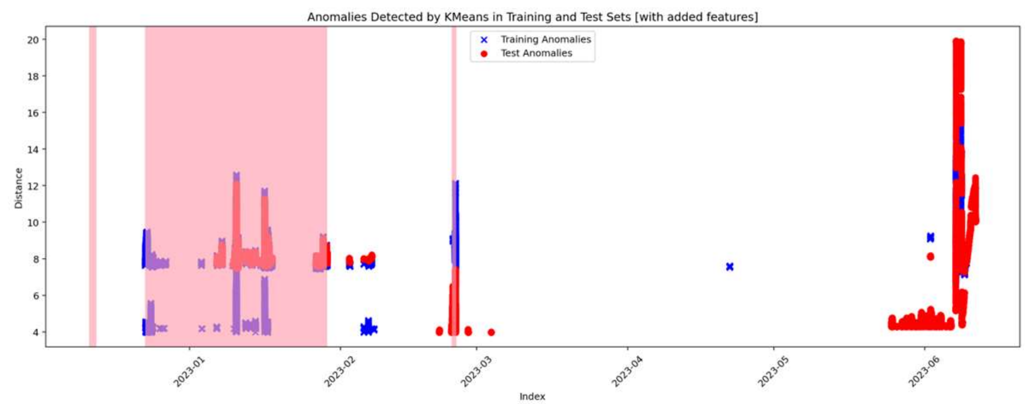
Results suggest that including pre-clustering features can be advantageous, especially in contexts where accuracy is critical and training times are relevant.

Preliminaries on Anomaly Detection (1)

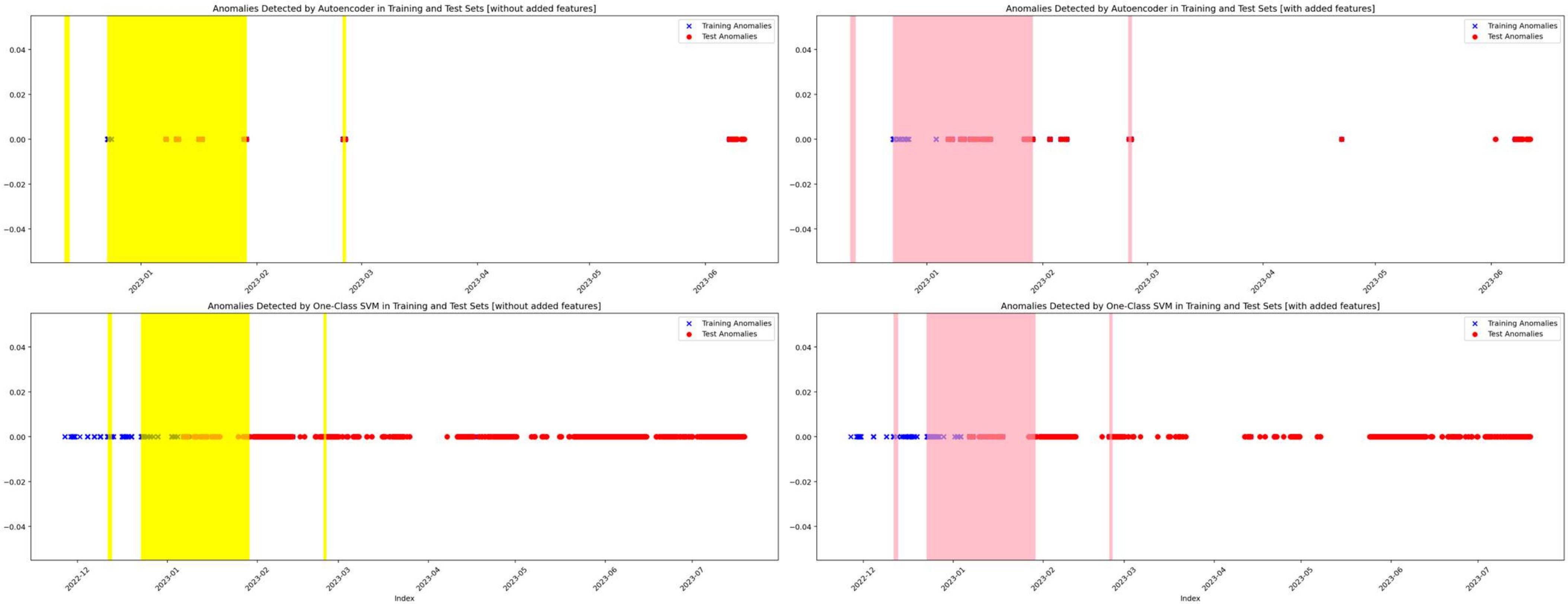




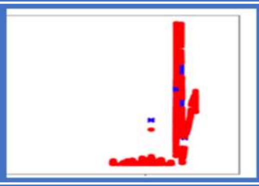




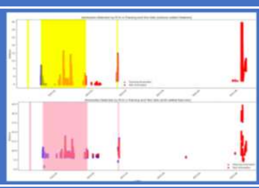
Preliminaries on Anomaly Detection (2)



Preliminary Questions on Anomaly Detection and next steps



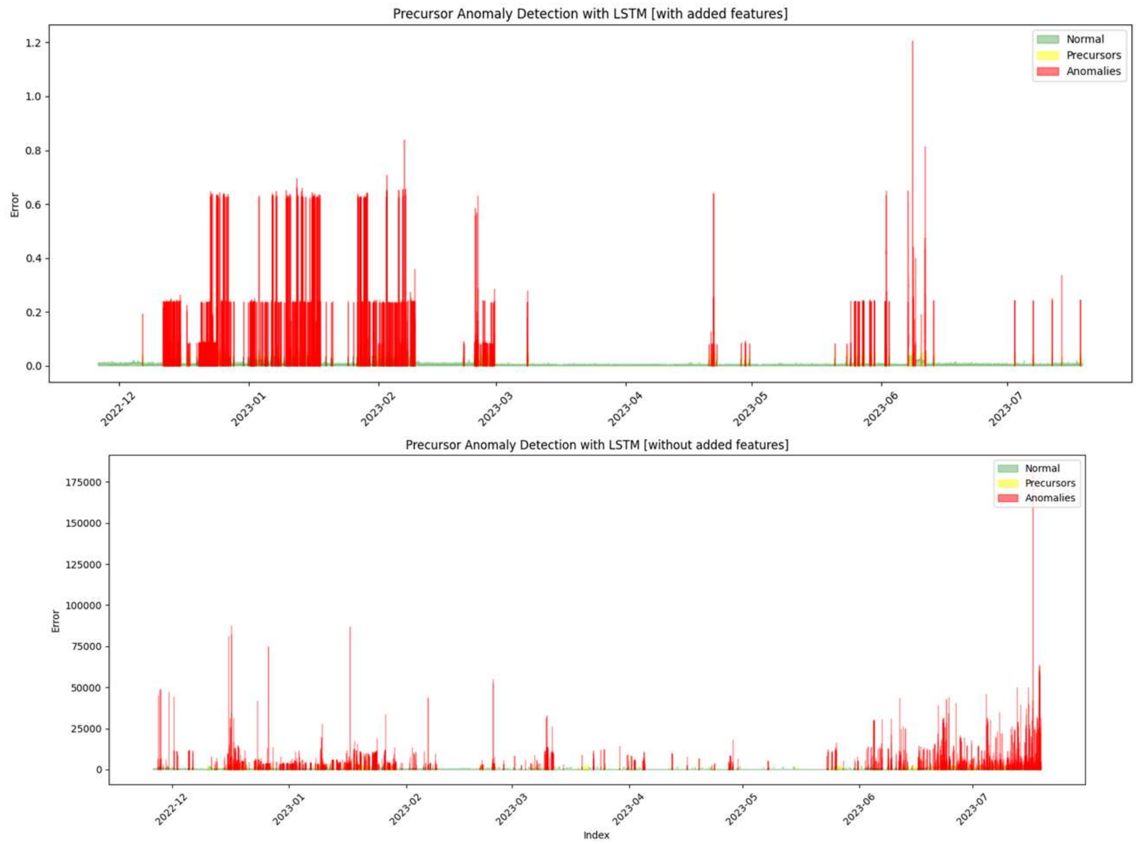
Why do all the methods identify the 'fault period' in June (not present in the NoC) ?



Why do the added features not seem to improve the accuracy in this case?

To provide the metrics of accuracy, such as F1-score, Recall, ROC, and PR

Preliminaries on Precursor Anomaly Detection



Precursor Anomaly Detection, next steps

Check for the most relevant methods from different strategies

Obtain the metrics for all the methods to be used

Identify iper-parameters that could improve the quality of the prediction

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
ICSC Centro Nazionale di Ricerca in HPC, Big Data and Quantum Computing

Thank You