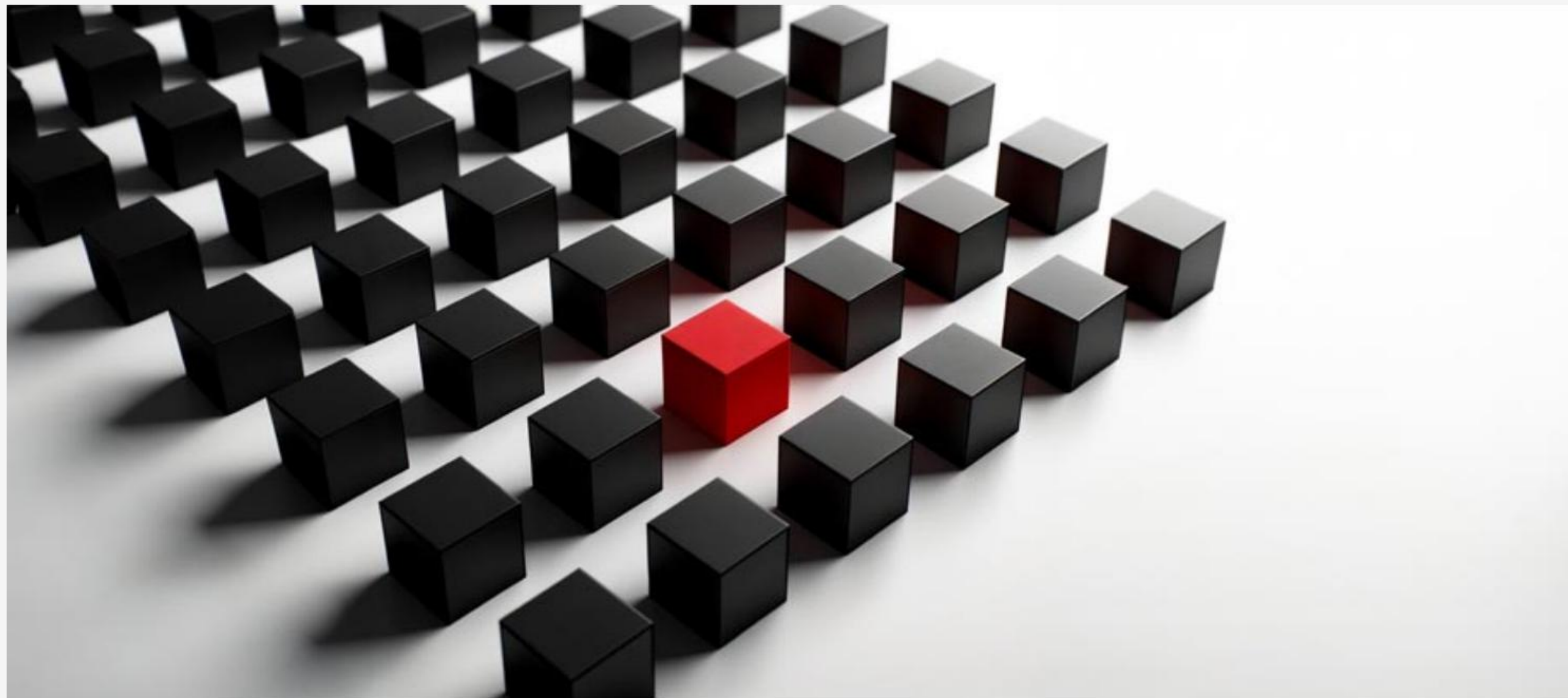


Industrial Anomaly Detection in manufacturing

a Supper & Supper Use Case



Project Goal

The goal of this project was to detect anomalous behavior in machinery as well as mechanical and industrial equipment **without prior information about what anomalous behavior consists of**. Preemptively detecting anomalies in manufacturing and production processes paves the path towards a more efficient future for industries and factories. In this project, we used state-of-the-art machine learning tools to ensure precise anomaly detection in industrial processes, allowing for **early identification** of such anomalous behavior.

What are anomalies in manufacturing?

Anomalies in manufacturing refer to deviations in the operation of a (manufacturing/industrial engineering/technical) system's operation from its intended or normal behavior. Such deviations can **decrease performance, leading to instabilities, security issues, defects, and even system failure**. Given the intricate dynamics of these systems, pinpointing the causes of these anomalies can be challenging.

Dataset used

The data consisted of **multivariate time series** collected from sensors installed on a machine test bed. Anomalous data was obtained from experiments in which the machine/system setup was **deliberately manipulated** over certain periods of time. The algorithms used work equally well when data is available only from the normal behavior of a control system – in which case anomalies in the datasets can be synthetically created to test the anomaly detection approach. The training set consists of **8,125 data points, of which 337 are anomalous**.

If we know which data points are anomalous, we can take advantage of this information using (semi-)supervised learning methods. In most applications, however, this information is unavailable, and we need to apply unsupervised learning. For this project, we used **unsupervised learning methods**, which means that the labels of the anomalous data points in our dataset are only used to test the performance of our methods, not to train our models.

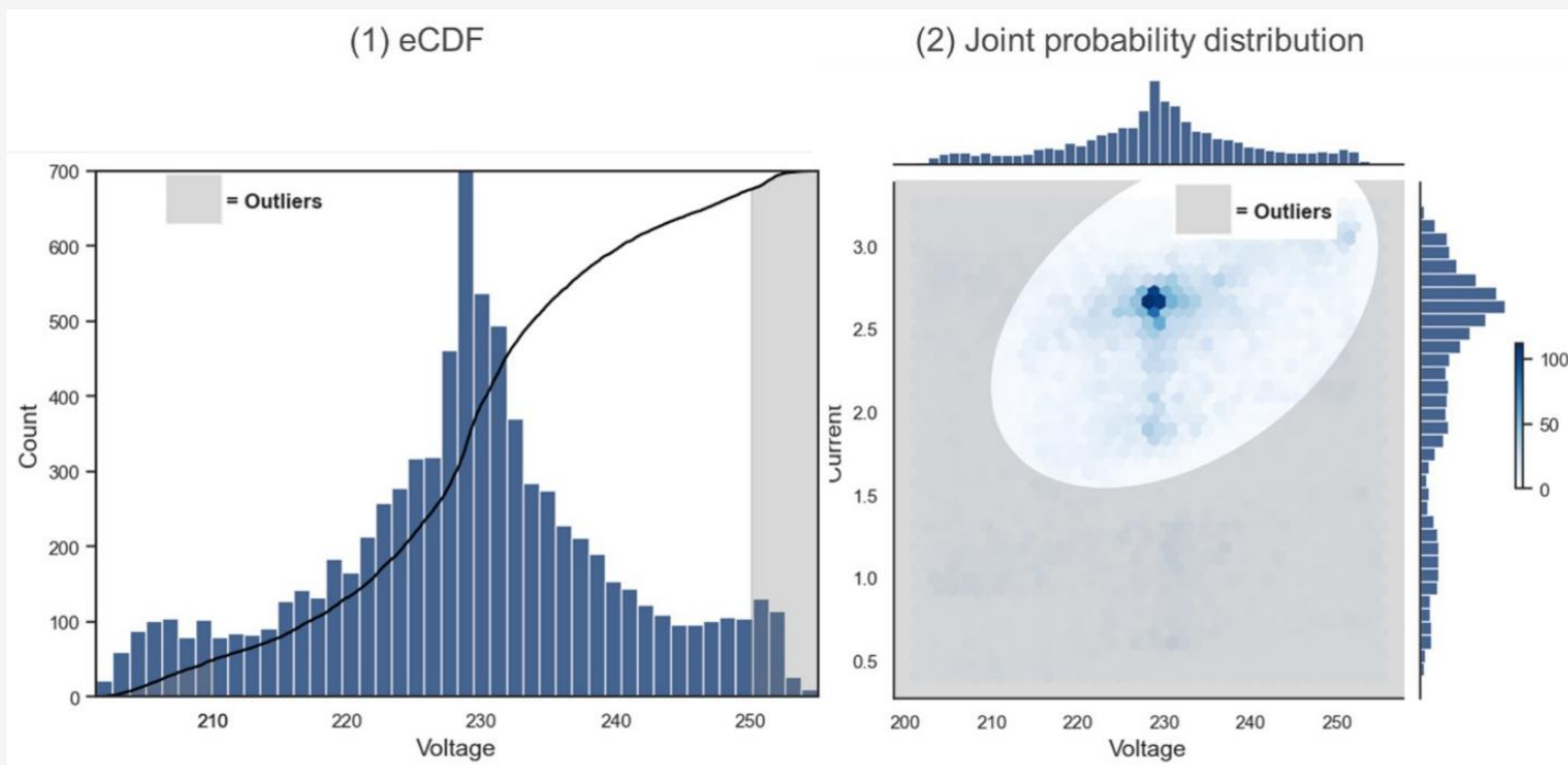
Challenges

1. One major challenge is that there are only few <<5% anomalous data points in the training set, presenting a scarcity of information sources to learn from.
2. Furthermore, there is a tradeoff between the number of **false positives (falsely detected anomalies)** and **false negatives (missed anomalies)** when choosing the best algorithm. In this project, as in many industrial applications, the cost of a missed anomaly is higher than that of a false alarm. Therefore, the optimal method needs to minimize false negatives while maintaining a good overall performance.
3. Lastly, many established machine learning methods are computationally expensive and time-consuming. The optimal method should be **efficient and lightweight** enough to run on embedded/edge devices.

Applied Methods (Implementation)

We approached the problem by testing established machine learning methods for anomaly detection (OneClassSVM, iForest) against state-of-the-art models that address the above challenges (ECOD, COPOD).

- **One class SVM**: detects anomalies by learning a decision boundary that groups data through classification into anomalous and non-anomalous
- **Isolation Forest (iForest)**: detects anomalies using binary trees
- **ECOD**: detects anomalies using empirical cumulative distribution functions (eCDFs)
- **COPOD**: detects anomalies using empirical copulas to obtain joint probability distributions



Statistical concepts behind (1) ECOD and (2) COPOD

To evaluate the algorithms, we used:

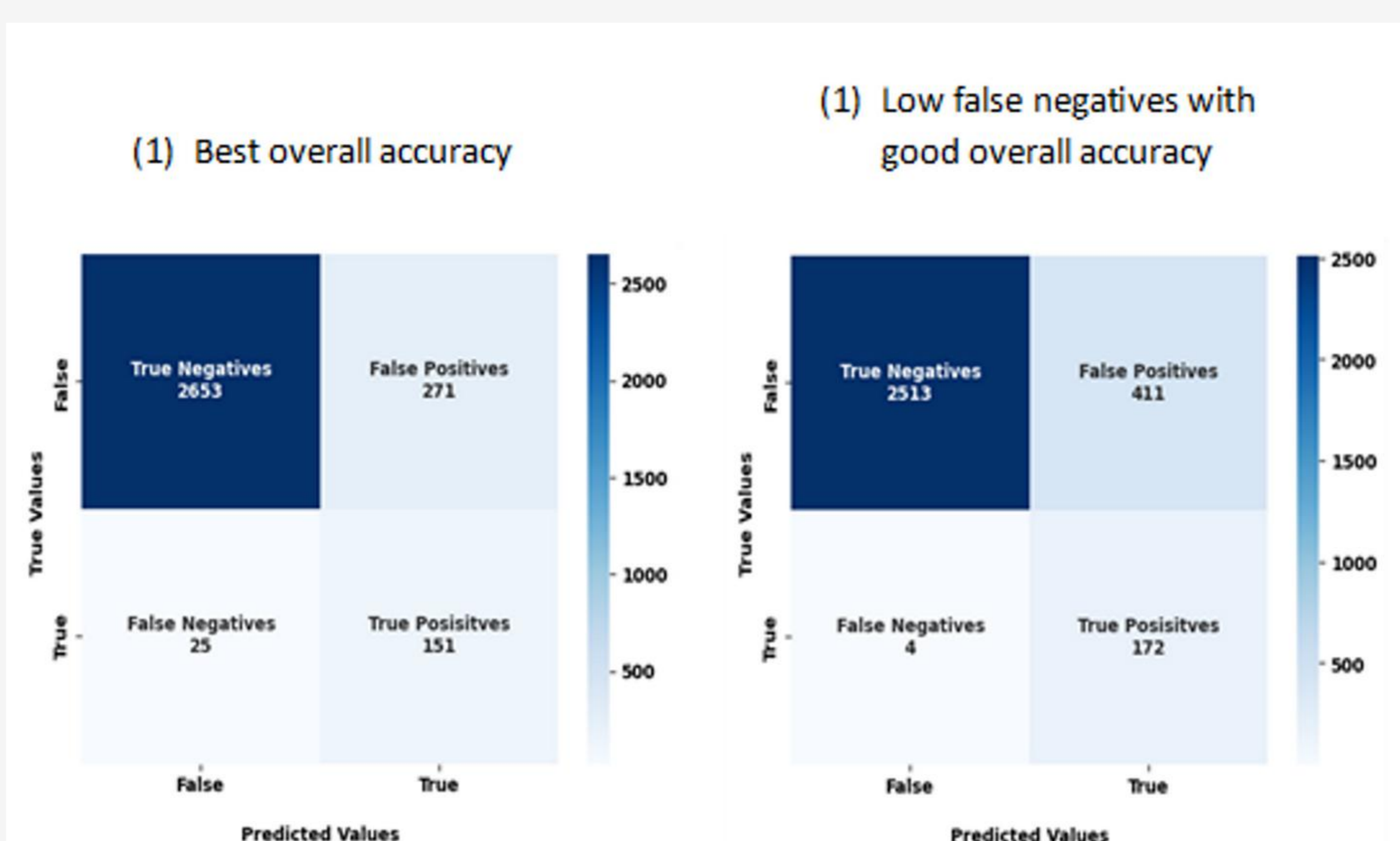
- the Missing Alarm Rate (MAR) = missed anomalies / all anomalies
- the False Alarm Rate (FAR) = falsely detected anomalies / all non-anomalies, and
- the F1 macro score: a measure of the overall accuracy of the model.

	FAR (%)	MAR (%)	F1 macro score
COPOD (F1 maximized)	9.3	14.2	0.73
COPOD (MAR minimized)	14.1	2.3	0.69
ECOD	12.7	8.0	0.69
iForest	12.1	16.5	0.68
OneClassSVM	13.7	38.6	0.62

Project outcome

The best performing models for our project of industrial anomaly detection in terms of overall accuracy, MAR and FAR were ECOD and COPOD. They are extremely lightweight and efficient and could therefore run on embedded devices.

- The COPOD model was optimized for (1) overall accuracy or (2) low number of false negatives (with good overall accuracy).
- The model optimized for a low number of false negatives missed only 2.3% of anomalies (MAR) and misclassified 14.1% of all non-anomalies (FAR).
- The model optimized for overall accuracy missed 14.2% of anomalies (MAR) and misclassified 9.3% of all non-anomalies (FAR).



Category

- Mechanical Engineering
- Predictive Maintenance

Download



Social Sharing

