Quantifying the Confidence of Anomaly Detectors in Their Example-Wise Predictions

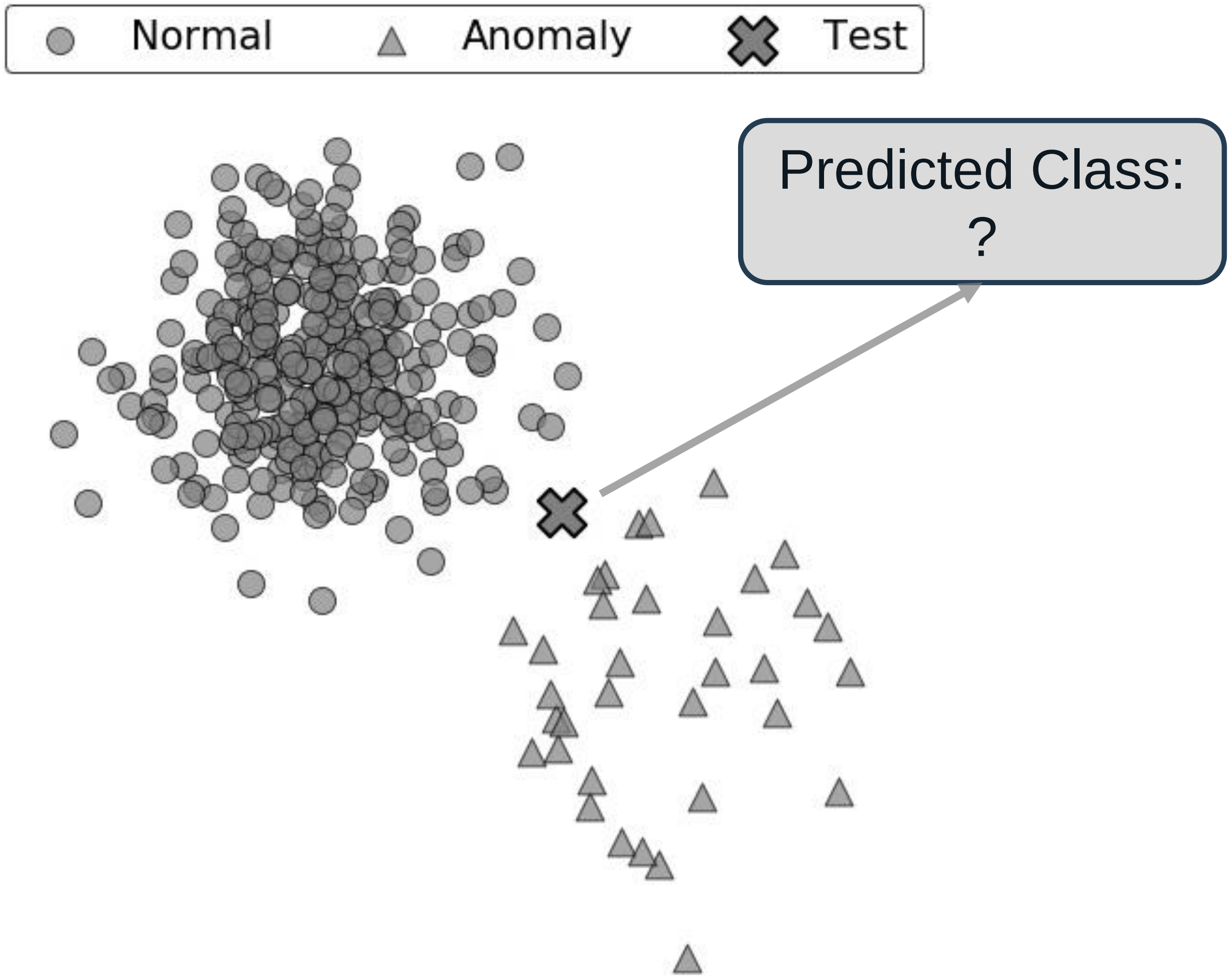
Lorenzo Perini, *Vincent Vercruyssen, Jesse Davis*

ECML - PKDD 2020

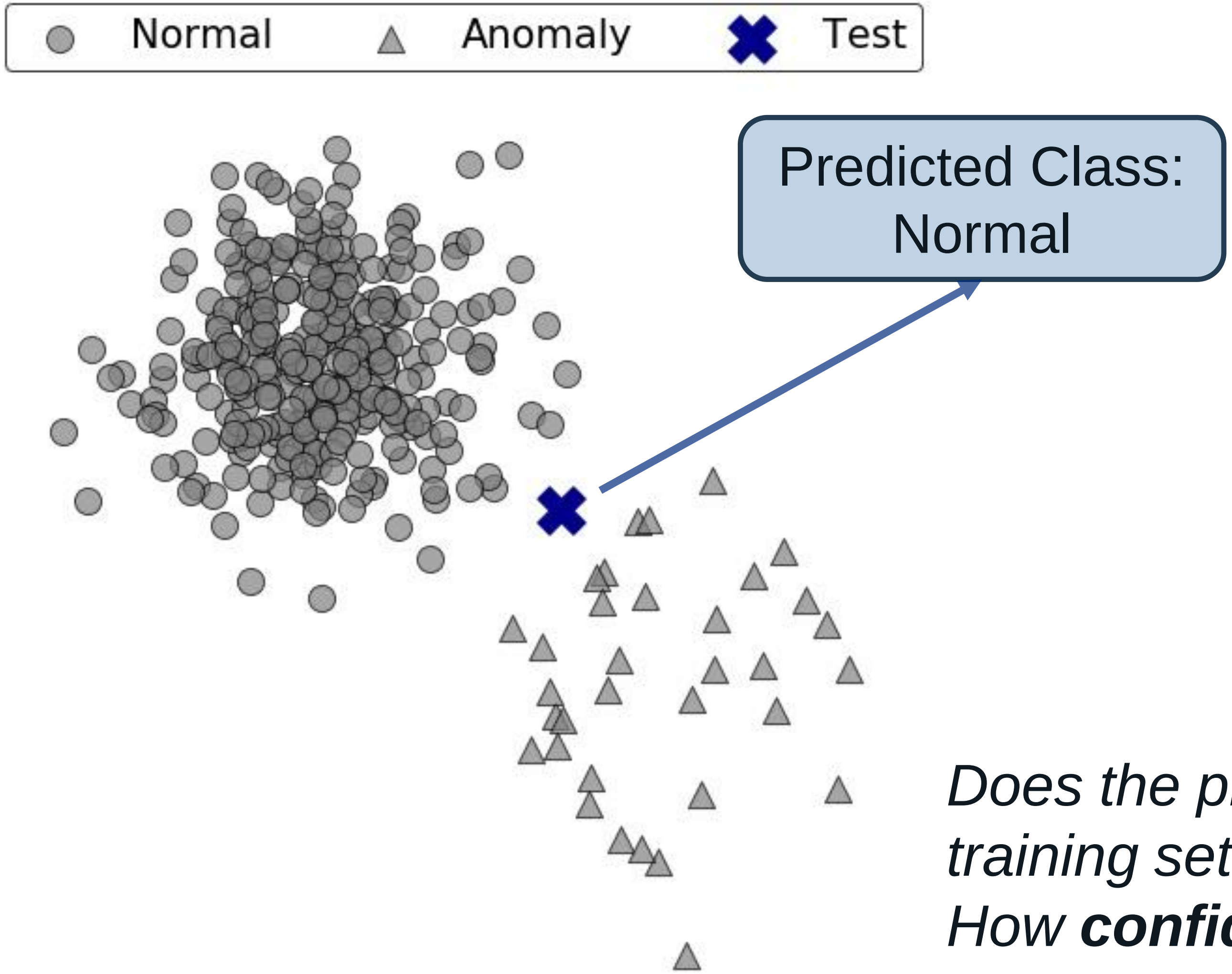
https://people.cs.kuleuven.be/~lorenzo.perini/

@LorenzoPerini95

Capturing Uncertainty in Anomaly Detection



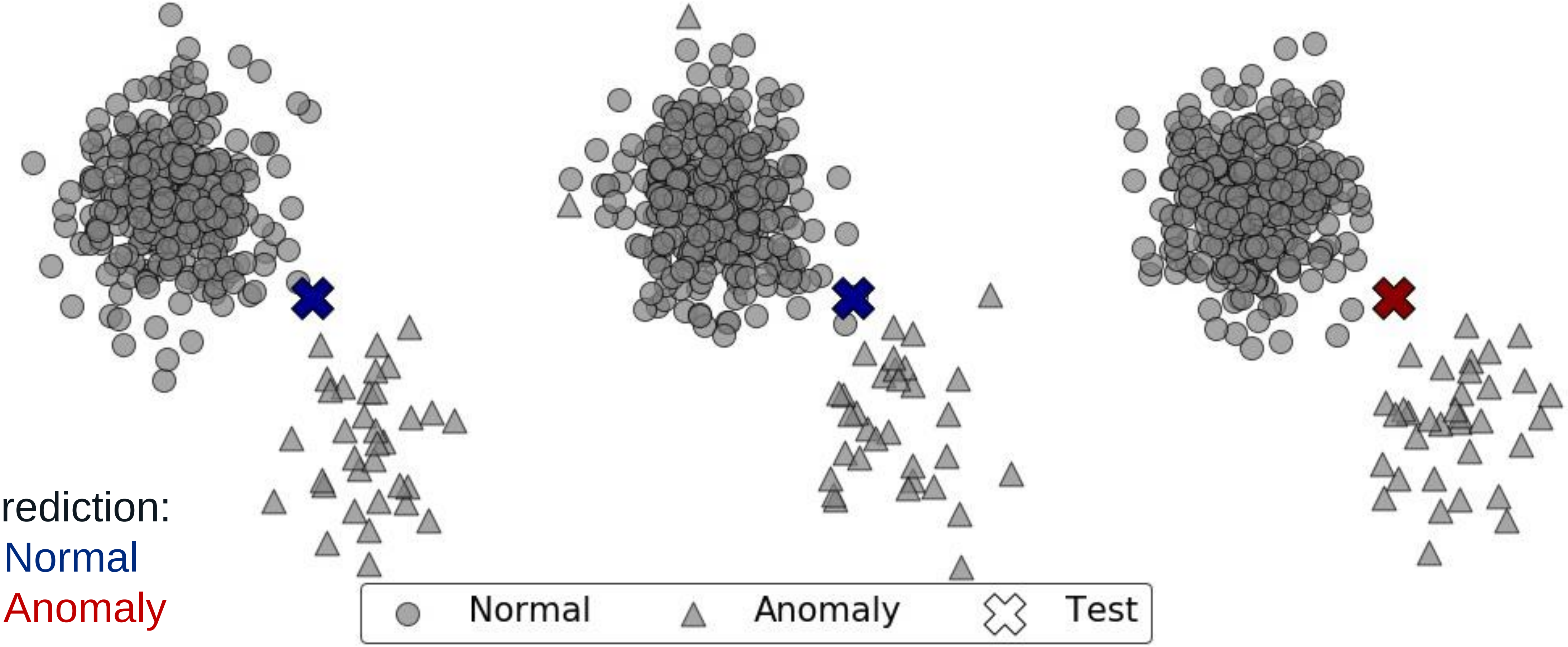
Capturing Uncertainty in Anomaly Detection



Does the prediction change when a different training set is drawn from the same distribution?
*How **confident** is the model in its prediction?*

What Is the Effect of Slightly Perturbing the Training Data on Anomaly Detector's Prediction for a Fixed Example?

Dataset 1 | Dataset 2 | Dataset 3

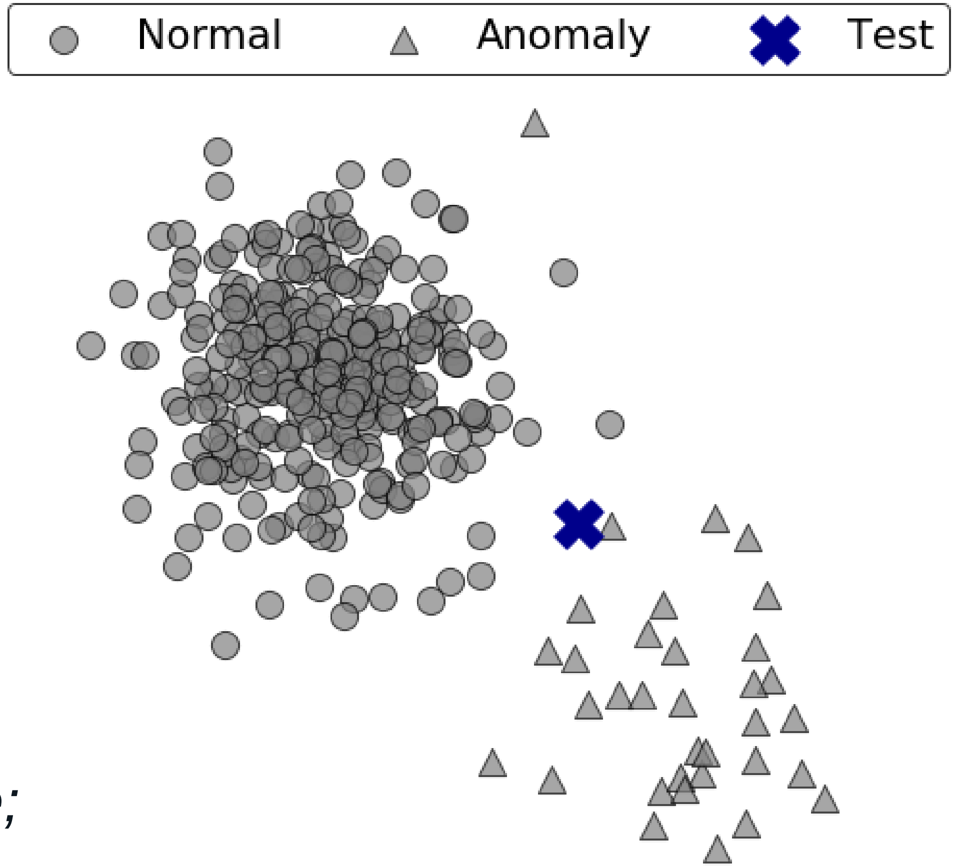


Prediction:
- Normal
- Anomaly

Our Goal Is To Capture the Uncertainty in Predictions

Out of 20 predictions:

- 50% Normal;
- 50% Anomaly.



Given:

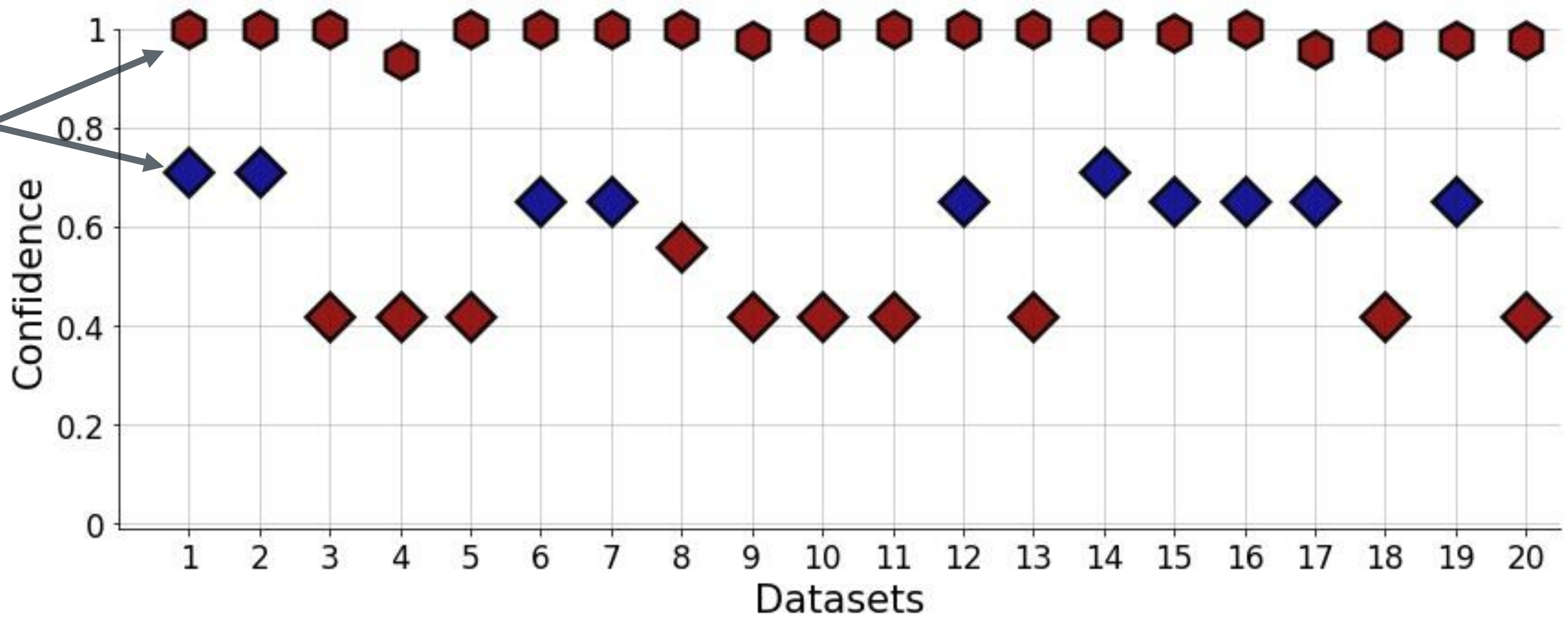
An anomaly detector and a test example;

Do:

Need of a measure of confidence in predicting the same class.

Confidence Allows Cross-Comparisons among Models

Which one would you trust more?



Prediction:
- Normal
- Anomaly

kNNO
IForest

How would you estimate
the *confidence* of a model
in its example-wise predictions?

End of the spotlight presentation

We Make 4 Contributions

1. Define the confidence in an example-wise prediction;

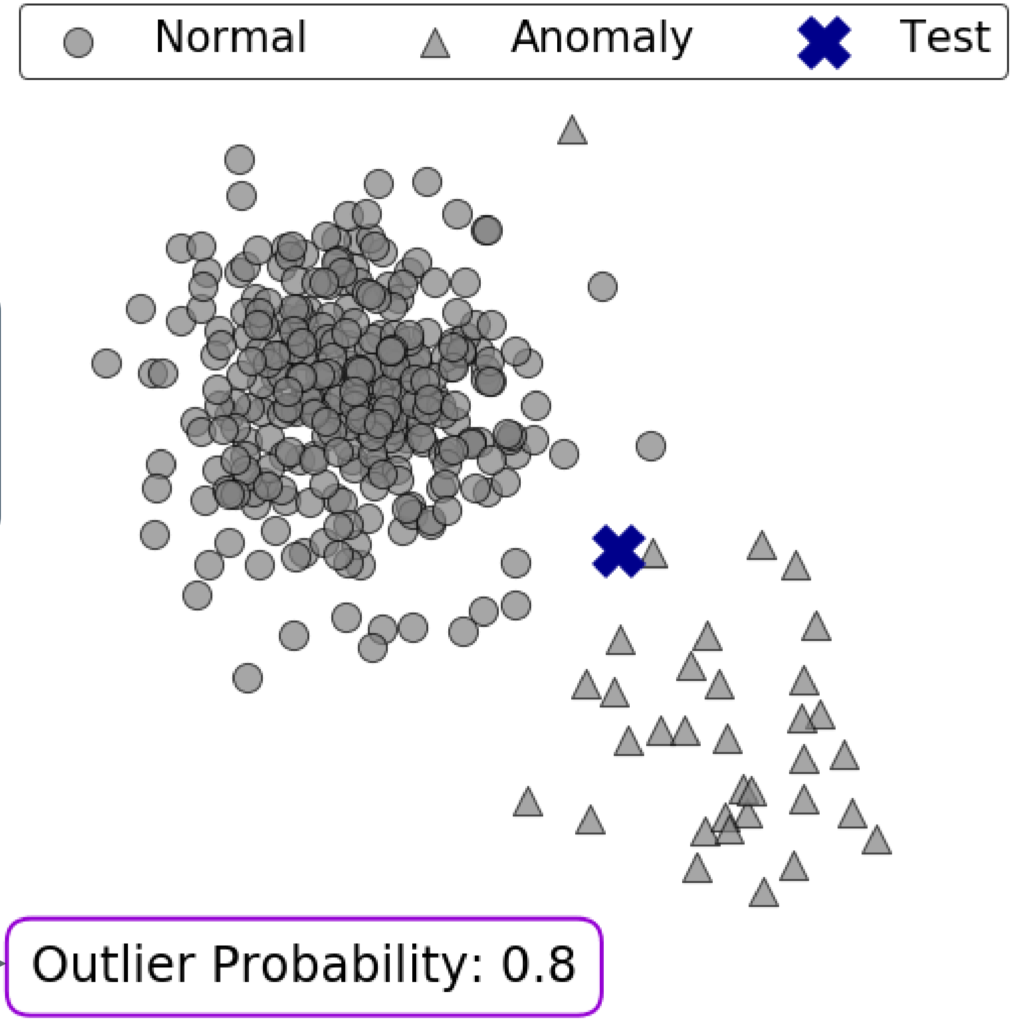
   ***Confidence*: the probability that a detector's prediction would change for any fixed example if a different training set was observed**

2. Propose ExCeeD: a 2-step approach for estimating the confidence:
   I. Convert anomaly scores to outlier probabilities using a Bayesian approach;
   II. Derive confidence scores by observing that different datasets would lead to different thresholds and, in turn, to different predictions;
3. Analyze the convergence behavior of our estimate;
4. Perform an extensive empirical evaluation on benchmark datasets.

Contribution 1: Defining the confidence

A measure of uncertainty in class prediction

Do Density Estimators Capture Uncertainty in Class Prediction?

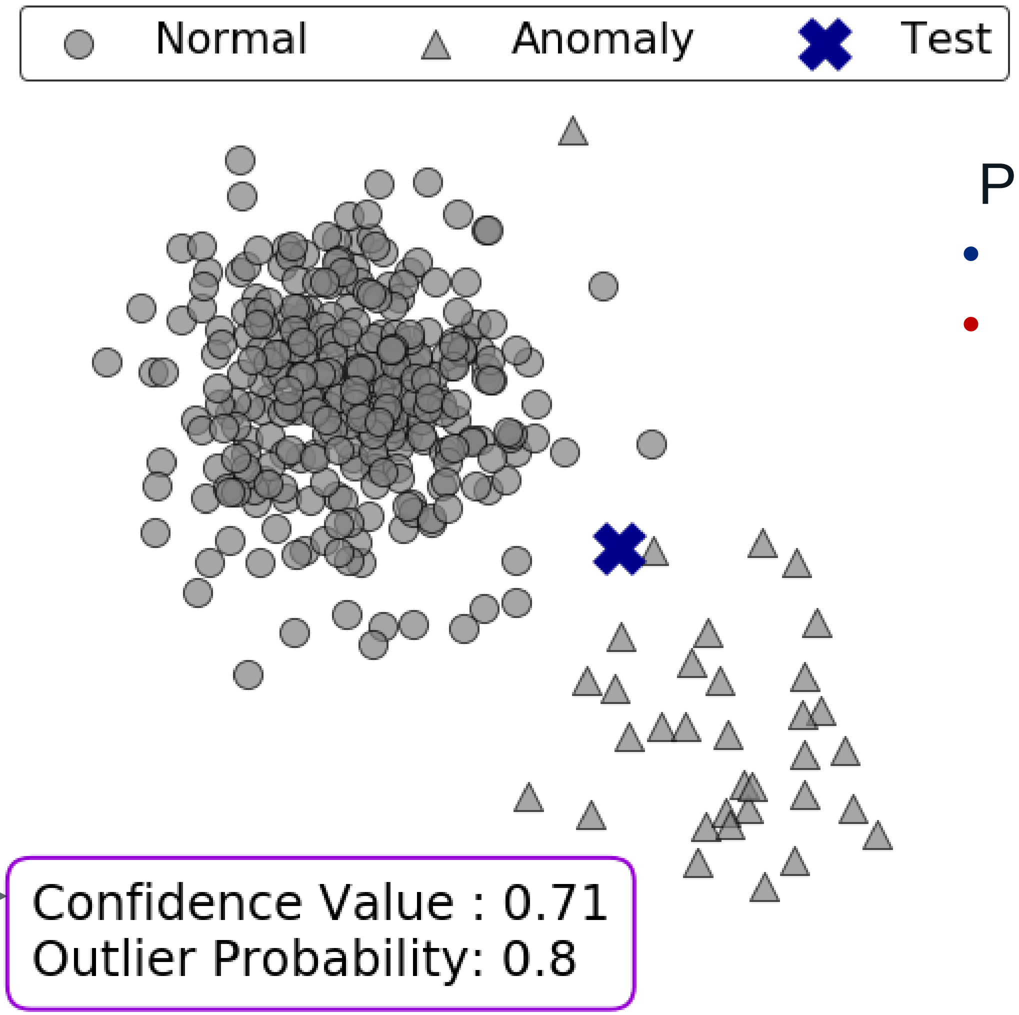


Although the predicted class changes many times, the outlier probability keeps being high

Prediction:

- Normal
- Anomaly

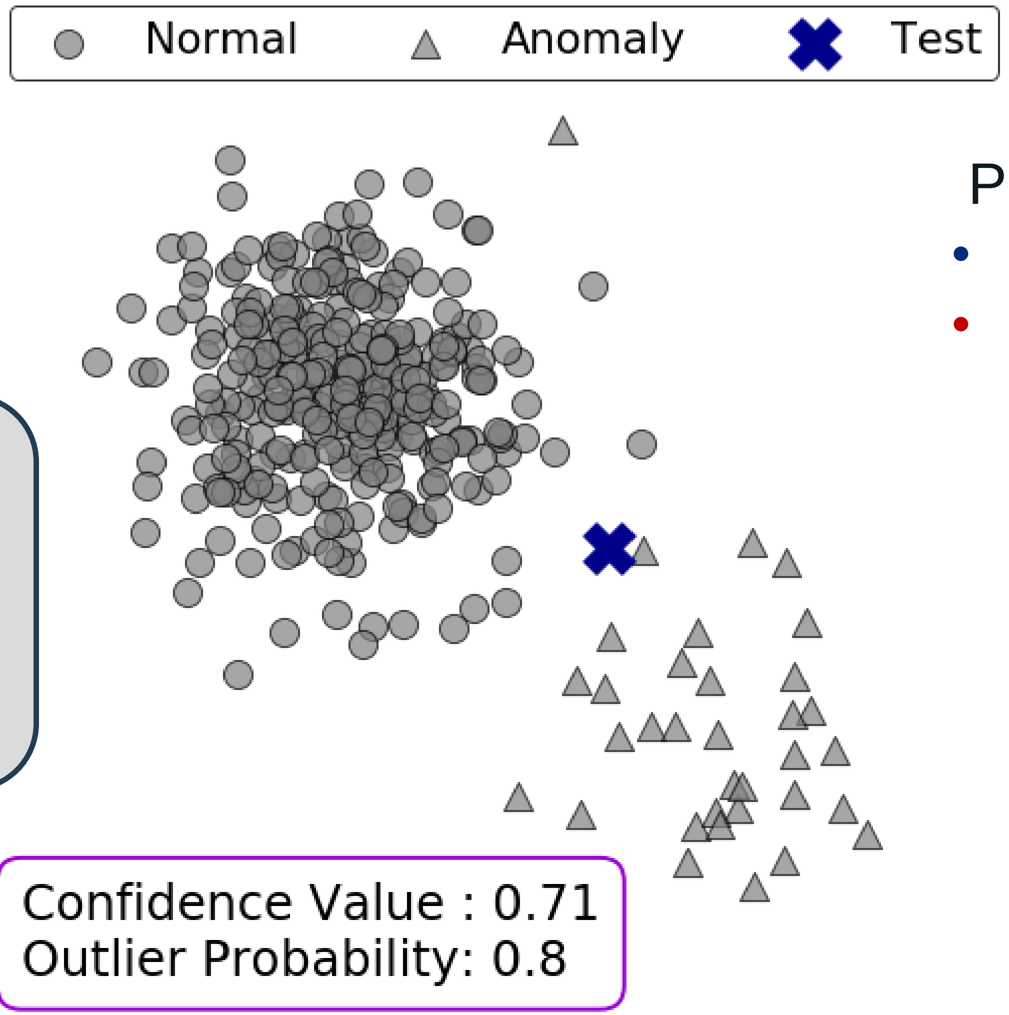
Confidence Is a Novel Measure of Consistency



- When the class is Normal, the confidence is around **60%**
- When the class is Anomaly, the confidence is around **40%**

Prediction:
- Normal
- Anomaly

Confidence Is NOT Outlier Probability because they measure fundamentally different things

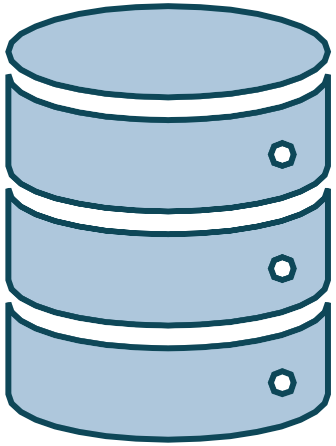


Prediction:
- Normal
- Anomaly

Outlier probability measures the probability of belonging to the anomaly class

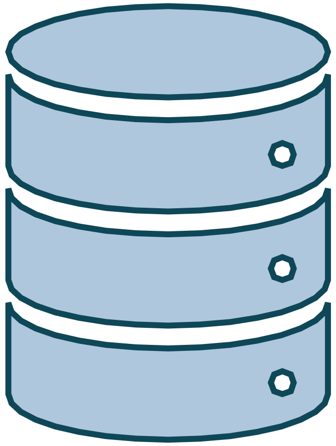
Confidence measures the consistency in predicting the same class

3 Steps of Standard Unsupervised Anomaly Detection



Unlabeled Data

3 Steps of Standard Unsupervised Anomaly Detection

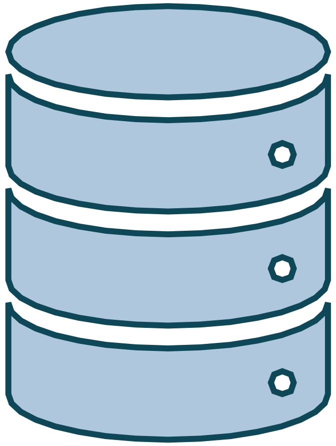


Unlabeled Data

Map

Anomaly Scores
s_4
s_t
s_1
s_6
…
s_n
…
s_5

3 Steps of Standard Unsupervised Anomaly Detection



Unlabeled Data

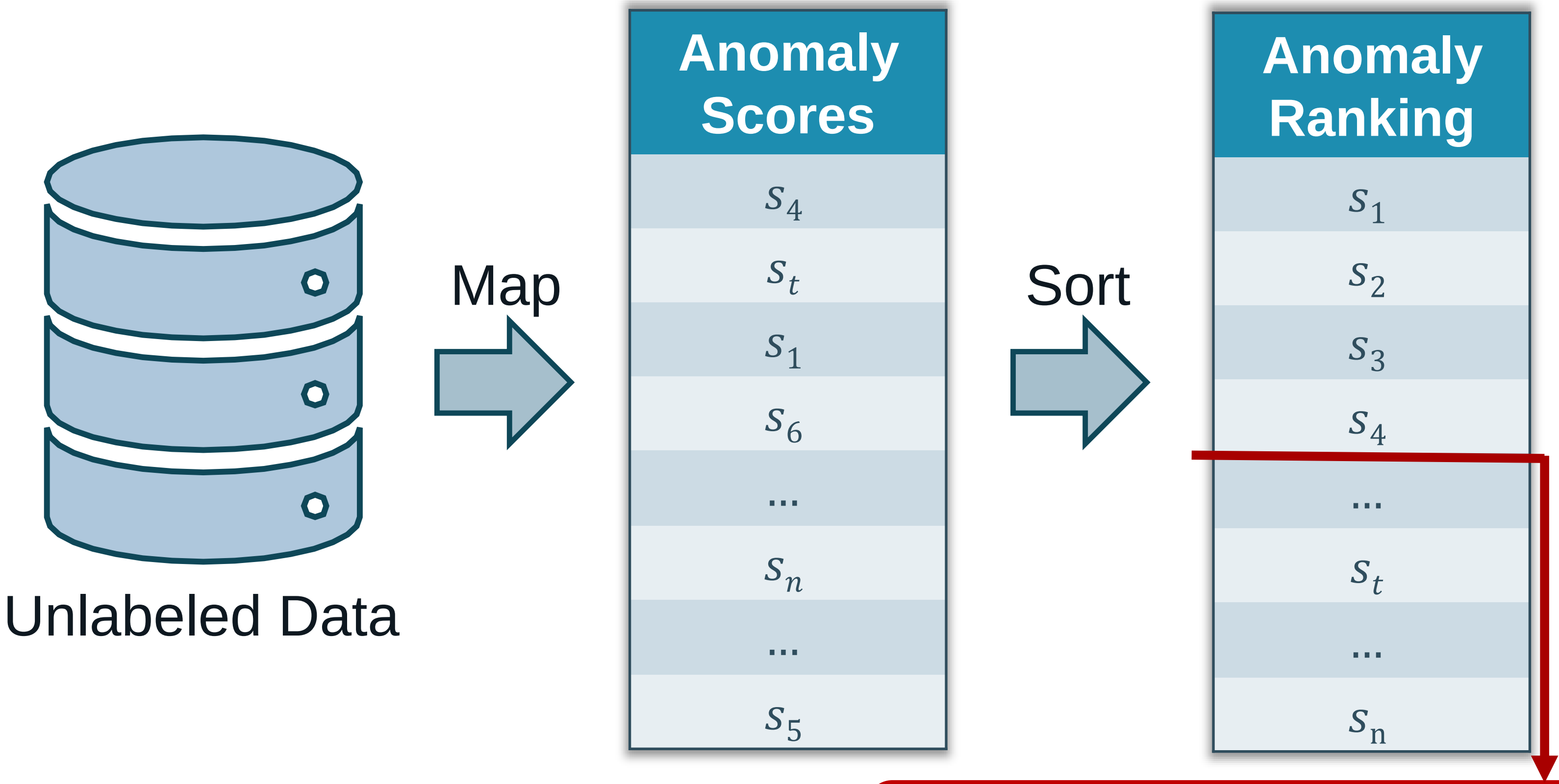
Map

Anomaly Scores
s_4
s_t
s_1
s_6
…
s_n
…
s_5

Sort

Anomaly Ranking
s_1
s_2
s_3
s_4
…
s_t
…
s_n

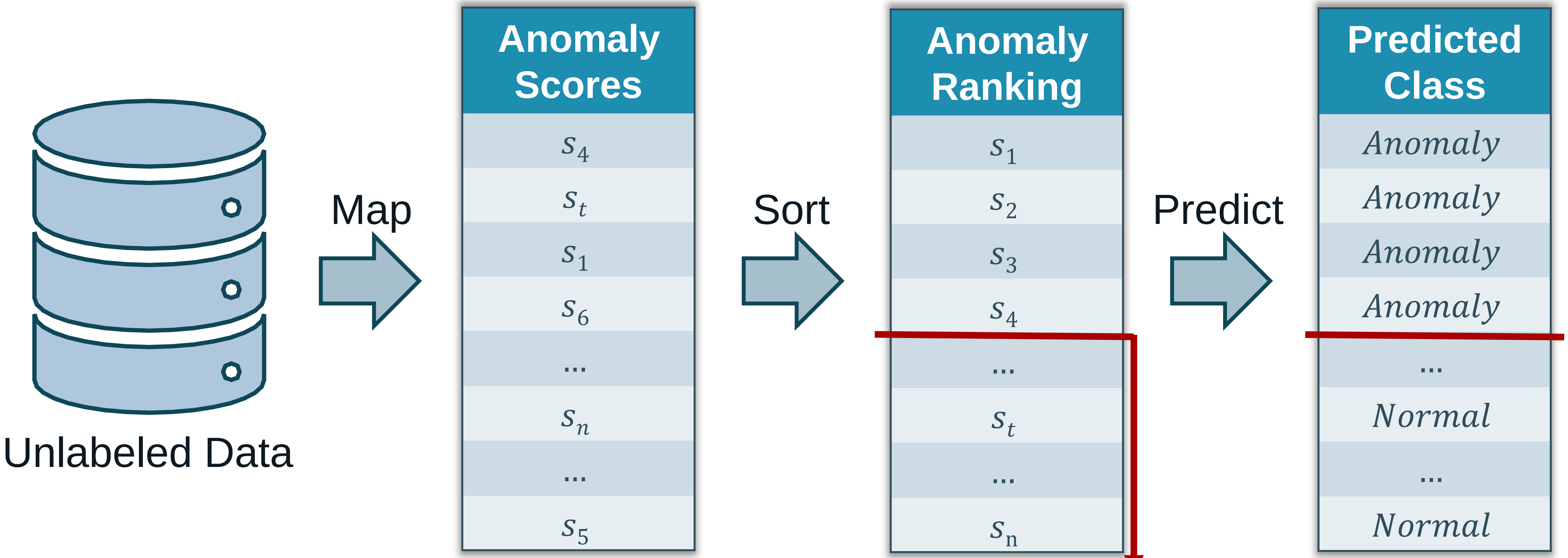
3 Steps of Standard Unsupervised Anomaly Detection



Threshold $= \boldsymbol{n \times \gamma}$

($\boldsymbol{\gamma}$ is the expected proportion of anomalies)

3 Steps of Standard Unsupervised Anomaly Detection



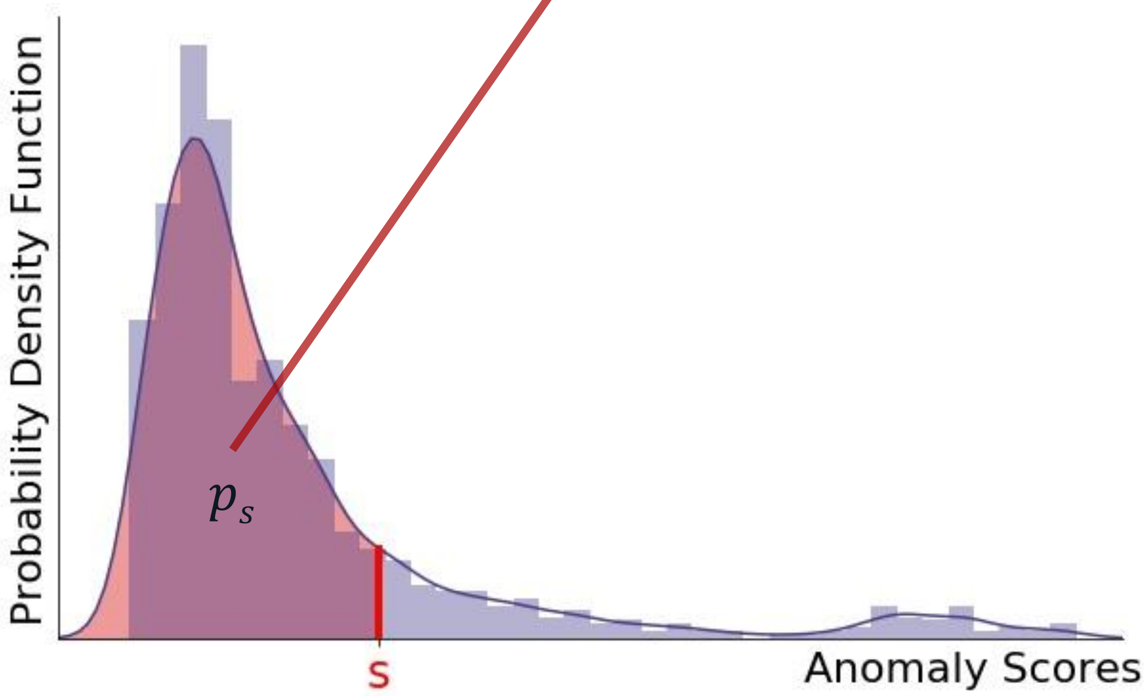
Threshold $= \boldsymbol{n \times \gamma}$
($\boldsymbol{\gamma}$ is the expected proportion of anomalies)

Contribution 2: ExCeeD

A two-steps approach for estimating the confidence

Step 1: Converting Anomaly Scores to Outlier Probabilities

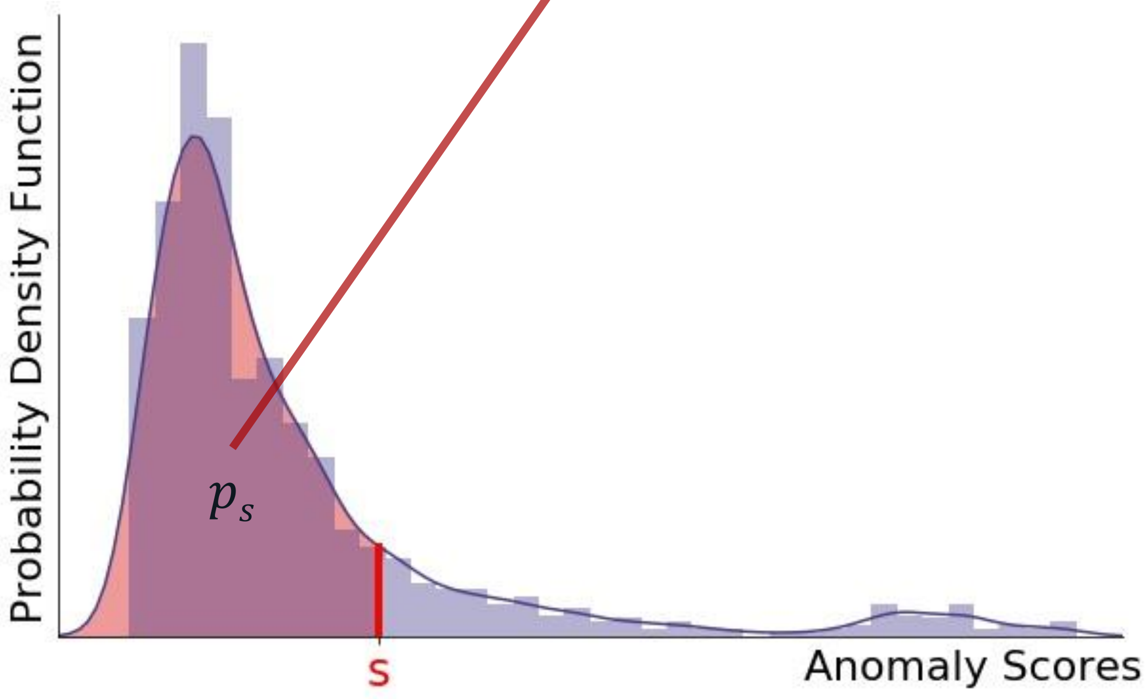
The larger the number of examples
with a lower anomaly score,
the more anomalous the example is.

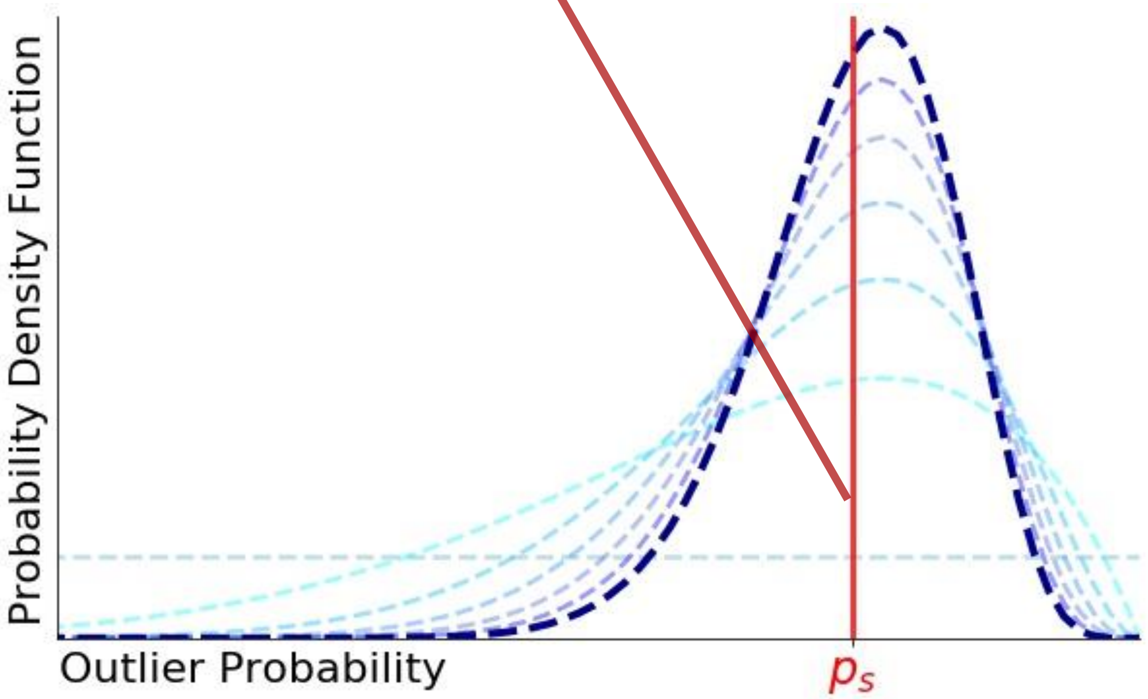


Step 1: Converting Anomaly Scores to Outlier Probabilities

The larger the number of examples
with a lower anomaly score,
the more anomalous the example is.

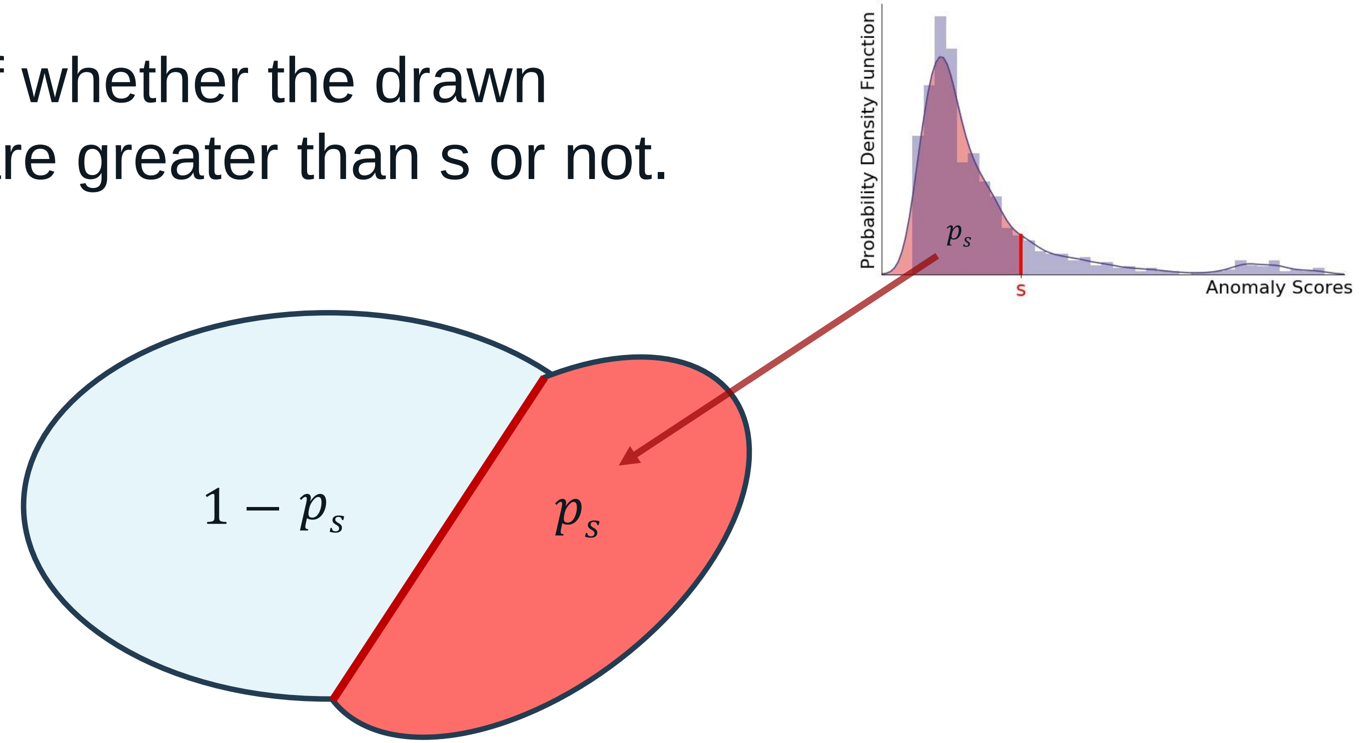
We need a Bayesian approach
because we need a smooth
measurement of the area.





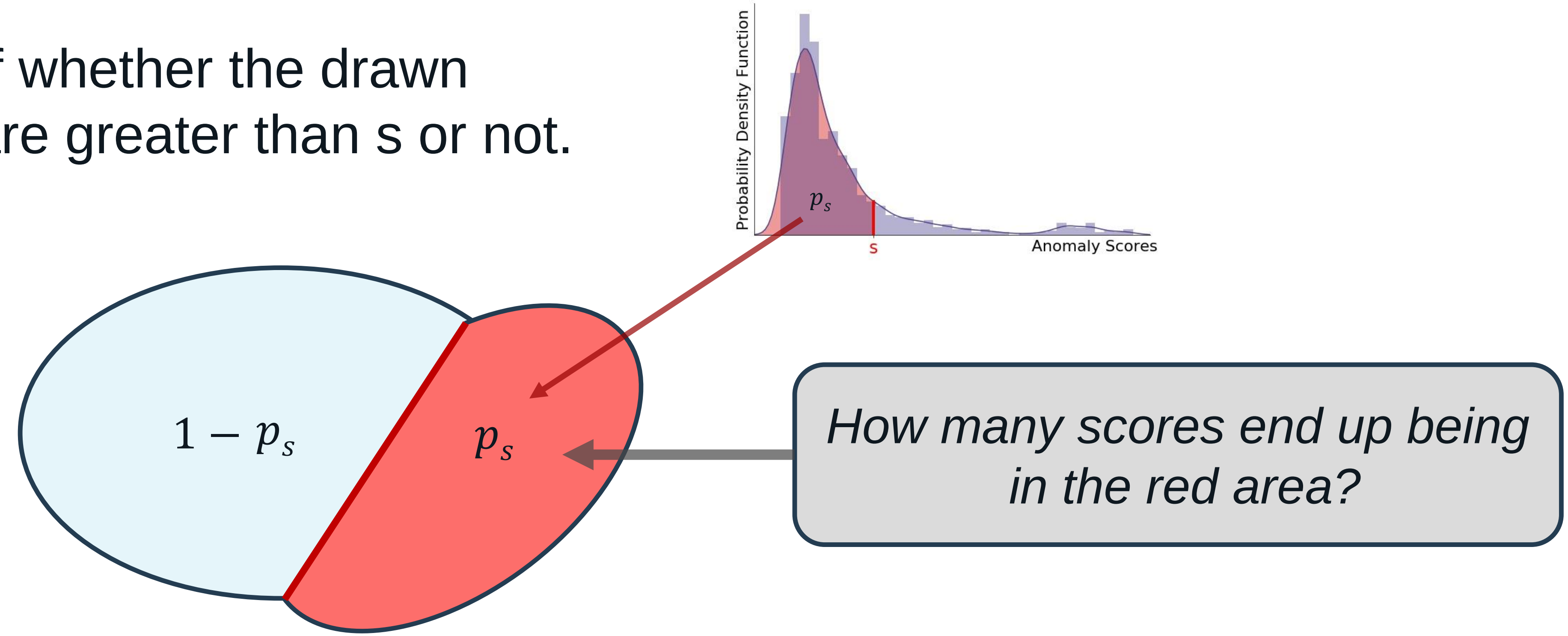
Step 2: Converting Outlier Probabilities to Confidence Scores

It is only matter of whether the drawn anomaly scores are greater than s or not.



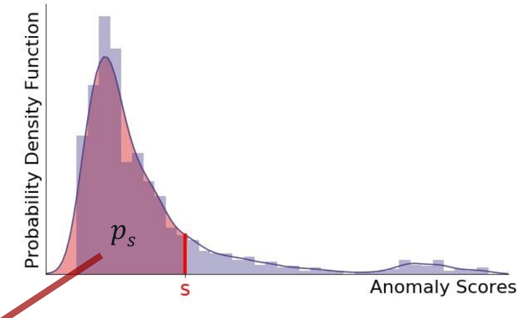
Step 2: Converting Outlier Probabilities to Confidence Scores

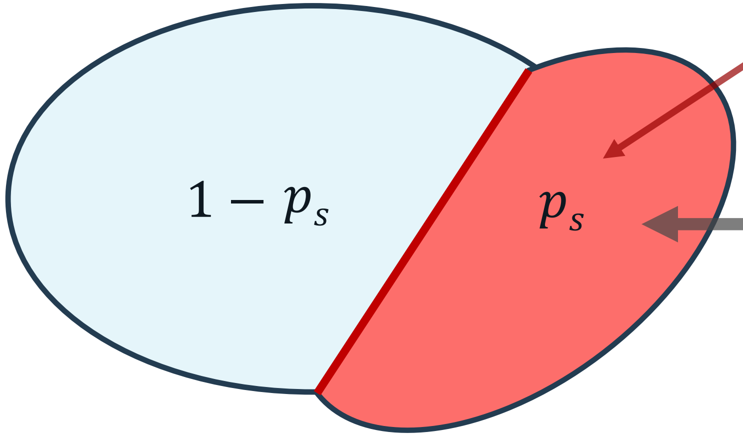
It is only matter of whether the drawn anomaly scores are greater than s or not.



Step 2: Converting Outlier Probabilities to Confidence Scores

It is only matter of whether the drawn anomaly scores are greater than s or not.





How many scores end up being in the red area?

ExCeeD estimates the probability that *‘enough’* examples fall inside the red area in order to keep the same prediction

Contribution 3: Convergence Analysis

How increasing the size of the training set affects ExCeeD’s confidence estimation

Convergence Behavior of Our Confidence Measure

Two cases according to $\boldsymbol{\gamma}$:

$\boldsymbol{\gamma \neq 0}$: training set contains anomalies

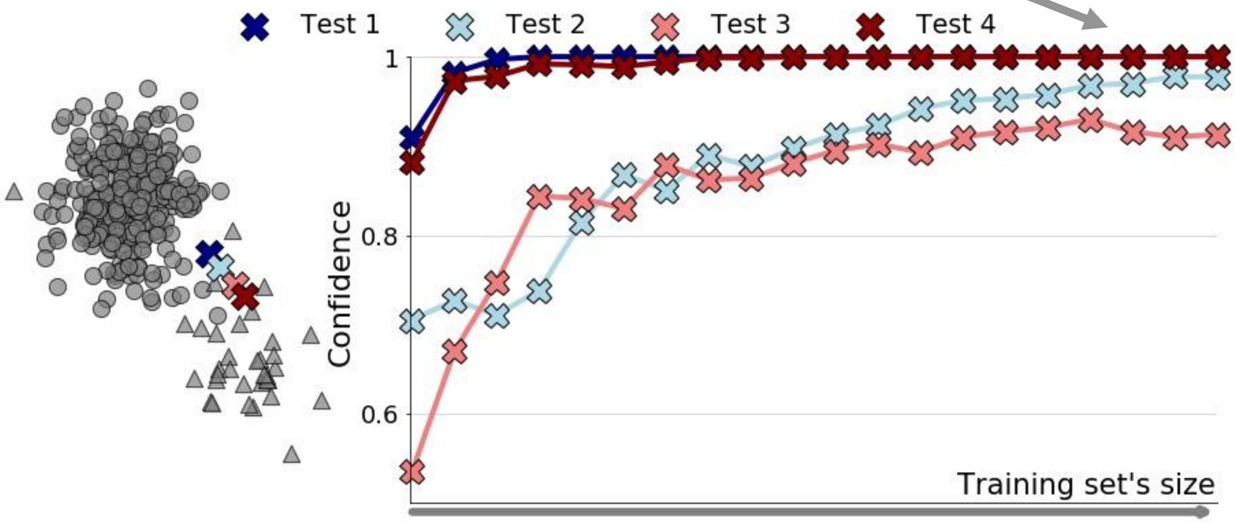
$\boldsymbol{\gamma = 0}$: training set with no anomalies

Convergence Behavior of Our Confidence Measure

Two cases according to $\boldsymbol{\gamma}$:

$\boldsymbol{\gamma \neq 0}$: training set contains anomalies

The confidence converges to 1



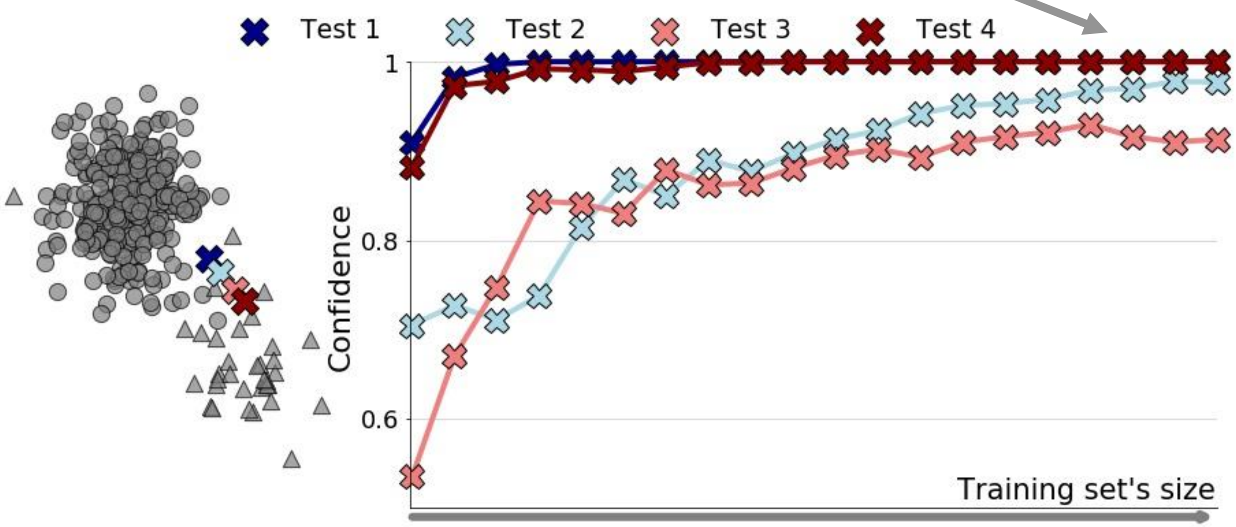
$\boldsymbol{\gamma = 0}$: training set with no anomalies

Convergence Behavior of Our Confidence Measure

Two cases according to $\boldsymbol{\gamma}$:

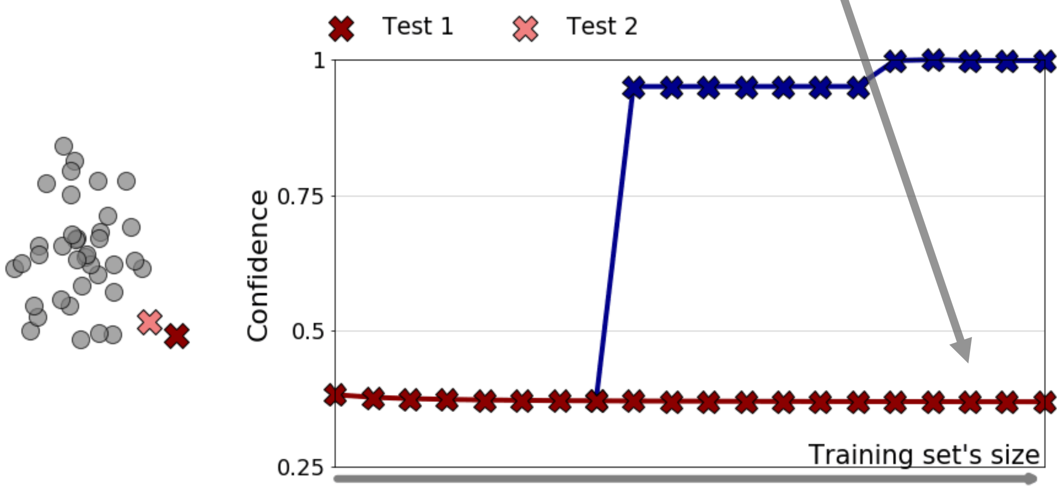
$\boldsymbol{\gamma \neq 0}$: training set contains anomalies

The confidence converges to 1



$\boldsymbol{\gamma = 0}$: training set with no anomalies

The confidence converges to $\boldsymbol{e^{-1} \approx 0.3679}$



Contribution 4: Experiments

An extensive empirical evaluation on benchmark datasets

A Large Experimental Comparison Shows that ExCeeD Recovers Confidence Matching Empirical Frequencies

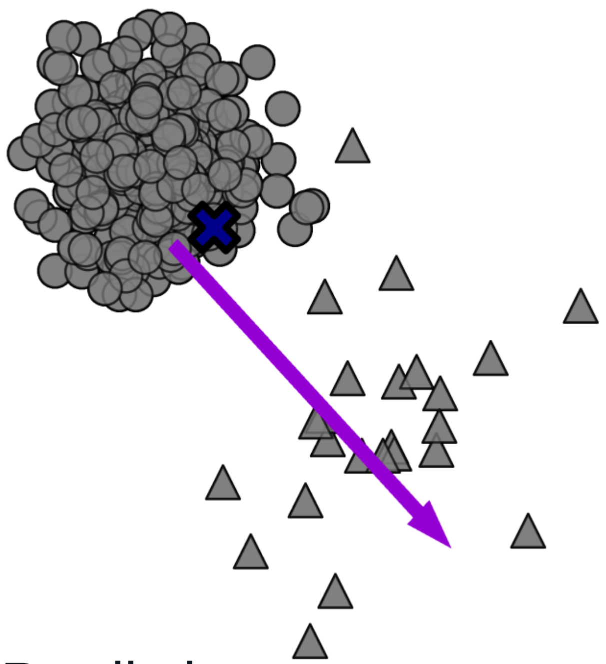
We compared ***ExCeeD*** with three types of baselines

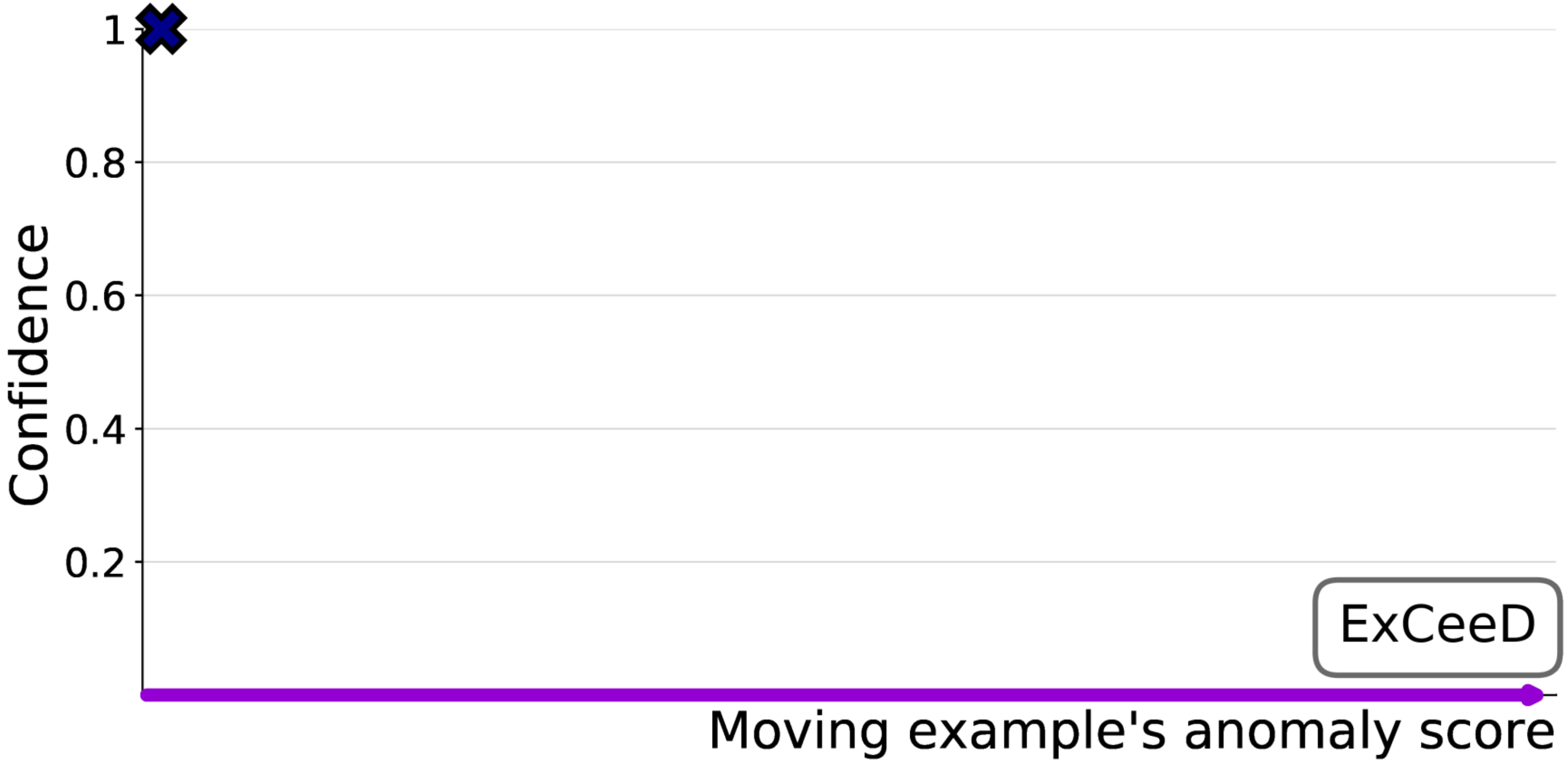
1. *Naive Baseline*, i.e. confidence always equal to 1;
2. *Outlier Probability methods*, i.e. apply existing methods instead of step 1;
3. *Calibration methods*, i.e. calibrated probabilities assuming to get a labeled dataset;

on ***21*** *benchmark datasets* and ***3*** *anomaly detectors*, and we found that:

1. Our outlier probability's estimate leads to more accurate confidence scores;
2. ExCeeD outperforms all the baselines, no matter of the anomaly detector used.

ExCeeD Captures Our Intuitions about How a Detector's Confidence in Its Prediction Varies as this Example Moves





Prediction:

- Normal
- Anomaly

In Conclusion, ExCeeD Is Not (Only) a Density Estimator

- We proposed a novel definition of *confidence* as the probability that the predicted class would change if a different training set was observed;

- We proposed *ExCeeD*, a method for estimating any anomaly detector's *confidence*;

- We proved that *ExCeeD*'s estimates of the confidence *converge*;

- Empirically, *ExCeeD* recovers *accurate* confidence scores.

All code and experiments are available online: *https://github.com/Lorenzo-Perini/Confidence_AD*

Contact us

name.surname@kuleuven.be

***Lorenzo Perini**, Vincent Vercruyssen, Jesse Davis*

ECML - PKDD 2020

https://people.cs.kuleuven.be/~lorenzo.perini/

@LorenzoPerini95