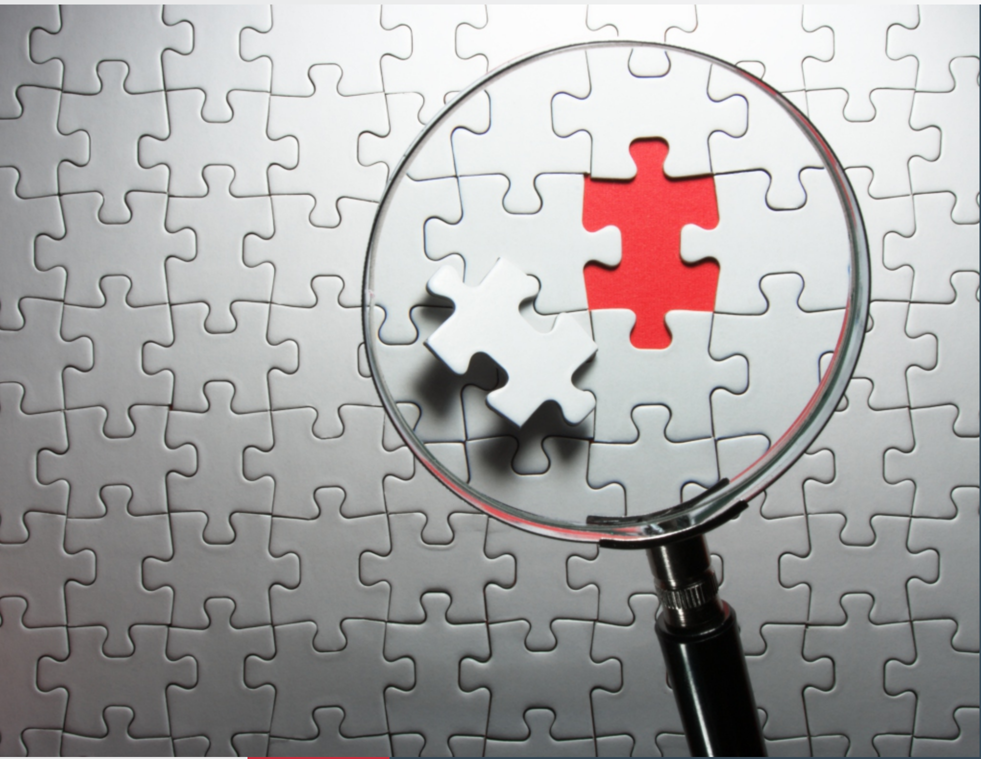
STATWOLF

ANOMALY DETECTION AND INTERPRETABILITY

An introduction to Anomaly Detection & Interpretability in real-time monitoring and business intelligence scenarios



ANOMALY DETECTION AND AI TRANSFORMATION

Anomaly detection is often the entry point for business digital transformation. Collecting logs and data from everyday activities is the first thing to do to embrace a data-driven strategy.

Once the data collected are enough, without the need for data supervision, Anomaly Detection provides a monitoring system that not only will be part of decision support system, but also allows to build more innovative and business impactful tools such as Predictive Maintenance, Fault Detection and Classification

Statwolf has developed solutions and efficient algorithms to capture anomalies in real-time monitoring scenarios and in Business Intelligence applications, allowing users to understand the black-box suggestions and to make on-the-fly root cause analysis.

Want to know more about Anomaly Detection and Interpretability in Anomaly Detection? Let us take to a quick introductory journey on the topic!

INTRODUCTION TO ANOMALY DETECTION

Anomaly detection is the process of identifying unexpected items, rare events, or strange patterns in data.

With the exponential growth of available data, industries can use data science tools to detect observations that are not aligned to normal data flow, which raise suspicions because of their significant differences from the majority of the data.

The difference found in the abnormal points can be a symptom of mutation in the data generation process (permanent, trend change or temporary), or simply identify an outlier.

For instance, Anomaly Detection can be used to identify

- defected units using quality assessment tests or machine logs;
- cyberattack attempts by looking at network traffic;
- fraud detection considering the transaction log.

Machine Learning algorithms are trained to understand which data patterns are normal: in this way, when 'strange' data come along, the Machine Learning module can spot them!

Three major categories of anomaly detection techniques exist: see them in the next pages!

1

SUPERVISED ANOMALY DETECTION

Supervised anomaly detection requires the availability of a historical dataset that contains both normal and anomalous data points, correctly labelled. Typically, supervised methods provide a better rate of anomaly detection thanks to their ability to encode any interdependency between variables and including previous data in any predictive model.

Some well-known algorithms that can be used for supervised AD:

Logistic Regression, Random Forest, Neural Network, SVC

2

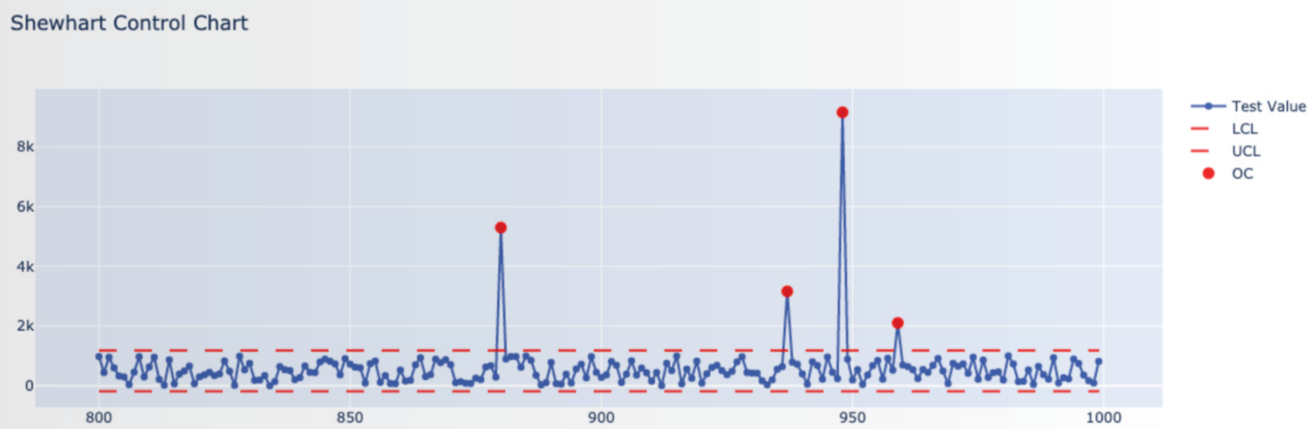
SEMI-SUPERVISED ANOMALY DETECTION

Semi-supervised anomaly detection techniques compute a set of statistical measures about the normal behaviour based on a given training data set consisting of normal points only. Then, new instances are tested to decide whether they have been generated from the same data flow or they are outliers, using Statistical Process Control (SPC).

SPC is a quality standard that is widely used in the manufacturing process. Item data are retrieved during the manufacturing run-time process and are compared to un upper and lower limit to determine if they are in the acceptable range of values, or they fall outside.

Some examples of semi-supervised AD algorithms:

Statistical Tests, Univariate Control Charts, Multivariate Control Chart



3

UNSUPERVISED ANOMALY DETECTION

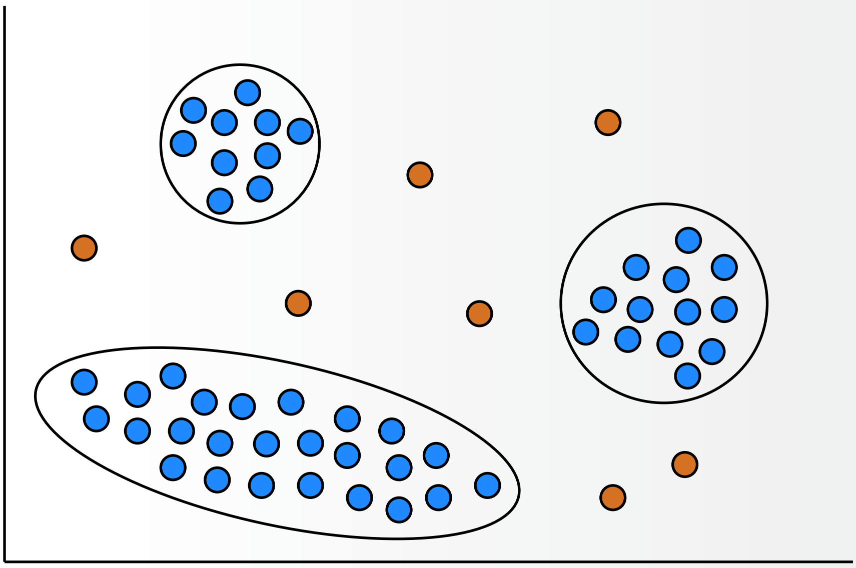
Unsupervised methods are used when there is no information about which observations are normal and which are not. These methods are based on the statistical assumption that most of the data are normal, and only a few observations aren't.

The most classic technique adopted in the unsupervised case is clustering: starting from the multivariate analysis of the dataset, it allows us to select and group similar elements into the same cluster. Then, cluster analysis allows us to place in the same set the elements that have the highest degree of similarity and to separate those that have the highest level of difference. The clustering procedure then detects the items that do not match with any cluster and label them as anomalous.

A newer, more sophisticated technique is the Isolation Forest. This algorithm is based on decision-tree aggregation, which isolate - as the name suggests - each point in the dataset based on its greater or lesser distance from other points. The basic principle is that abnormal observations are easier to isolate than normal ones.

Some examples of unsupervised AD algorithms:

Kmeans, Isolation Forest, PCA, TSNE

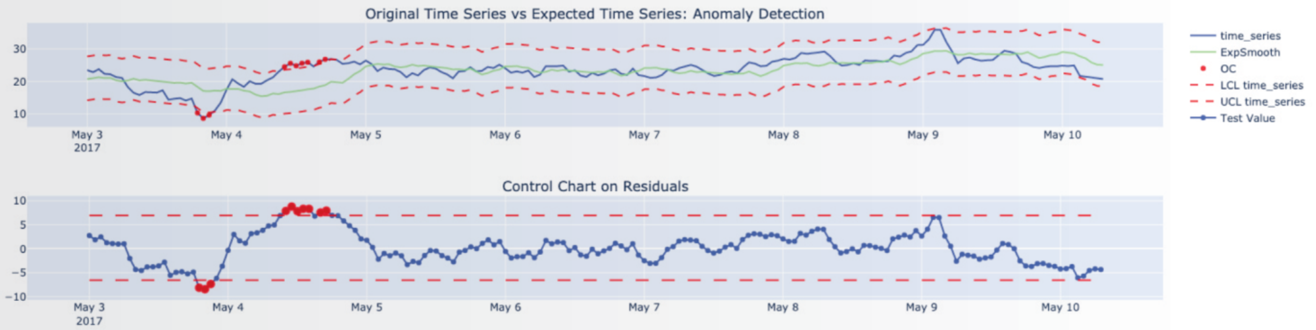


TIME SERIES ANOMALY DETECTION

Anomaly detection for time-series is a challenging AD task, because of the natural behaviour of the data series. Trend and seasonal components should be removed from the collected values. Alternatively, a time series smoothing or forecasting algorithm could be used instead of trend-seasonal decomposition. Usually, to address AD for time series we first learn a model to represent the series, then we apply classic AD techniques to prediction residue.

Some examples of AD algorithms for time-series:

Recurrent Neural Network, ARIMAX, Exponential Smoothing combined with Semi or Unsupervised Anomaly detection



INTERPRETABILITY FOR ANOMALY DETECTION

Having solved the problem of finding abnormal observations in our data, the remaining question is: why are they different from normal data?

To answer, it is necessary to understand why the algorithm has deemed that observation as abnormal. This branch of ML is called Interpretability. Interpretability is the degree to which a human can consistently predict the model's result. The higher the interpretability of a machine learning model, the easier it is for someone to comprehend why certain decisions or predictions have been made. Notice that interpretability is crucial to make ML a valuable resource for decision support.

Algorithms that are simple to understand for humans (a.k.a. white-box-models) are widely used in industries but in some cases, they do not succeed to catch up with the necessary performances, so more complex but at the same time more powerful algorithms are used (a.k.a. black-box-models). In these cases, an explanatory algorithm is applied to give users knowledge about the model decision function.

In addition to understanding the general operation of the model, it is possible to study in detail the individual predictions of the model to understand why each observation is anomalous.

If a production line has a high number of defects and the reason is not clear, an abnormal production state will recur. If a unit is labelled as an outlier and the reason is not clear, then it's not possible to proceed to a repair and therefore to a recovery of the produced unit. In other words, to turn ML results into actionable insight we need root cause analysis, and interpretability can come to the rescue.

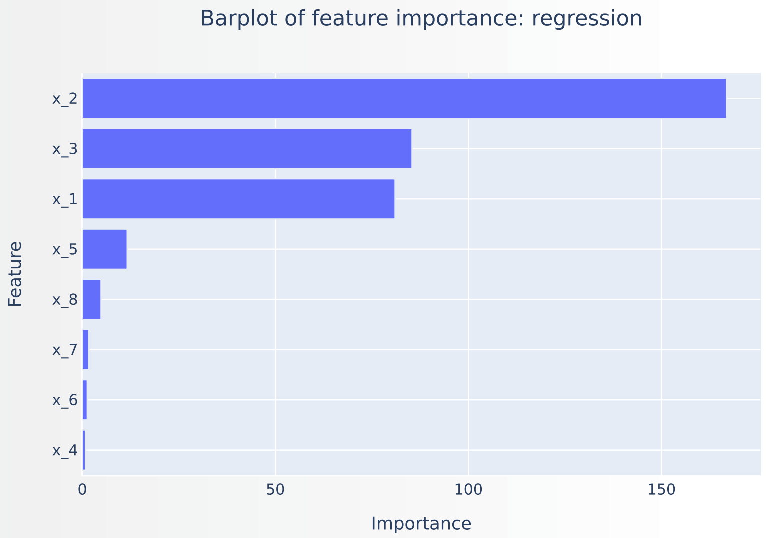
For instance, interpretability can provide an importance score for each feature, or a what-if approach that results in a simple and comparable explanation. This information can be turned into corrective actions to improve the production process, detect components of a machine that are close to failure or find the cause of a defect.

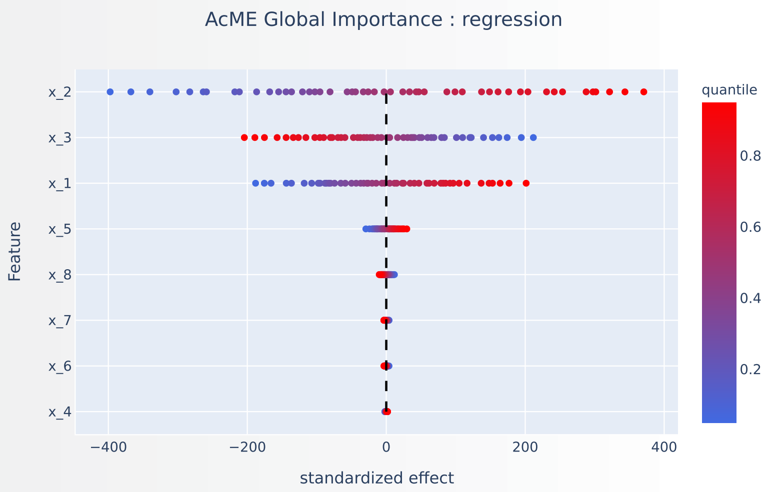
Some examples of interpretability techniques:

SHAP, Permutation Importance, AcME, LIME

ACME

AcMe is a new model interpretability technique developed by Statwolf. It is a simple and fast procedure that studies the model behaviour observing the changes in the model output caused by using different quantiles of each input feature. To evaluate the impact of the outputs' changing, we introduce a new measure, named standardized effect, that keeps in count both the changing direction and the overall variable impact amplitude. Standardized effects are also used to compute the final scores that represent the importance of the features. AcME calculation is extremely fast, making it particularly adaptable in applications where time efficiency is crucial, like high-frequency production industry or real-time monitoring. In addition, the resulting visualizations are designed to be concise and easy to read also for users with limited or no background in Machine Learning, as most decision support systems target users.





This initiative has been funded by 2014/2020 POR FESR Regione Veneto Project 'ExplAIn 4.0: Sviluppo di soluzioni interpretabili di Intelligenza Artificiale per l'Industria 4.0'
Bando per il sostegno di progetti di ricerca che prevedono l'impiego di ricercatori, Asse 1, Azione 1.1.1

STATWOLF

www.statwolf.com

datascience@statwolf.com

2014/2020 POR
FESR / REGIONE DEL VENETO
UNIONE EUROPEA
REGIONE DEL VENETO