eXplainable Artificial Intelligence (XAI) in Industrial Diagnostic: Interpretable Unsupervised Anomaly Detection

Gian Antonio Susto

Associate Professor, Università degli Studi di Padova, Italy

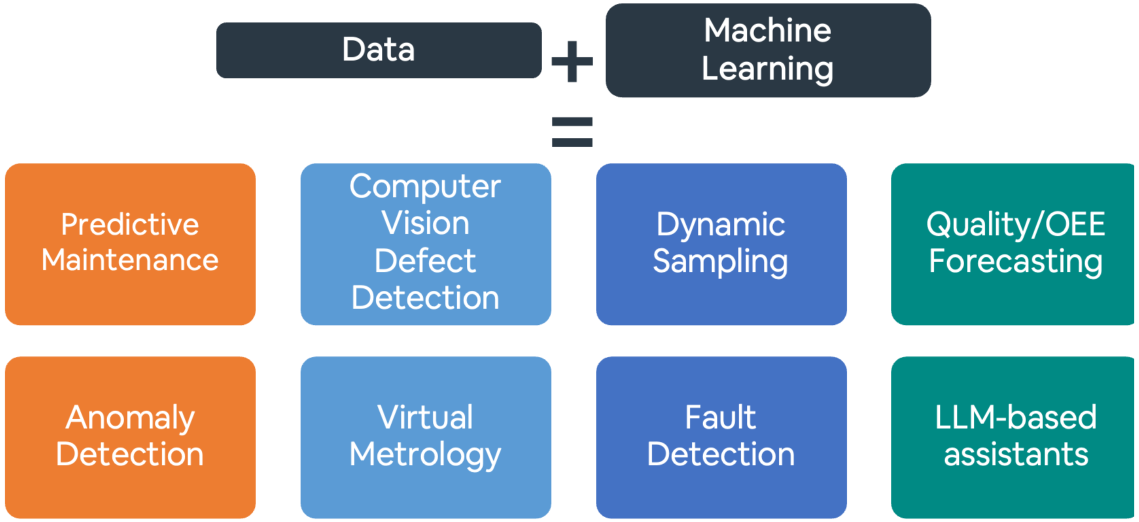
September 3rd, 2024

@ Lausanne – 8th Intelligent Maintenance Conference

The speaker

- Associate Professor of Machine Learning (ML) & Control @ Università degli Studi di Padova, Italy
- Leading the Artificial intelligence, Machine learning & Control (AMCO) research group. What we do: Explainable Artificial Intelligence, Reinforcement Learning, Unsupervised Learning, Industrial Applications, Fairness, Active Learning...
- Co-founder @ Statwolf: software company developing industrial Machine Learning-based solutions

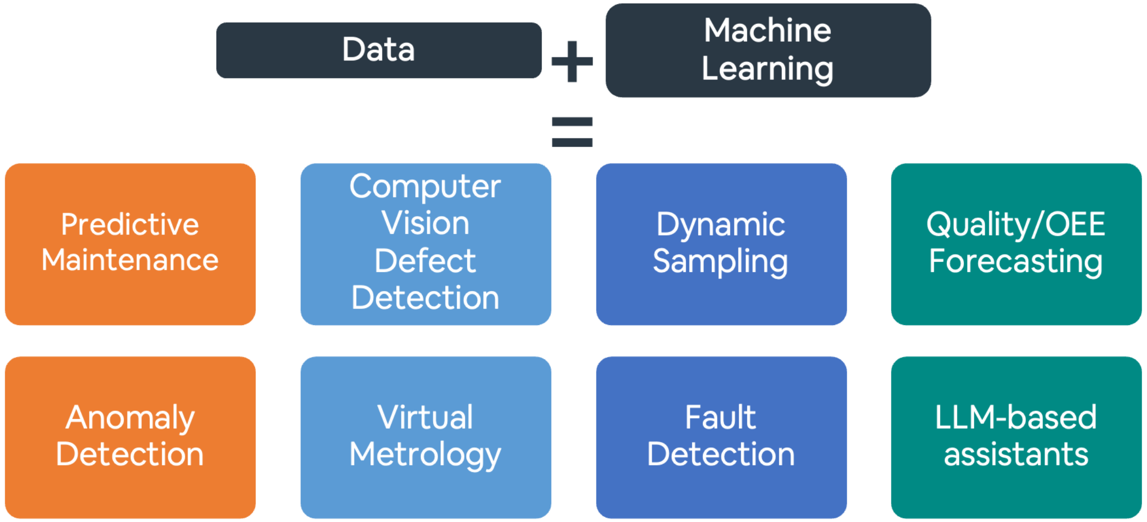
Goals



Goals:

- More efficient process monitoring
- More accurate process control
- Reduced downtime
- Improved maintenance/service operations
- Improved quality/reduced scrap products
- Faster/more accurate detection of defects
- ...

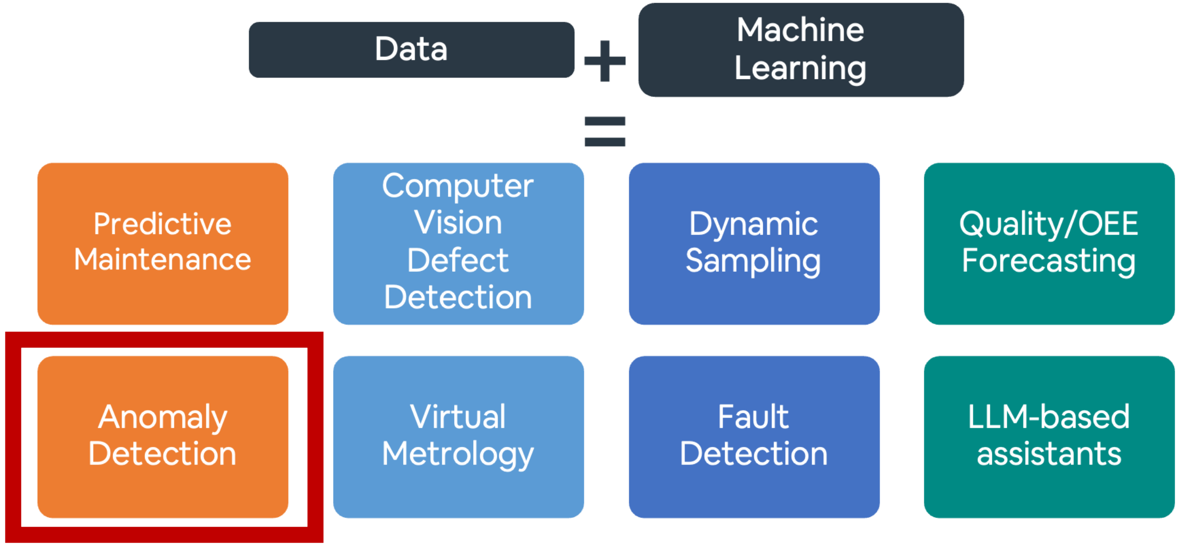
Challenges



<u>Challenges:</u>

1. Data availability: ie. Lack of tagged/sufficient/consistent data;
2. The 'human-in-the-loop factor': ie. Lack of trust in ML solutions, difficulties in translating ML suggestions into actions and business impact;
3. Lack of ML-ready architectures: ie. no interoperability of systems, no tool to properly monitor and improve ML solutions (MLOps);
4. No off-the-shelf solutions: while some tasks can be considered 'solved', others required customized approaches;
5. Scalability: ie. Is it worth to develop a ML solution if I cannot easily scale it to several machines/production sites/customers?
6. ...

Challenges



The rest of this talk!

Challenges:

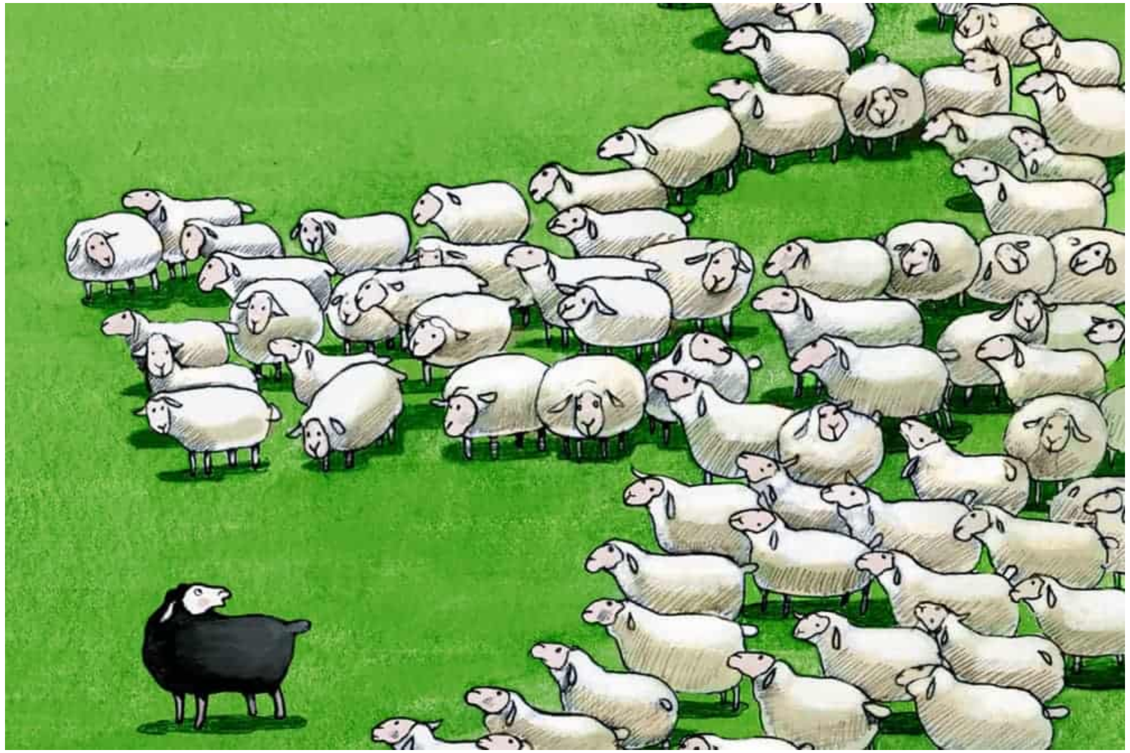
1. Data availability: ie. Lack of tagged/sufficient/consistent data;
2. The 'human-in-the-loop factor': ie. Lack of trust in ML solutions, difficulties in translating ML suggestions into actions and business impact;
3. Lack of ML-ready architectures: ie. no interoperability of systems, no tool to properly monitor and improve ML solutions (MLOps);
4. No off-the-shelf solutions: while some tasks can be considered 'solved', others required customized approaches;
5. Scalability: ie. Is it worth to develop a ML solution if I cannot easily scale it to several machines/production sites/customers?
6. ...

Anomaly/Outlier Detection

What is an anomaly/outlier?

'An outlier is an observation that deviates so much from other observations as to arouse suspicion that it was generated by a different mechanism' [1]

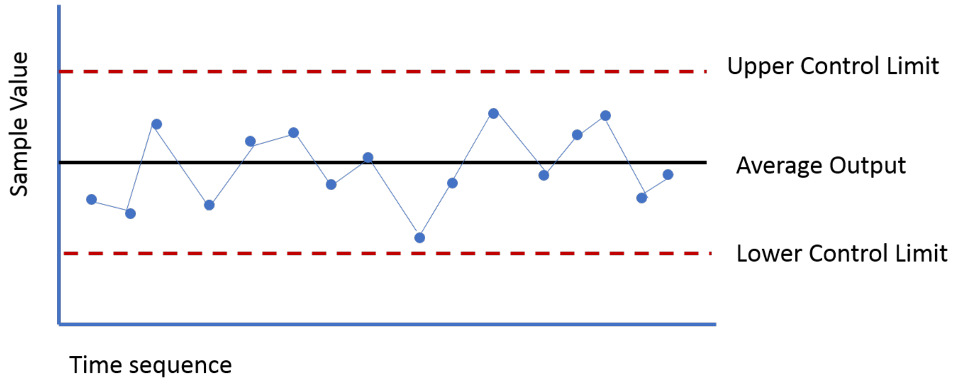
We want to detect anomalies for monitoring/quality purposes and trigger maintenance/corrective actions



[1] D. M. Hawkins, Identification of outliers, vol. 11., Springer, 1980

Process Monitoring in (most) Industries

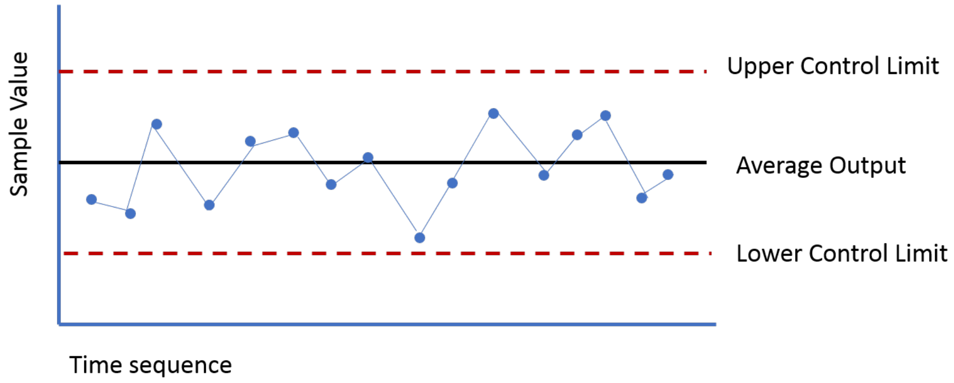
- One technique to rule them all...
- Univariate control charts (CC) are the standard approach to deal with process monitoring

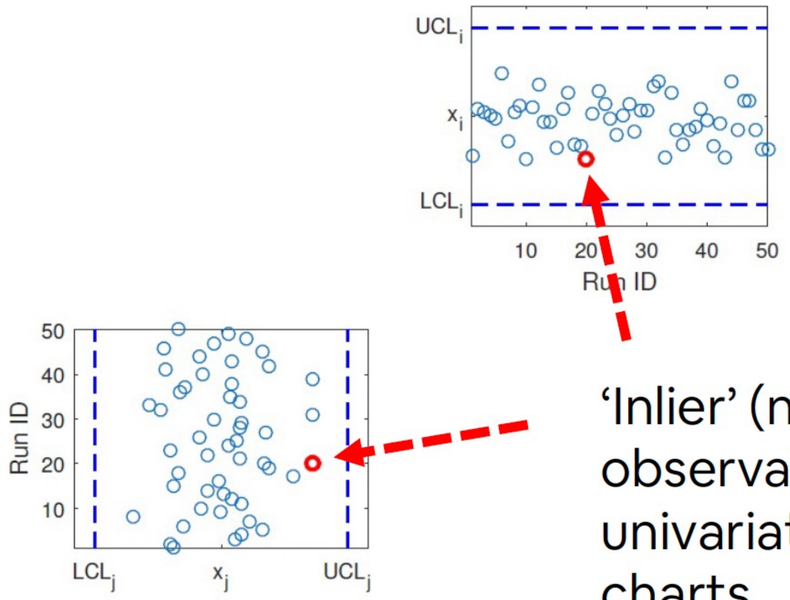


Process Monitoring in (most) Industries

- One technique to rule them all...
- Univariate control charts (CC) are the standard approach to deal with process monitoring
- Univariate CC Limitations:

1. They are unable to capture multivariate anomalies



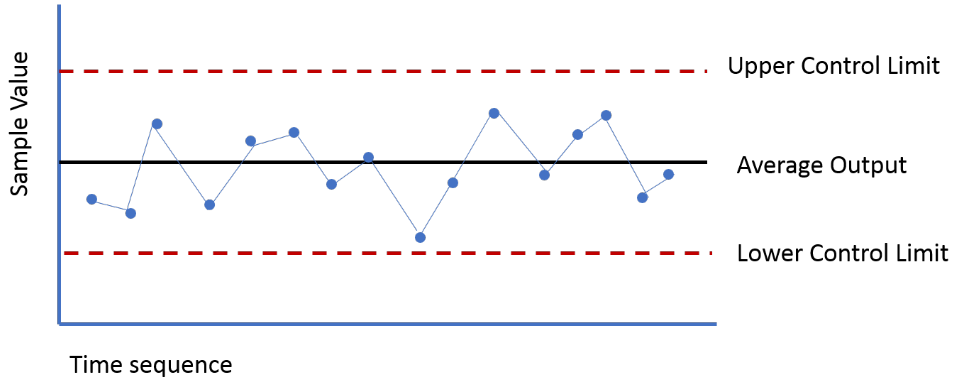


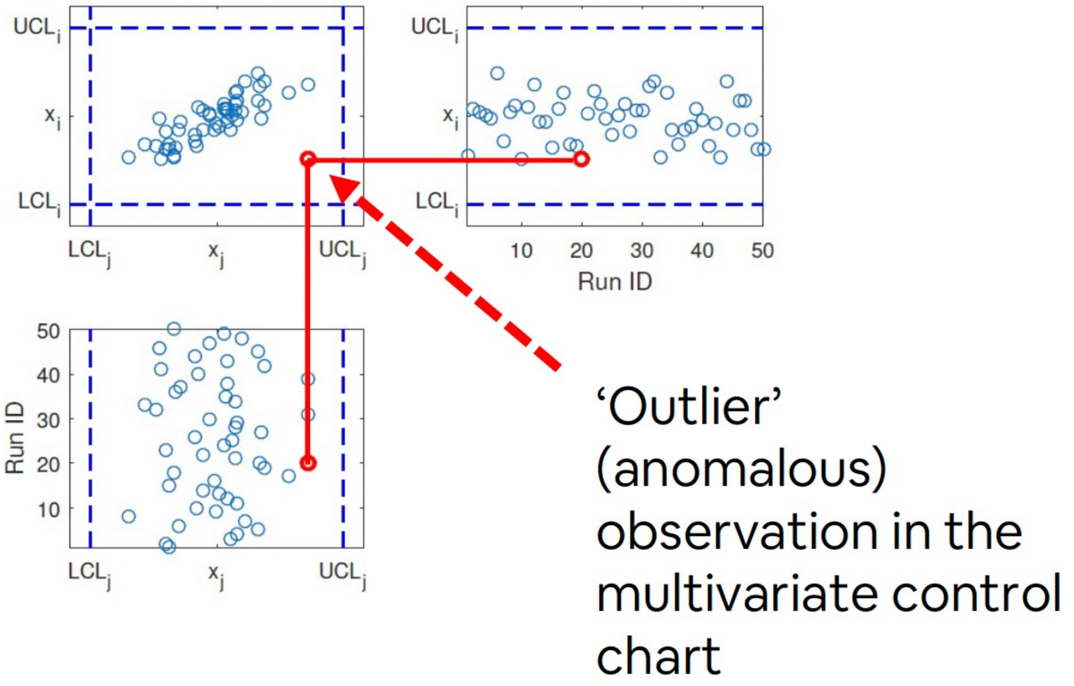
‘Inlier’ (normal) observation in univariate control charts

Process Monitoring in (most) Industries

- One technique to rule them all...
- Univariate control charts (CC) are the standard approach to deal with process monitoring
- Univariate CC Limitations:

1. They are unable to capture multivariate anomalies

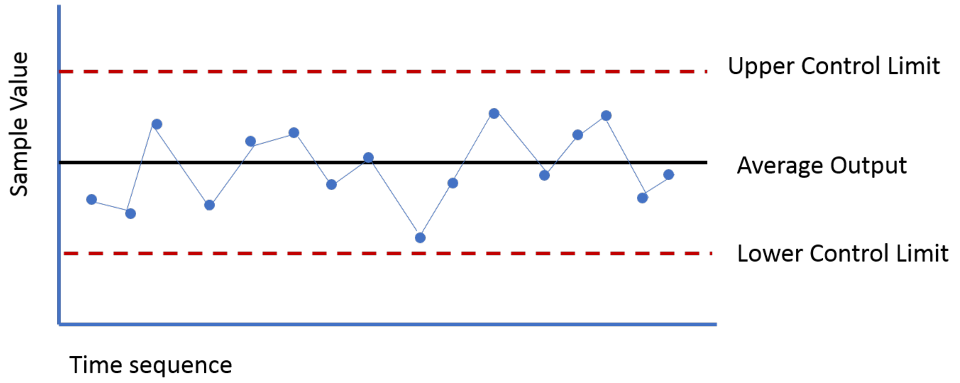




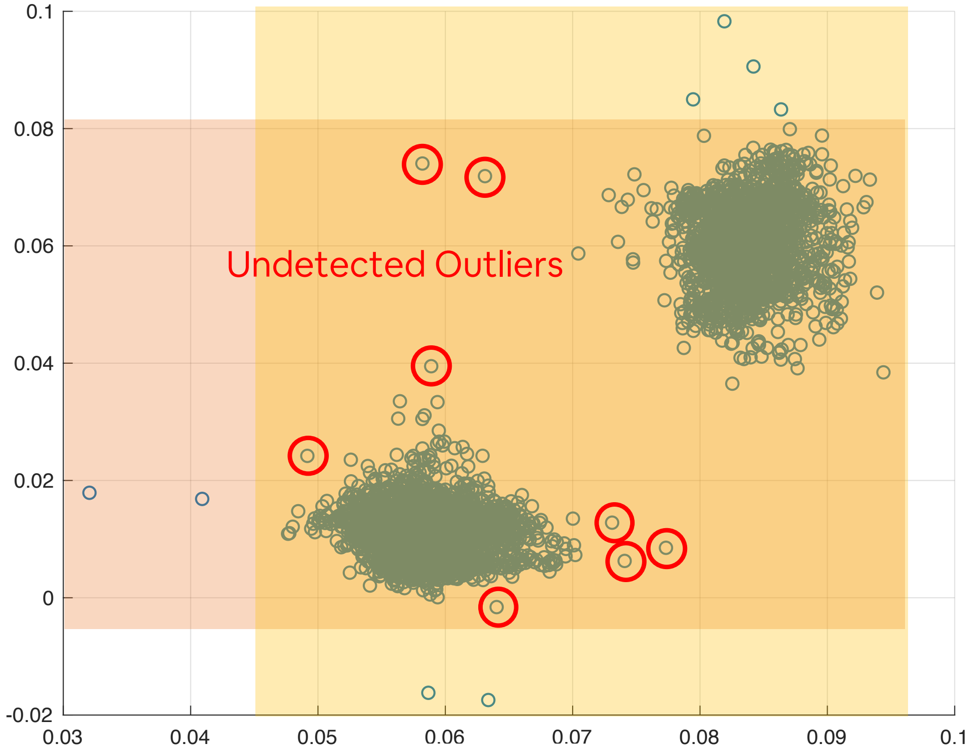
Process Monitoring in (most) Industries

- One technique to rule them all...
- Univariate control charts (CC) are the standard approach to deal with process monitoring
- Univariate CC Limitations:

1. They are unable to capture multivariate anomalies
2. They rely on Gaussian/unimodal distribution of underlying data



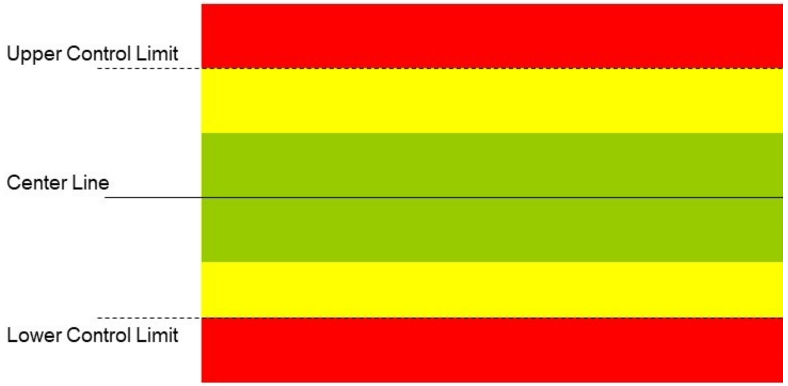
Real production data!



Process Monitoring in (most) Industries

- One technique to rule them all...
- Univariate control charts (CC) are the standard approach to deal with process monitoring
- Univariate CC Limitations:

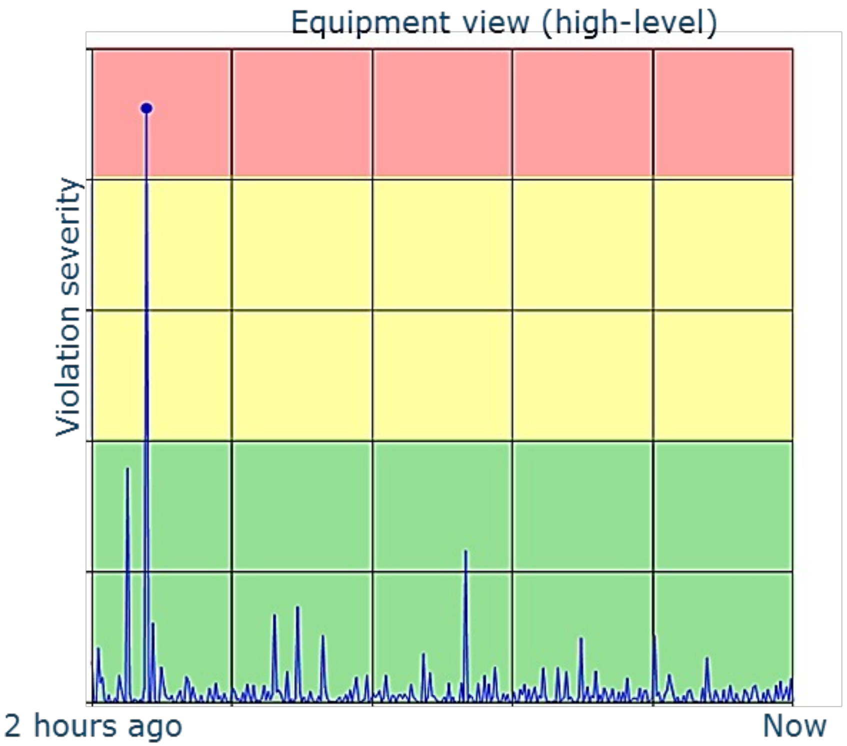
1. They are unable to capture multivariate anomalies
2. They rely on Gaussian/unimodal distribution of underlying data
3. Too many to handle!



Multivariate Unsupervised Anomaly Detection

Multivariate Unsupervised AD approaches provide 'anomaly scores': unique quantitative indicators able to represent the degree of 'outlierness' of complex systems with many variables

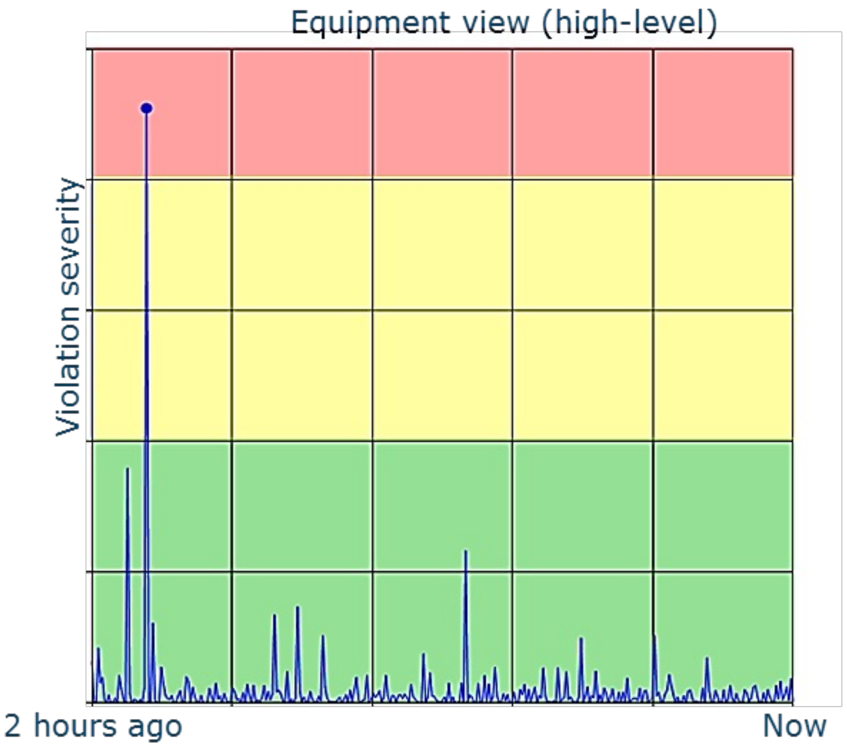
1. No labelled data are required
2. Dozens/Hundreds of sensors variables can be considered at the same time
3. No need for gaussian/unimodal distributions



Multivariate Unsupervised Anomaly Detection

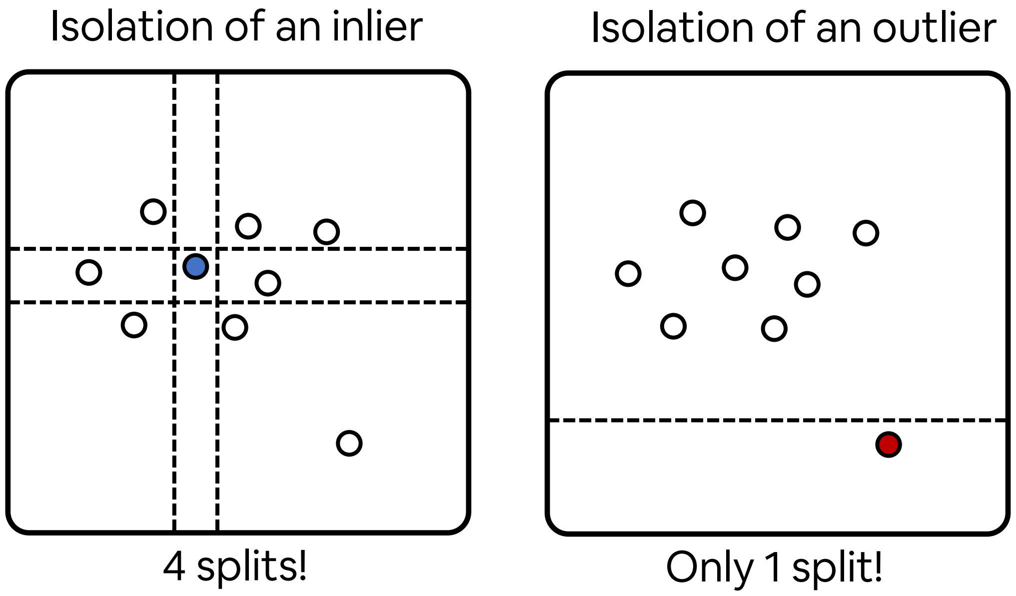
Many approaches for tabular data (data where rows are observations and columns are variables) [2]:

- Density-based methods (e.g. LOF, DBSCAN)
- Distance-based methods (e.g. kNN)
- Clustering-based methods (e.g. CBLOF)
- Neural Networks (e.g. Autoencoder)
- Isolation Forest
- ...



[2] PyOD (Python library for detecting outlying objects) https://pyod.readthedocs.io/en/latest/

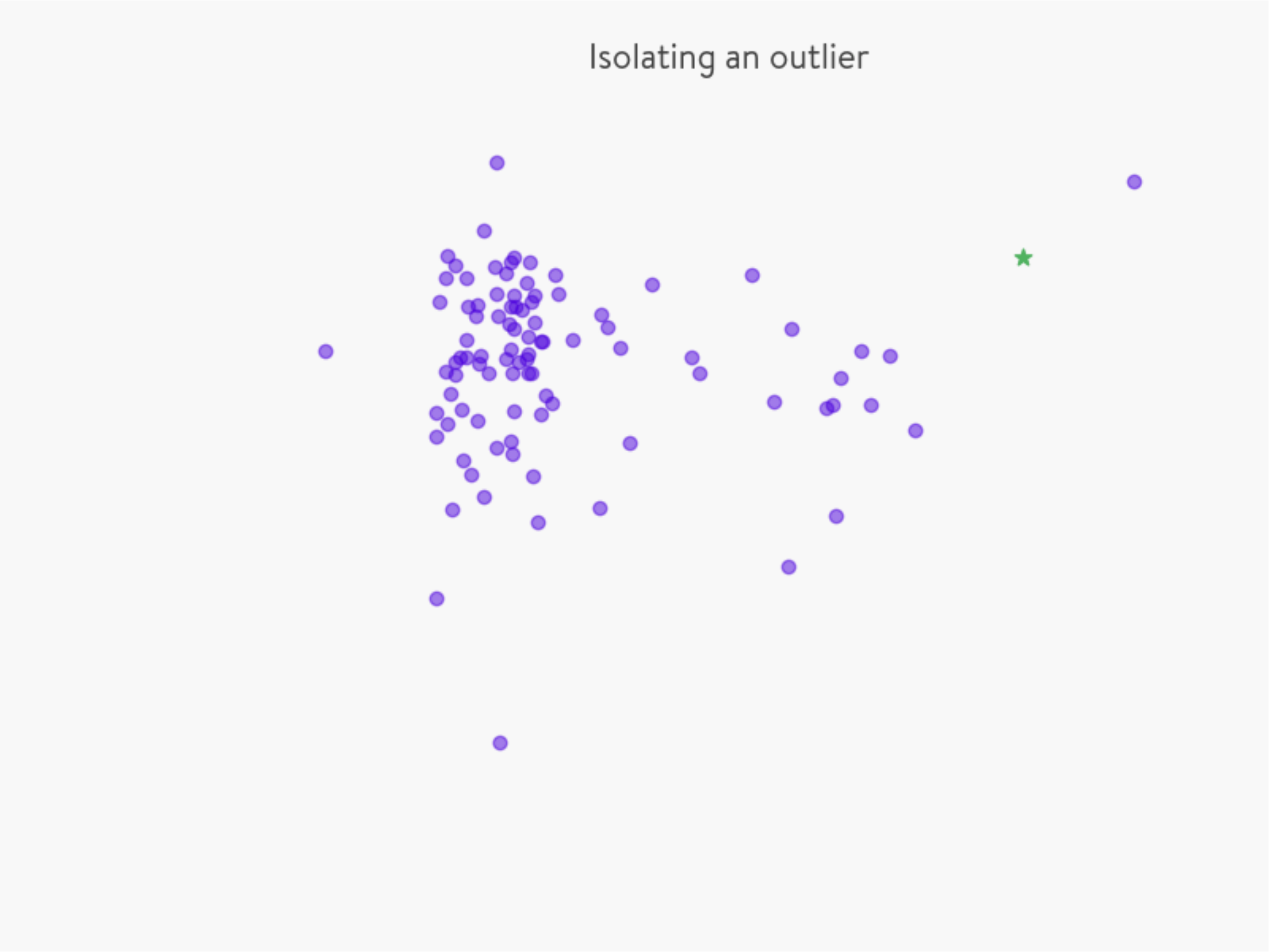
Isolation Forest [3]



- Efficient algorithm that outperforms other AD methods in several domains [4]
- Based on a partitioning procedure (that creates isolation trees) and on the idea that outlier and inlier are differently affected by such procedure

[3] Liu et al. (2012). Isolation-based anomaly detection. *ACM Trans. on Knowledge Discovery from Data 6*(1), 1-39.
[4] Ma et al. (2023) The need for unsupervised outlier model selection: A review and evaluation of internal evaluation strategies.

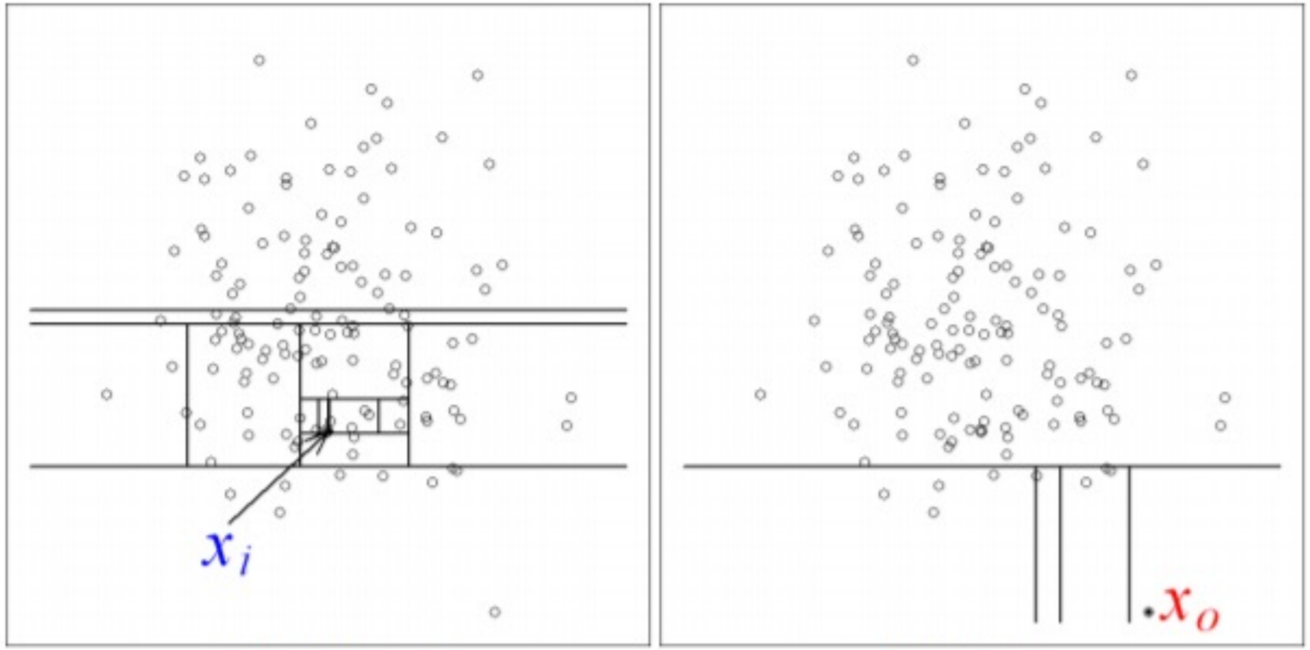
Isolation Forest [3]



- Efficient algorithm that outperforms other AD methods in several domains [4]
- Based on a partitioning procedure (that creates isolation trees) and on the idea that outlier and inlier are differently affected by such procedure

[3] Liu et al. (2012). Isolation-based anomaly detection. *ACM Trans. on Knowledge Discovery from Data 6*(1), 1-39.
[4] Ma et al. (2023) The need for unsupervised outlier model selection: A review and evaluation of internal evaluation strategies.

Isolation Forest [3]

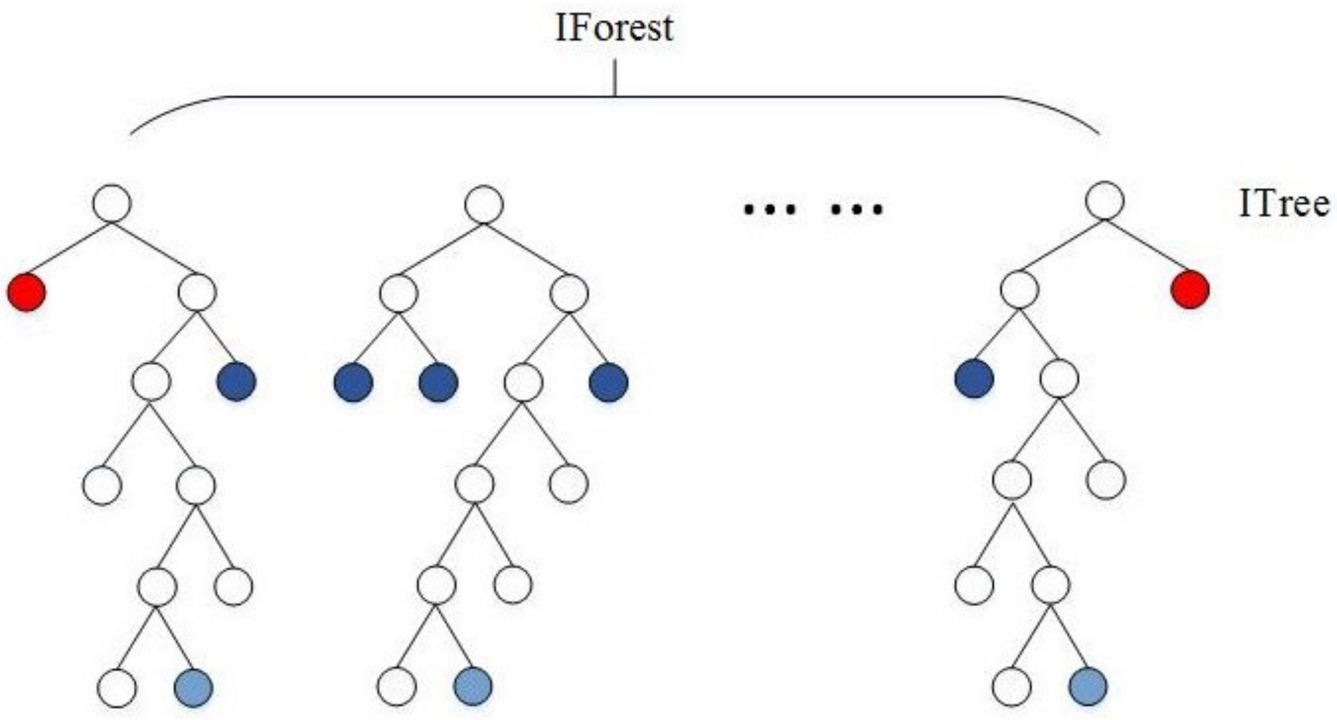


(a) Isolating x_i (b) Isolating x_o

- Efficient algorithm that outperforms other AD methods in several domains [4]
- Based on a partitioning procedure (that creates isolation trees) and on the idea that outlier and inlier are differently affected by such procedure

[3] Liu et al. (2012). Isolation-based anomaly detection. *ACM Trans. on Knowledge Discovery from Data 6*(1), 1-39.
[4] Ma et al. (2023) The need for unsupervised outlier model selection: A review and evaluation of internal evaluation strategies.

Isolation Forest [3]



- Efficient algorithm that outperforms other AD methods in several domains [4]
- Based on a partitioning procedure (that creates isolation trees) and on the idea that outlier and inlier are differently affected by such procedure
- An ensemble approach: anomaly score computed as mean of the depth over the various isolation trees

[3] Liu et al. (2012). Isolation-based anomaly detection. *ACM Trans. on Knowledge Discovery from Data 6*(1), 1-39.
[4] Ma et al. (2023) The need for unsupervised outlier model selection: A review and evaluation of internal evaluation strategies.

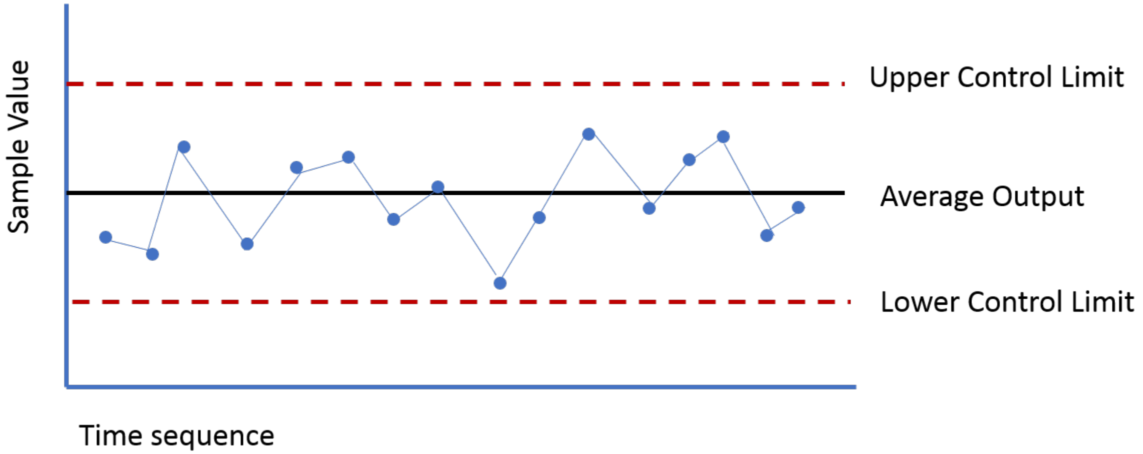
Univariate Control Chart vs Unsupervised AD

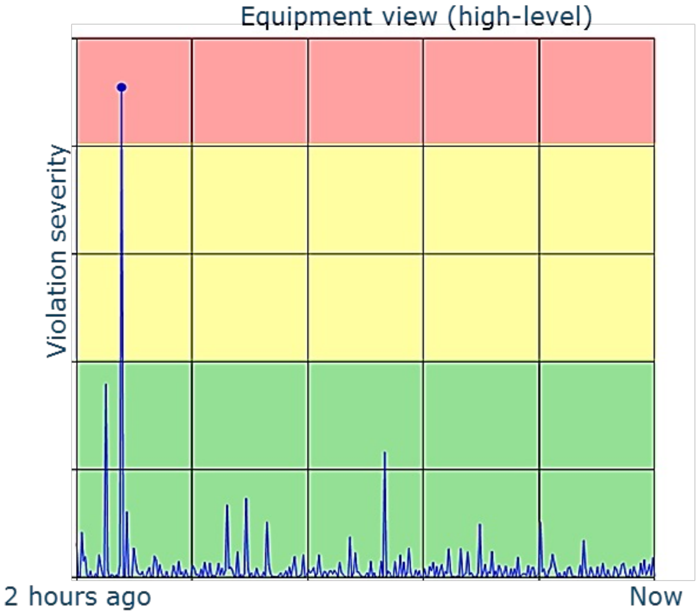
To enable Decision Making information should be:

complete

concise

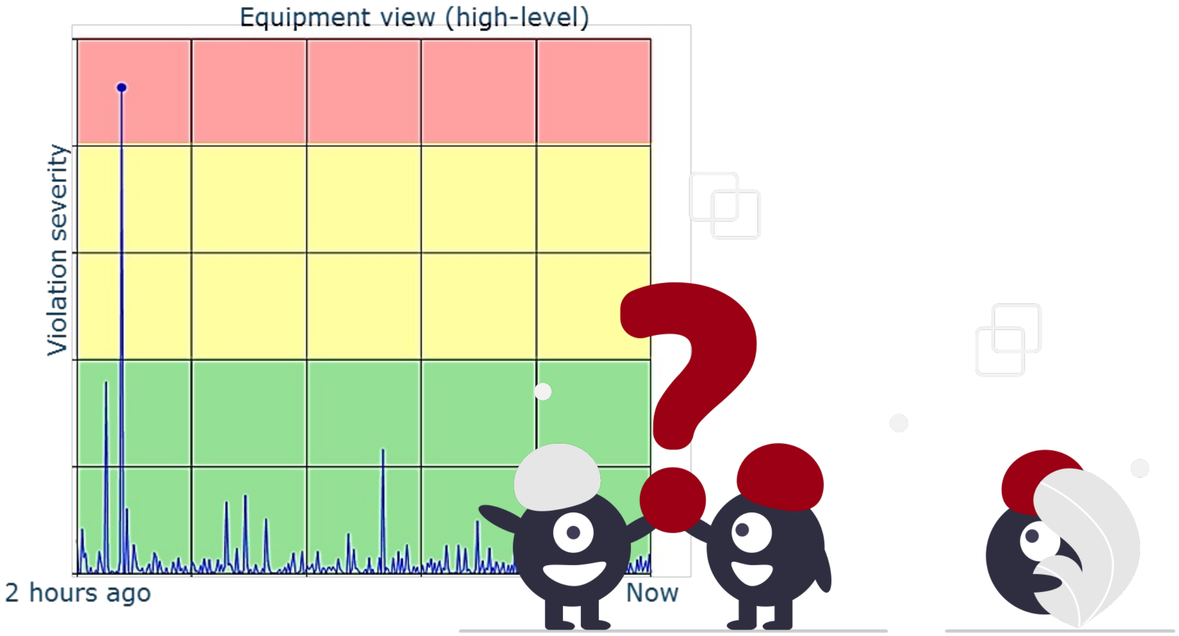
interpretable

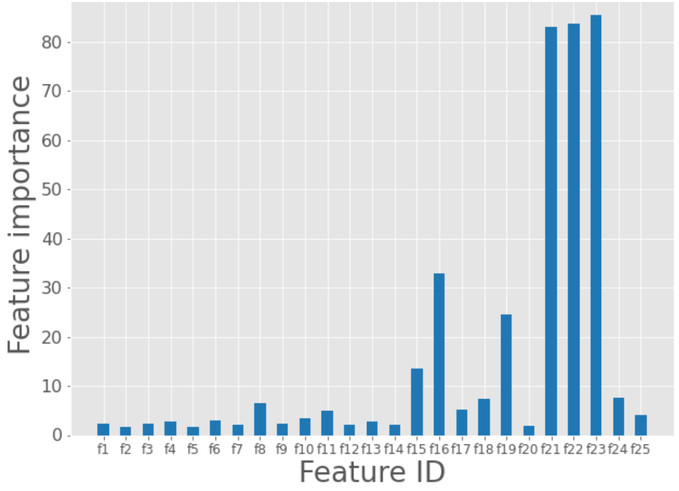




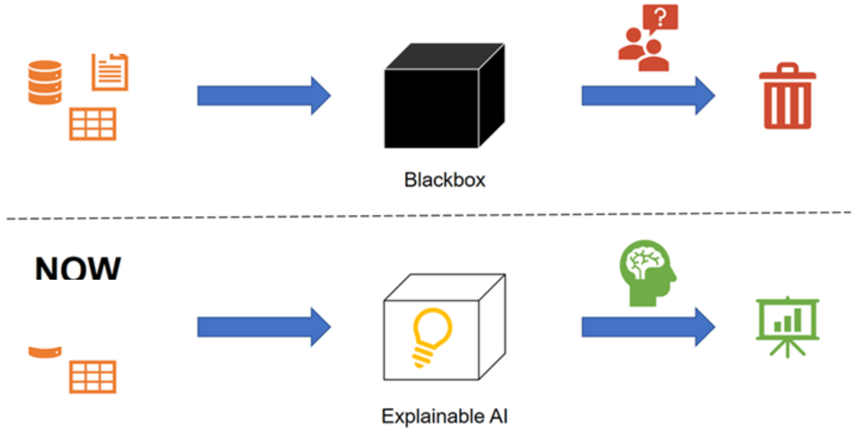
I've got the Anomaly Score: now, what?

- Thanks to the Anomaly score users are alerted of potential anomalous situation, however it is up to them to discover potential troubles
- It would be nice to ease the Root Cause Analysis to provide additional information, like feature rankings...





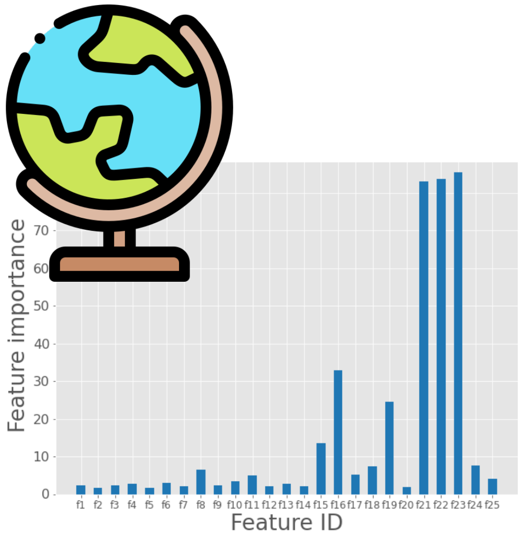
Depth-based Isolation Forest Feature Importance (DIFFI) [5]

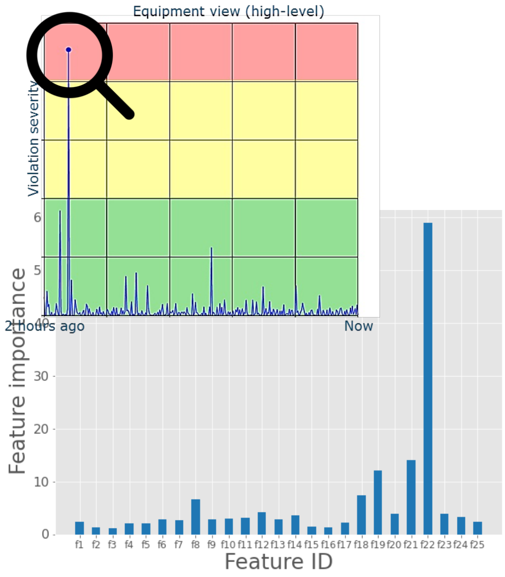


DIFFI is an eXplainable Artificial Intelligente (XAI) approach designed for the Isolation Forest

DIFFI provides a variable ranking for:

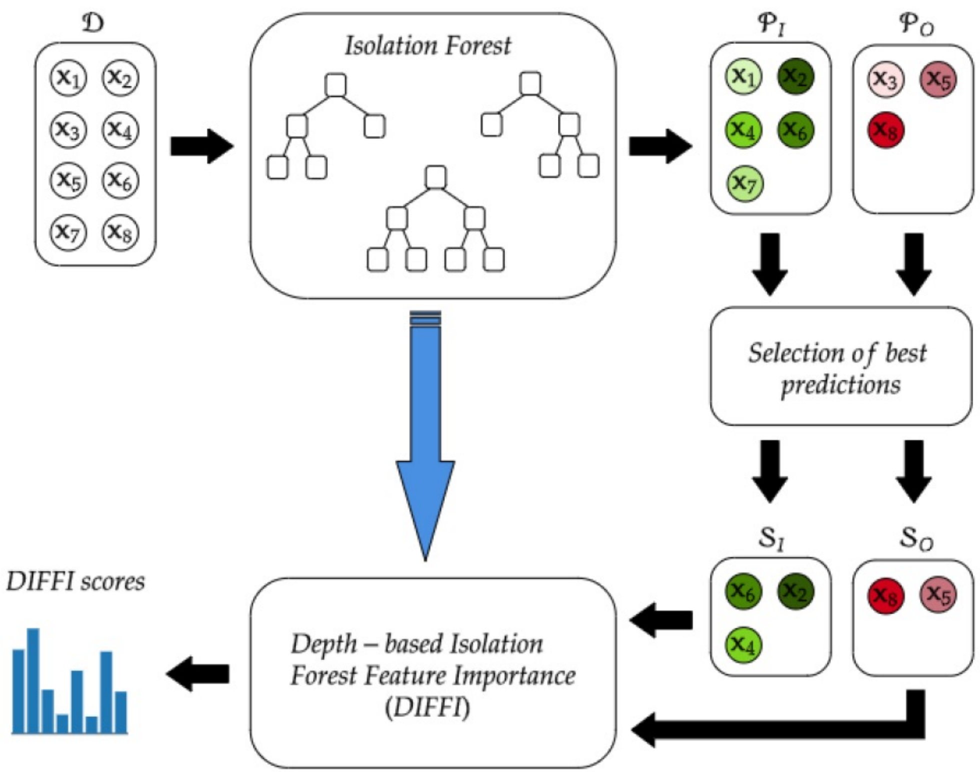
- Global Explainability (ie. what variables are important for the whole Isolation Foreset model)
- Local Explainability (ie. what variables are important for a particular prediction)





[5] Carletti, M., Terzi, M., & Susto, G. A. (2023). Interpretable Anomaly Detection with DIFFI: Depth-based feature importance of Isolation Forest. *Engineering Applications of Artificial Intelligence*, *119*, 105730.

Depth-based Isolation Forest Feature Importance (DIFFI) [5]



DIFFI provides a variable ranking that:

- Does not require true labels (other XAI approaches do!)
- Low computational cost
- No tuning

IDEA: mark a feature as "important" if

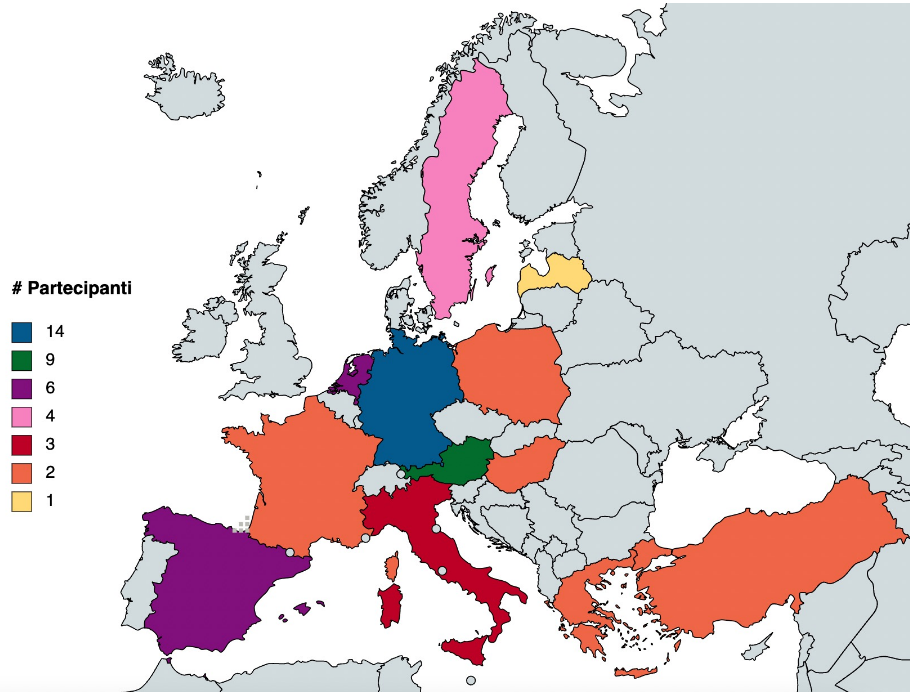
- it induces isolation of outliers at small depths (i.e. near the root)
- At the same time, does not contribute to the isolation of inliers

[5] Carletti, M., Terzi, M., & Susto, G. A. (2023). Interpretable Anomaly Detection with DIFFI: Depth-based feature importance of Isolation Forest. *Engineering Applications of Artificial Intelligence*, *119*, 105730.

For technical details

AIMS5.0 (AI for Sustainable Manufacturing): case study in semiconductor manufacturing

- EU Project on AI for Manufacturing
- 53 partners, 12 countries
- Main collaborations: LFoundry, Infineon, Statwolf, IDEKO



Dashboards Ride Monitoring with Anomaly Detection

Boss Statwolf

Timeframe

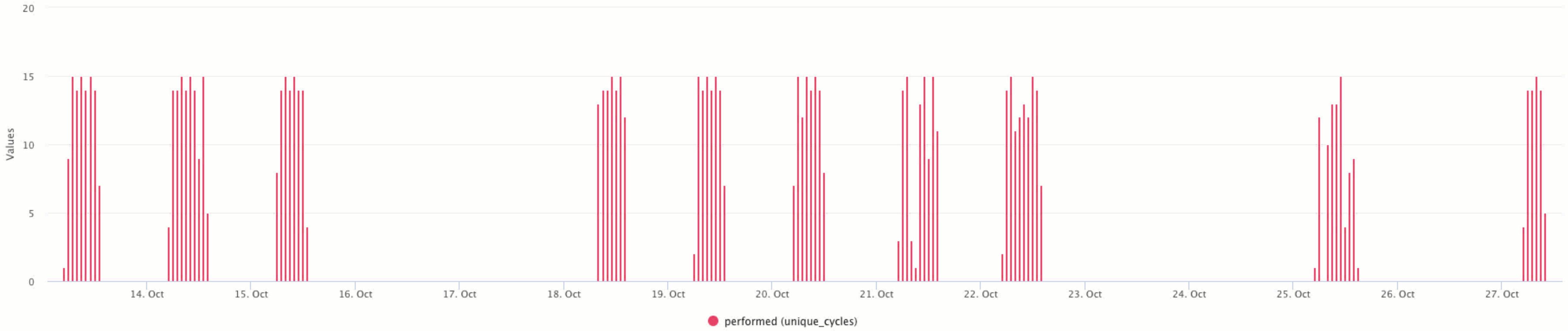
All Time

Capacity

Capacity is measured in number of cycles.

980	980	0	0
Total Cycles	**Performed Cycles**	**Stopped Cycles**	**Estop Cycles**

Total Hourly Capacity



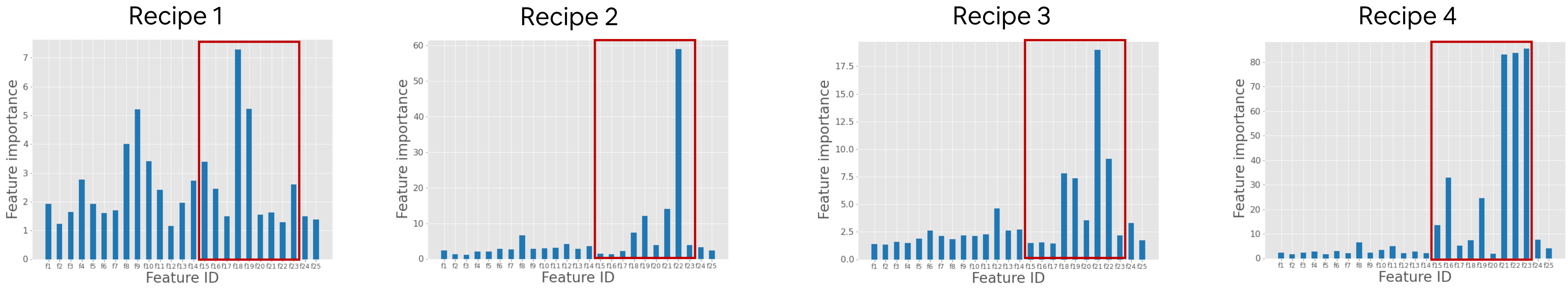
Duration

Duration is the sum of Working Time, Stopped Time, Downtime and Waiting Time.

DIFFI (global) - Evaluation

EXPERIMENTAL RESULTS: Anomaly Detection for Chemical Vapor Deposition monitoring problems (102k observations, 25 features, 4 recipes) [6]

Prior knowledge: based on domain knowledge, the indices of the most important features (most of which has a physical interpretation) are 15, 16, 17, 18, 19, 20, 21, 22, 23.

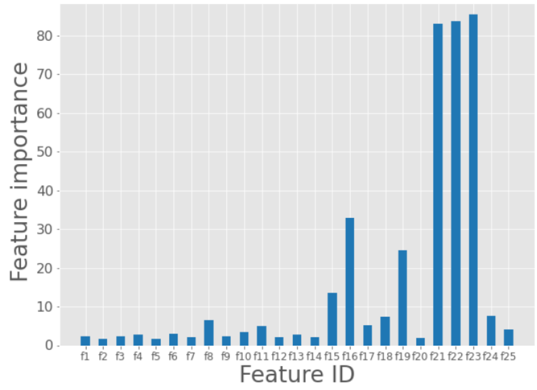


[6] Carletti, M., Maggipinto, M., Beghi, A., Susto, G. A., Gentner, N., Yang, Y., & Kyek, A. Interpretable anomaly detection for knowledge discovery in semiconductor manufacturing. In Winter Simulation Conference (WSC) (pp. 1875-1885). IEEE.

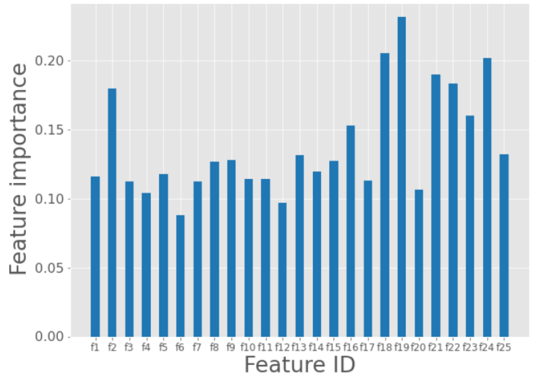
DIFFI (global) - Evaluation

EXPERIMENTAL RESULTS: Anomaly Detection for Chemical Vapor Deposition monitoring problems (102k observations, 25 features, 4 recipes)

Prior knowledge: based on domain knowledge, the indices of the most important features (most of which has a physical interpretation) are 15, 16, 17, 18, 19, 20, 21, 22, 23.



DIFFI



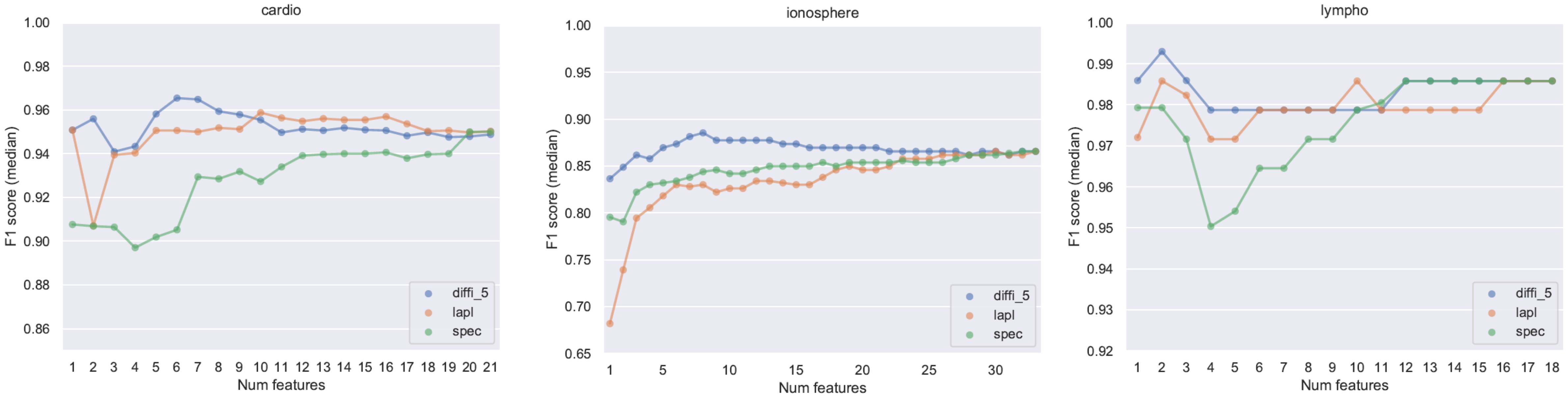
Permutation Importance (PIMP)

Table 1: Execution times comparison: DIFFI vs PIMP.

Dataset	Num samples	DIFFI exec time [s]	PIMP exec time [s]
Recipe 1	956	0.37	9.12
Recipe 2	34716	0.41	272.38
Recipe 3	3580	0.38	18.75
Recipe 4	13459	0.38	55.73

DIFFI (global) - Evaluation

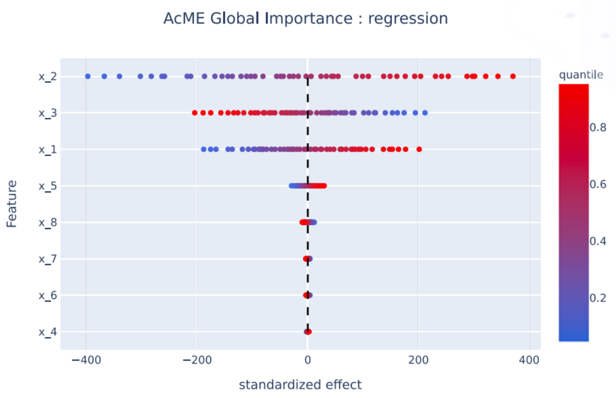
EXPERIMENTAL RESULTS: Using as proxy task, unsupervised feature selection

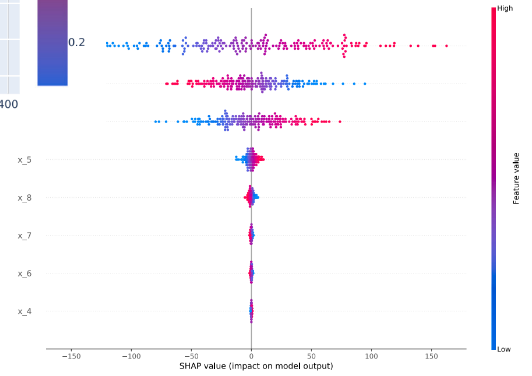


We are comparing with Laplacian Score (He et al. 2006) and SPEC (Zhao et al. 2007) on a tagged dataset

Conclusive remarks

- Starting now with advanced diagnostic? Unsupervised anomaly detection is the perfect entry point!
- XAI approaches can be a fundamental tool for the adoption of AI solutions. Don’t forget about the users!
- If you have unsupervised monitoring tasks when you need interpratibility (or you want fast interpretability) try DIFFI or let us know!
- Users will also need information in valuable time
- We developed approaches that can provide SHAP-like explanations in ‘reasonable’ time [7] to foster adoption of AI-based decision support systems





[7] Dandolo, D., Masiero, C., Carletti, M., Dalle Pezze, D., & Susto, G. A. (2023). AcME—Accelerated model-agnostic explanations: Fast whitening of the machine-learning black box. *Expert Systems with Applications*, *214*, 119115.

Thank you! Danke! Grazie!

eXplainable Artificial Intelligence (XAI) in Industrial Diagnostic: Interpretable Unsupervised Anomaly Detection

Gian Antonio Susto

Associate Professor, Università degli Studi di Padova, Italy

September 3rd, 2024

@ Lausanne – 8th Intelligent Maintenance Conference

Bonus Slides

DIFFI (global)

1) For each feature f, we compute a *Cumulative Feature Importance* for outliers ($I_O(f)$) and for inliers ($I_I(f)$).
We iterate over all trees of the forest and over all predicted outliers and inliers, respectively. Each time feature f appears in the path, update $I_O(f)$ or $I_I(f)$ according to the following update rules

For predicted outliers

$$I_O(f) \longleftarrow I_O(f) + \boxed{\frac{1}{d_{leaf}(i,t)}} \cdot \boxed{\lambda_O(v)}$$

Depth of the leaf node for *i-th* sample in t-*th* tree

Imbalance induced by the split

For predicted inliers

$$I_I(f) \longleftarrow I_I(f) + \frac{1}{d_{leaf}(i,t)} \cdot \lambda_I(v)$$

2) Compute the global feature importance score for the generic feature f:

$$GFI(f) = \frac{I_O(f)}{\boxed{C_O(f)}} \Big/ \frac{I_I(f)}{\boxed{C_I(f)}}$$

Counters: # of times feature *f* appeared in the paths of outliers or inliers

Local-DIFFI

Local variant of the global DIFFI method to produce local feature importance scores for predicted outliers

- Update rule: $I_O(f)$ $I_O(f) + \frac{1}{d_{leaf}(i,t)} - \frac{1}{d_{max}}$

  where d_{max} is the maximum depth of the isolation trees

- Local feature importance score: $LFI(f) = \frac{I_O(f)}{C_O(f)}$

No imbalanced induced, no inlier/outlier (we are not training on this data point)

Local-DIFFI - Evaluation

EXPERIMENTAL RESULTS: glass dataset

Regular points: windows glass (classes 1-4 of the original dataset)

Anomalous points: headlamps glass (class 7 of the original dataset)

Feature "RI": refractive index

Features "Na", "Mg", "Al", "Si", "K", "Ca", "Ba", "Fe": sodium, magnesium, aluminum, silicon, potassium, calcium, barium, iron concentrations

Prior knowledge: headlamps glass should have high concentrations of aluminum (used as a reflective coating) and barium (heat resistant properties)

Local-DIFFI - Evaluation