Anomaly detection

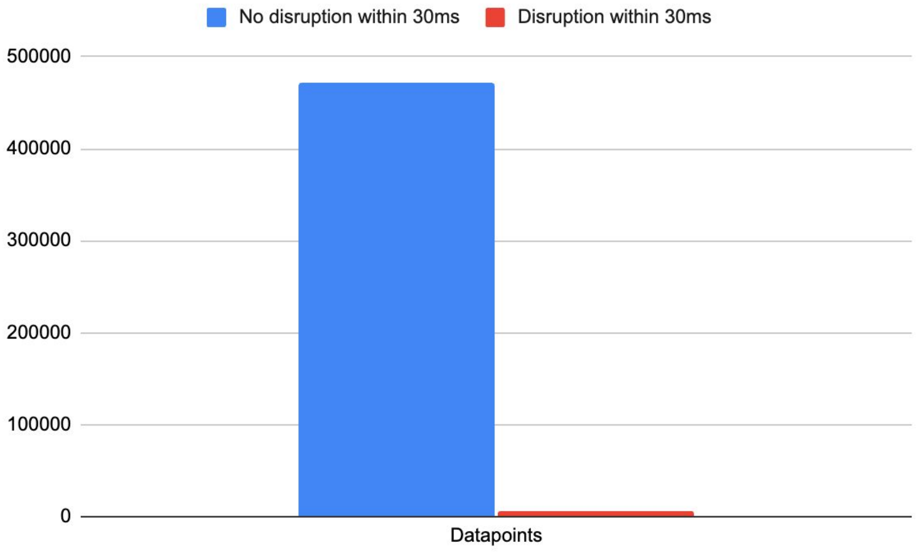
Problem

Detection of disruptions in a fusion reactor within next 30ms.

The identification of disruptions allow to implement and act control measures to mitigate damages to infrastructure.

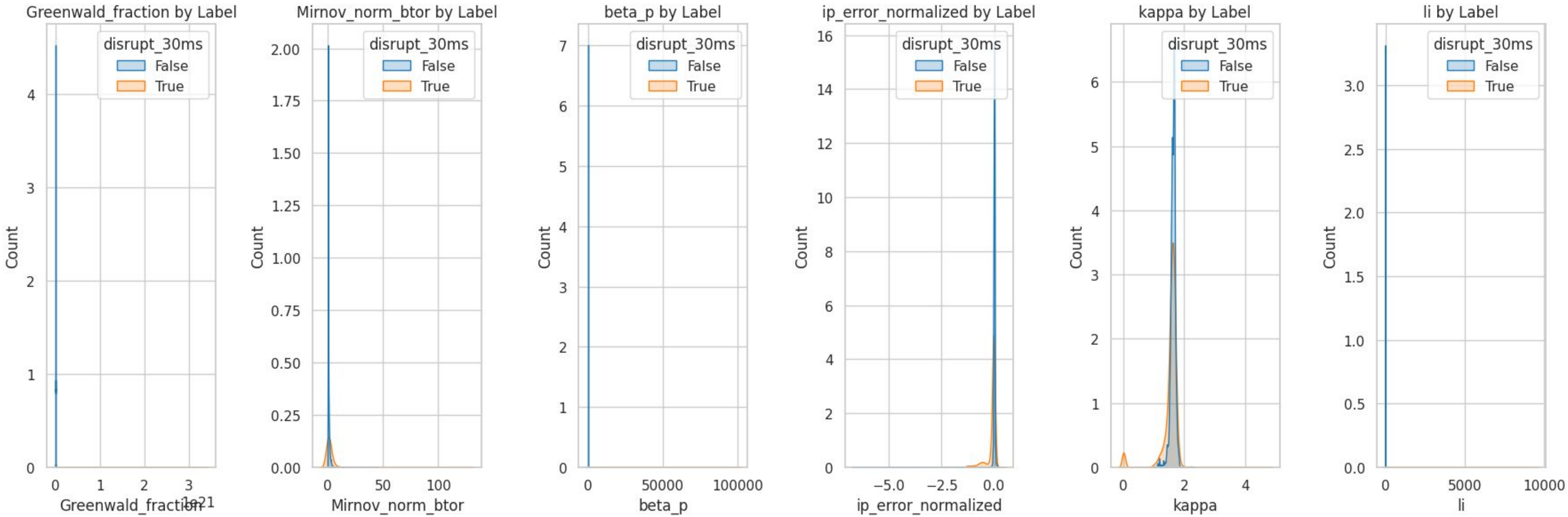
Available Data

The data used is highly imbalanced, with many more states of the reactor not leading to disruption. This is expected, as this task is an anomaly detection task. Still, this underrepresentation of data of interest could be a challenge for the Machine Learning algorithm, as it needs many observations of the event to be classified to adequately learn how to identify it.



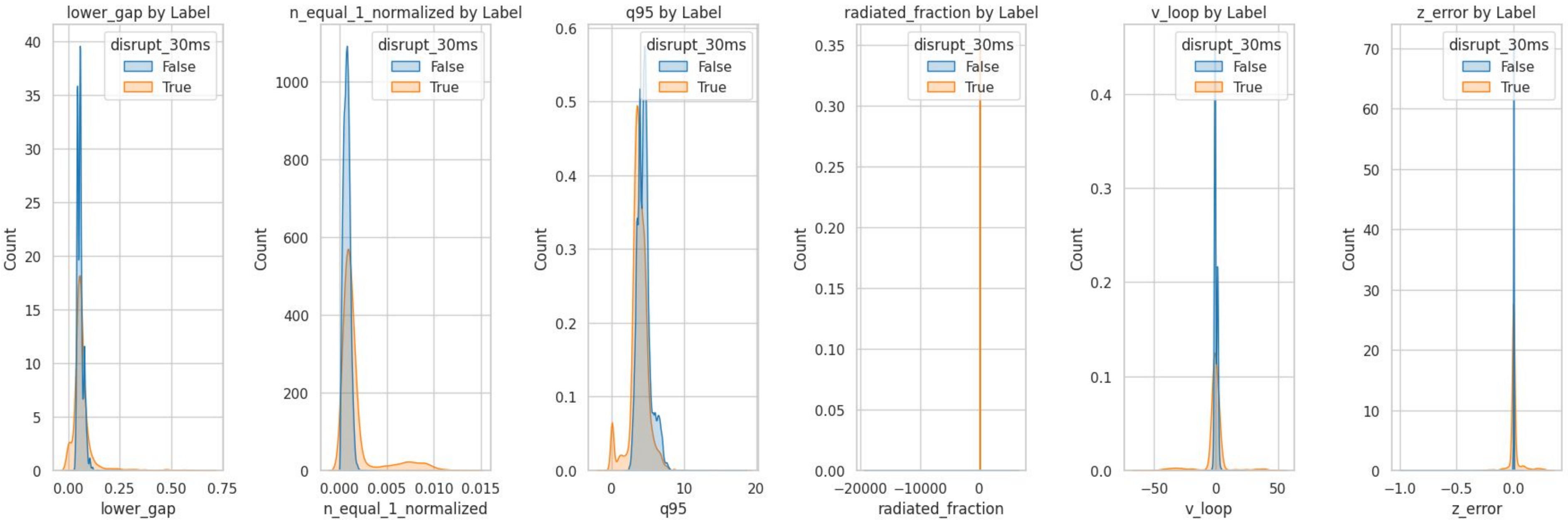
Available Data

No significant difference in the features of the reactor state to identify disruption

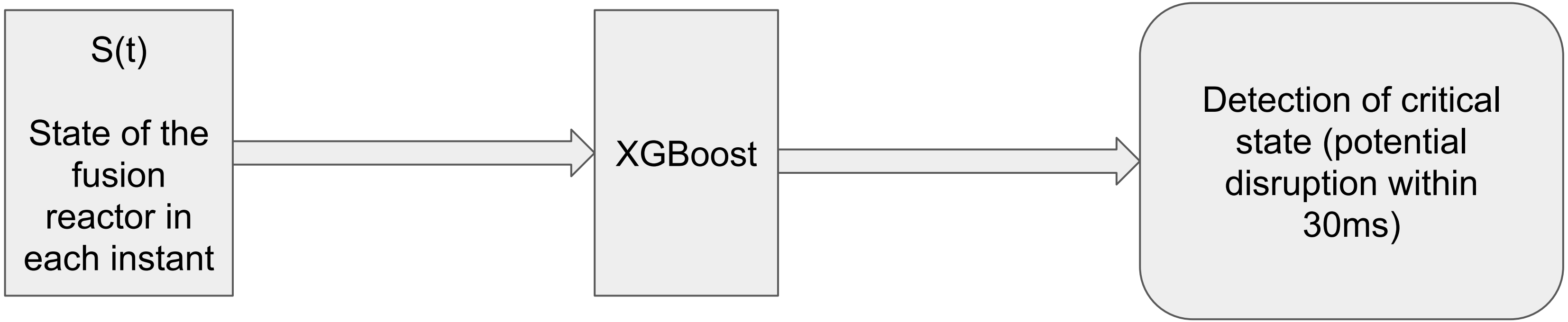


Available Data

No significant difference in the features of the reactor state to identify disruption



Approach



The algorithm classifies each state of the fusion reactor using the data observed in each instant. After the algorithm receives the data related to the state of the reactor, it provides a classification of the state in one of two classes:

- Disruption within 30ms (1)
- No disruption within 30ms (0)

Performance

The algorithm obtains performance on novel data as follows:

- Recall: 58%

The Recall is defined as the ratio between identified disruptions and the total number of disruptions; therefore, the algorithm correctly identifies 58% of the disruptions.

So, the algorithm allows mitigation and acts to prevent or mitigate the impact of 58% of the disruptions.

Performance

The algorithm obtains performance on novel data as follows:

- Precision: 64%

Precision is defined as the fraction of actual disruptions among the identified ones. Thus, in 64% of the cases when the alert will be triggered, an actual disruption would have happened, producing false alarms only in 36% of cases.

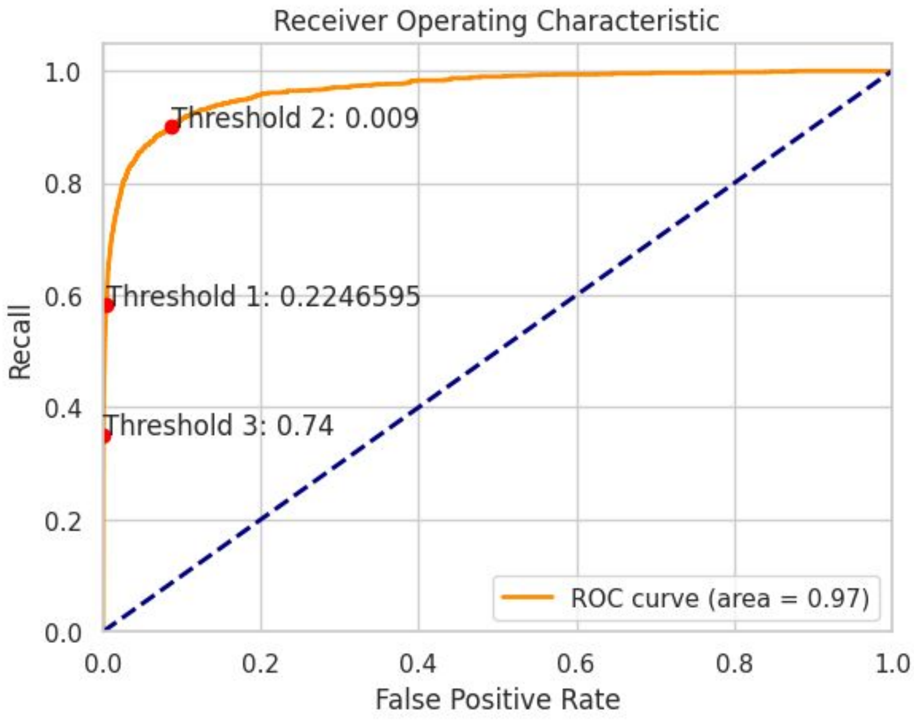
Performance

Furthermore, the device is easily customizable to the business needs, to be able to be more suitable to the team balance between false alarms and sensibility of the device.

The algorithm for each instance provides a probability score. This score is then analysed using a decision threshold to classify the event.

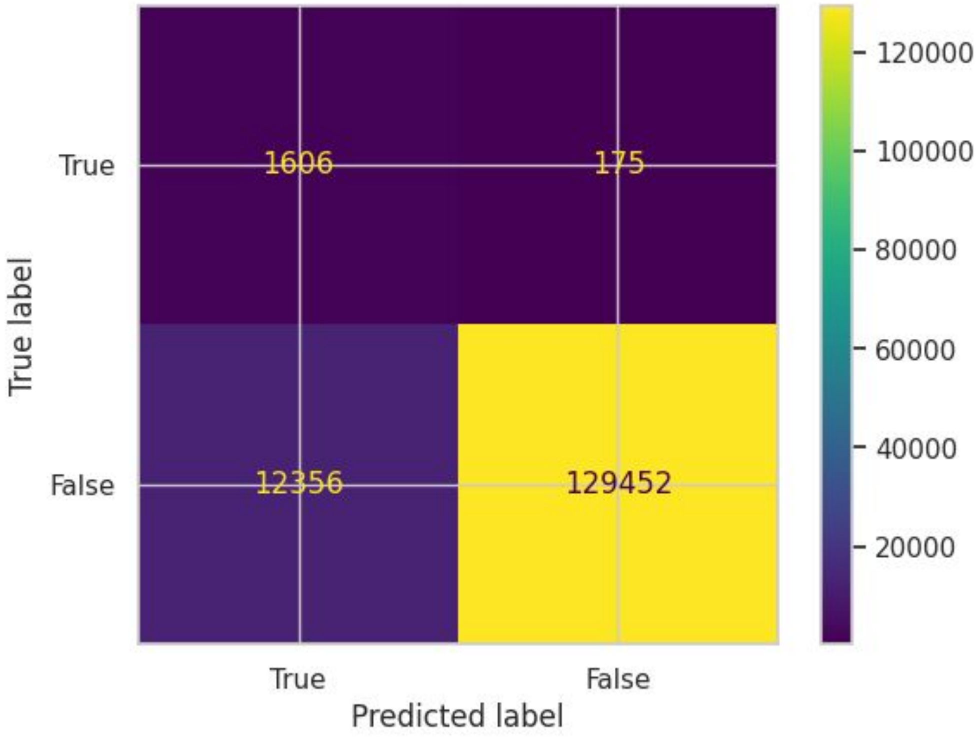
This decision threshold is configurable to adapt to different needs quickly. Threshold 1 obtains the performance seen in the previous slides.

In this plot, the False Positive Rate represent the ratio between false alarms and the total number of observations without disruptions.



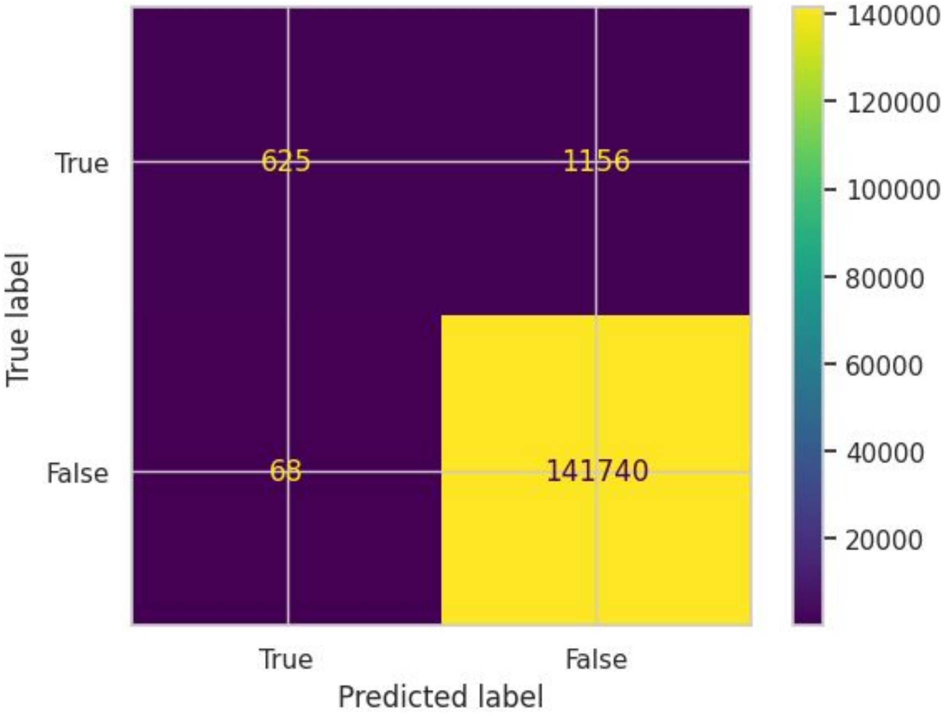
Performance

The customisation of the threshold enables the change of the algorithm behaviour, allowing it to be overly sensitive. In this case, using Threshold 2, it obtained a Recall score of 90%, identifying 9 out of 10 disruptions but with a higher grade of false alarms (up to 88%).



Performance

Customising the threshold also allows it to be overly conservative like this one obtained using Threshold. 3. It obtained a Precision score of 90%, providing only one false alarm out of 10 alarms, but with a higher grade of missed disruption identification (up to 65%).



Robustness

The results shown before are all obtained from data that the algorithm never used to train, thus representing an actual representation of the data visualisation of the model as it was in production, receiving new unseen data. Therefore, considering data without significant changes with respect to the one used to build the model, the model is expected to perform as represented in the previous slides.

Benefits

The implementation of this model allows for an easy and quick deploy, and allows to have detection of disruption potentially saving damages to the infrastructures and operativity interruptions. With the configuration initially proposed, this reduction may be applied in the 58% of the disruptions. Still, as it can be seen, it can be customised to prevent even more disruptions, reducing the events, therefore reducing direct costs, infrastructure management, infrastructure maintenance and operativity interruptions, increasing the efficiency of the reactor.