

Time-series anomaly detection methods for unlabeled data.

Project Participants

Delaney Caceres dcaceres4 Researcher/Project Member	Cynthia Di cdi1 Researcher/Project Member
---	---

Project Description

Compare 3 anomaly detection algorithms for time series by utilizing F1 score, accuracy, and precision. This is an important problem because labeling unlabeled datasets gets incredibly tedious, especially if it's long. This is also important because networks and how we process data evolves very quickly, so a model must be updated with new training data which would thus involve gathering data and labeling it appropriately to avoid model drift. We will use skills learned in class to evaluate and compare time-series anomaly detection methods.

Related work:

- <https://arxiv.org/html/2504.02999v1>
- <https://pmc.ncbi.nlm.nih.gov/articles/PMC8752013/>

Goals:

- Researching the best time series algorithms/methods to compare unsupervised learning results to detect anomalies in network data.

Data

- An unlabeled public dataset for unsupervised training
- A labeled public dataset for testing
- Using our model, we'll also look at network statistics like bandwidth, packet counts and size, traffic volume, number of unique addresses, etc.

Deliverables

- A Jupyter notebook comparing different time series algorithms and their results
- A project report of our findings