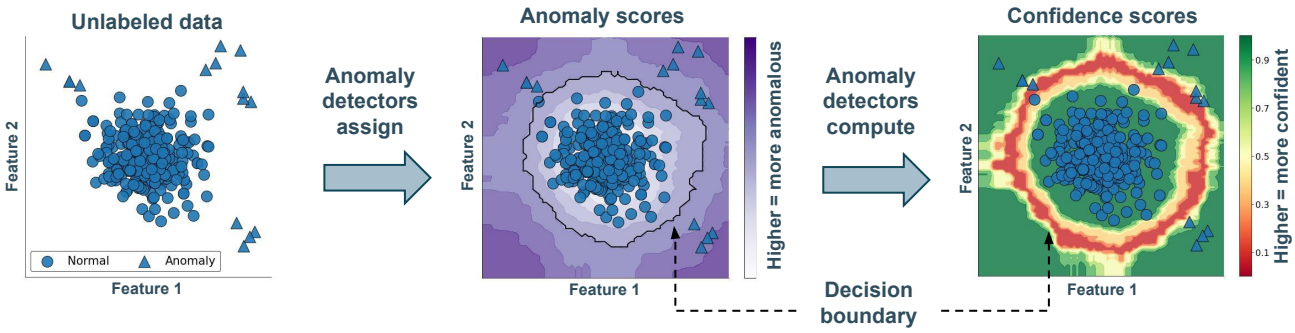
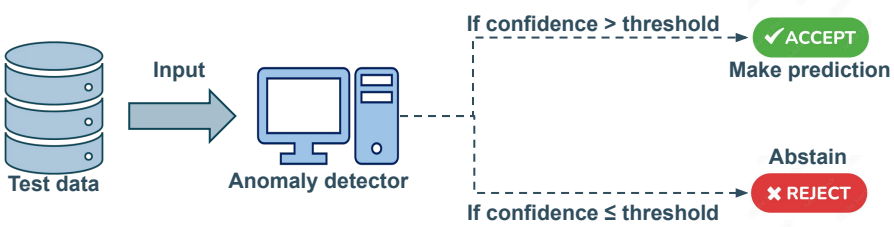
Unsupervised Anomaly Detection with Rejection

***Problem*: Unsupervised anomaly detectors have high uncertainty close to their decision boundary**



***Task*: Introduce a reject option to the anomaly detector, i.e. find a pair (confidence, rejection threshold)**

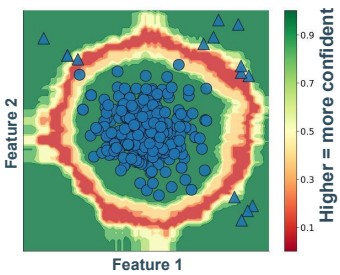


***Introducing a reject option with no labels is challenging*:**
1) traditional confidence metrics quantify how likely predictions are correct;
2) setting a rejection threshold requires evaluating the detector's performance.

} **Need labels!**

Our approach RejEx uses a *stability-based confidence metric* and sets a *constant rejection threshold* to 1-ε

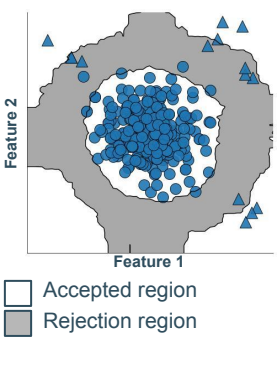
Theoretical analysis of ExCeeD[1], a stability-based unsupervised metric



We prove that:
1. Only few instances have confidence ≤ 1-ε;
2. Such instances surround the decision boundary;

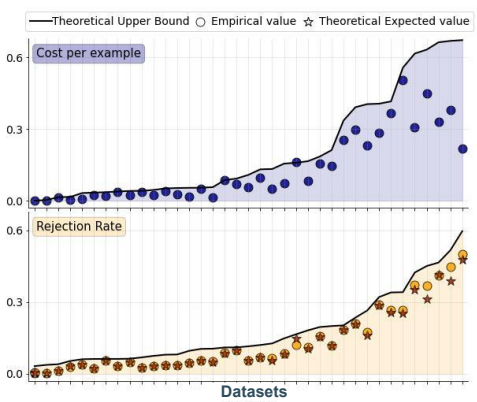
Contribution 1

We reject all instances with confidence ≤ 1-ε



Contribution 2

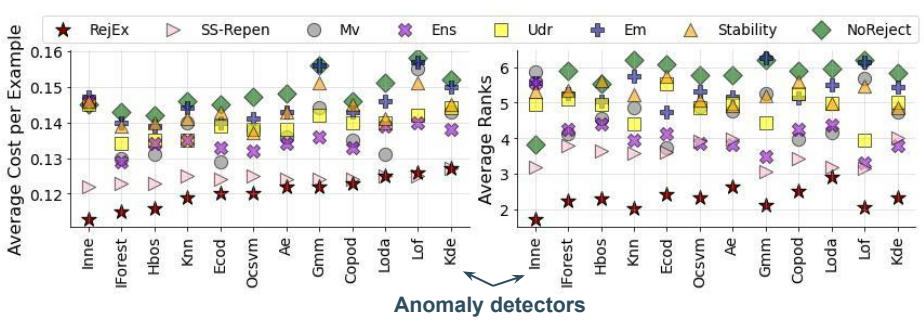
Our approach has several guarantees



Contribution 3

Experiments (34 datasets, 12 detectors) show that RejEx obtains a *lower (better) cost* in the majority of cases

Baselines:
- Self-Supervised (SS-Repen)
- Unsupervised metrics (Mv, Ens, Udr, Em, Stability)
- No rejection (NoReject)



$C_{fp} = 1$
$C_{fn} = 1$
$C_r = \gamma$

[1] Perini L., Vercruyssen V., Davis J., Quantifying the confidence of anomaly detectors in their example-wise predictions, ECML 2020.

Lorenzo Perini, Jesse Davis

@LorenzoPerini95

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

Check out the paper!

DTAI KU LEUVEN

fwo