Question: What experience on improving the performance of machine learning-based systems does exist in terms of addressing the anomalies (rare event with significant consequences), which may pose considerable impact on technological and/or economic performance of the power system? How should we distinct anomalies and data outliers between each other?

Improving the performance:

The widespread use of information and telecommunication technology in power systems has produced vast amounts of data that improves system observability, monitoring and controllability. Artificial intelligence (AI) technology has made great progress in many industries. And thus, over several decades, research on anomaly mining has received increasing interests due to the implications of these occurrences in a wide range of disciplines. Anomaly detection, which aims to identify rare observations, data points, events, or observations that are being different from the rest of the data points or observations, is among the most vital tasks in the world, and has shown its power in preventing detrimental events. They are patterns in the data that do not conform to expected behavior. As such, anomalies represent either noise or outliers in the data. As these outliers deviate from the nominal data, they typically translate to some problems. The detection task is typically solved by identifying outlying data points in the feature space and inherently overlooks the relational information in real-world data.

Historically, anomaly detection was solely utilized to find and remove anomalies from the data, thus ensuring good quality of data before feeding them to models. However, over time, the interest grew for investigating the anomalies themselves, what they represent, and what caused them. With advances in AI and increasing importance for anomaly detection and prevention in various domains, artificial neural network approaches enable the detection of more complex anomaly types while considering temporal and contextual characteristics. And there are a number of results shown from the application of various anomaly detections (from simple to complex deep learning based approaches) using machine learnings. An example below:

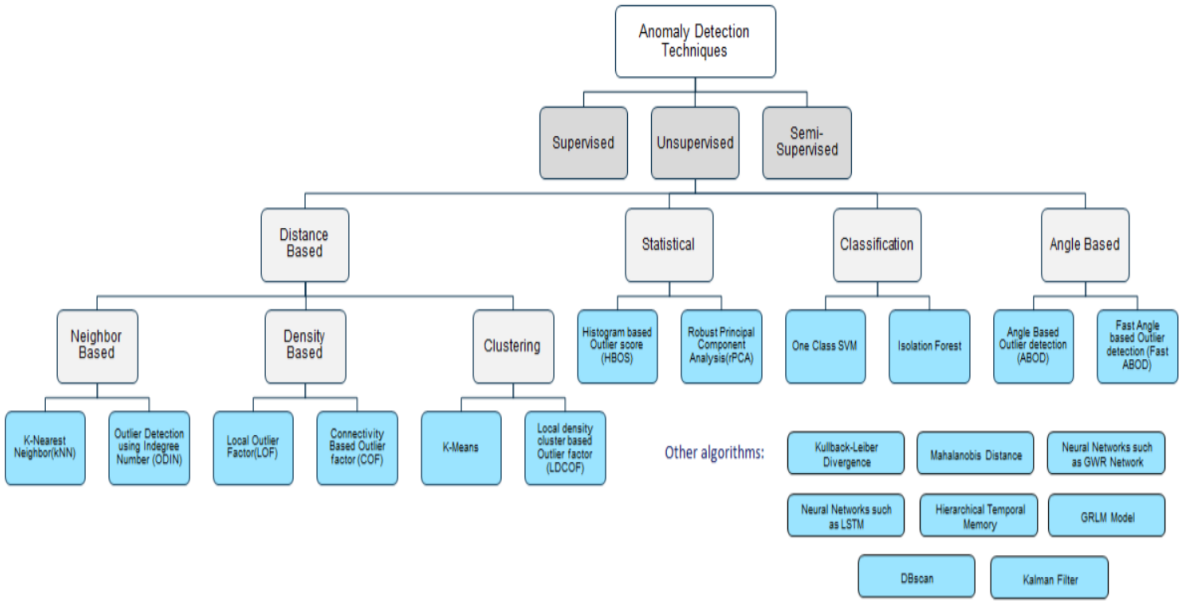

Fig 1. Various anomaly detection applied

Thus, the Organization across industries face a key dilemma to solve the problem:

- there's a set of equipment /system able to record in-time values of data from sensors.
- On other side, the industrial company targets unplanned downtime & improve productivity by using the data captured above.

To solve this the key questions to be addressed are as follows:

- Where is the abnormal behavior & when it occurs?
- Where and when have failures happened?
- How to predict unplanned failure before occurrence?

Data anomaly detection relies on the assumption that anomalies are rare events, and they differ considerably from normal behavior. Detection process needs a context of normal behavior to recognize any abnormal value. Time series data provides the context with a sequence of values over time. This context establishes a baseline for a normal behavior, to identify unusual patterns or outliers. Enterprise data anomaly detection works with three different setting- point, contextual & collective anomalies. Machine learning algorithms accelerate the process and improve accuracy over a period through iterative learning. Thus, the experience shows the adoption of the following to improve performance with considerable impact both in research & practice:

- Noise and poor data quality: affect distinguishing outliers from normal records, reducing the effectiveness of anomaly detection in identifying real outliers
- Data preprocessing- is the crucial step for treating data before it is fed to the model
- Modeling normal behavior with acceptable variance helps identify anomalies more accurately.
- Streaming data volume affects the processing speed and thus it is essential to detect data drift and outliers in real time to provide early warnings
- Deeper understanding of the data- by collaborating with the domain experts (not just the AI experts), for connecting data, insights, and algorithms uncovers a deeper understanding of data to identifying anomalies correctly
- Finally, choosing the right model & its architecture (or ensemble of models) with for the specific data being used. It's clear that the same methods do not perform well in all cases, thus, cannot be generalized across different use-cases in the power system data.
- Also, the performance metrics is specific to the context of the analysis

Anomalies and data outliers:

The taxonomy of anomalies applies for all investigated application areas: It can be characterized based on various aspects such as focus point (e.g. a certain actuator of a production machine), measurability, (non-)linearity and temporal behavior. Depending on the application, different focus points are possible. It can imply e.g. a direct affectation of the whole system's dynamics or an interference on the level of (individual) sensors and actuators. Anomalies can either be directly measurable or they have to be observed using some kind of indirect state estimation. Furthermore, they can either show linear or nonlinear characteristics.

Outliers are data instances which do not seem to readily fit the behavior of the remaining data or a resulting model. Though many machine learning algorithms intentionally do not take outliers into account, or can be modified to explicitly discard them, there are times when outliers themselves are where the money is. In reality, there is no universal definition of an outlier; could you imagine trying to define a specific rule delineating what would be geographically "too far" to be true, and which would apply to all fraud detection scenarios

similar to the case above? Even if we can agree what an outlier is, depending on the application we may not want to remove it, just be aware of its presence.

But even saying that we are interested in being notified of detected outliers, there are all sorts of ways to go about doing this like using simple descriptive statistical methods such as identifying low-dimensional data points which fall a particular multiple of standard deviations from the mean in the normal distribution. It can be adopted if that works for the specific problem and its data. But there are plenty of other approaches to be adopted.