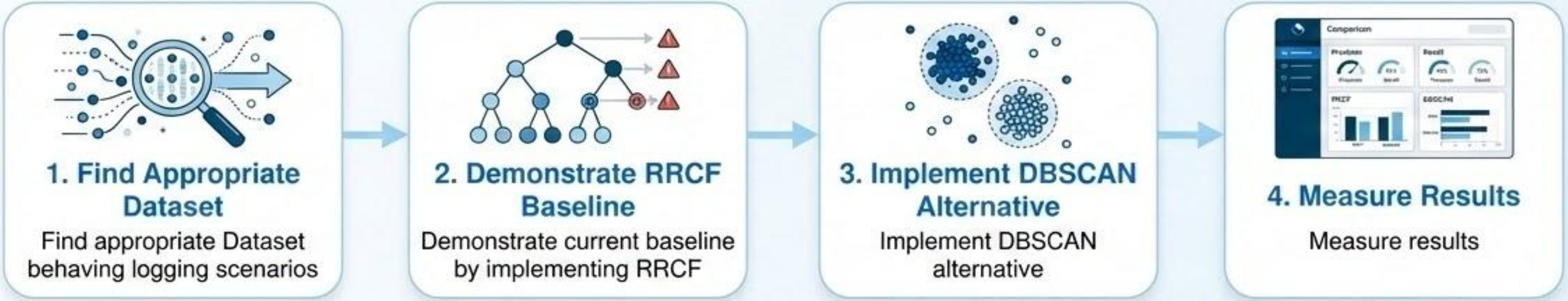
Time Series Anomaly Detection using DBSCAN

Evaluation Approach

The current Random Cut Forest (RCF) monitoring solution has a high false discovery rate (~40%), increasing operational overhead and diminishing system reliability. This has led to alert-desensitization, causing critical incidents to be missed among numerous false positives.

Evaluate a Density-Based Spatial Clustering of Applications with Noise (DBSCAN) tool to boost alerting precision without sacrificing recall.



1. Find Appropriate Dataset

Find appropriate Dataset behaving logging scenarios

2. Demonstrate RRCF Baseline

Demonstrate current baseline by implementing RRCF

3. Implement DBSCAN Alternative

Implement DBSCAN alternative

4. Measure Results

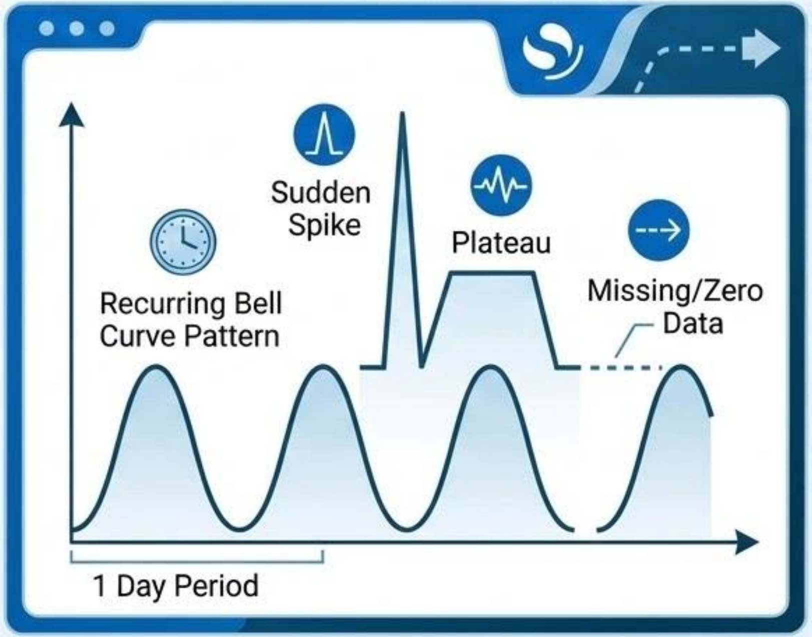
Measure results

Dataset Requirements

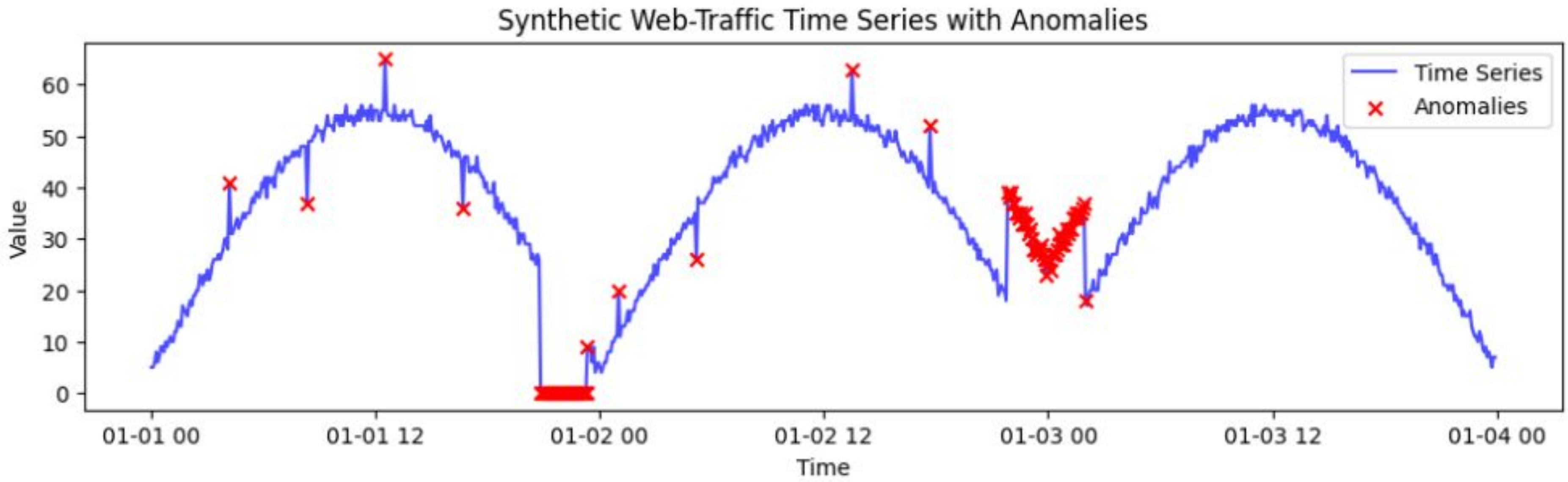
The time series must reflect the following criteria:

- Reflects a recurring normal/bell curve pattern
- Period of 1 day, to simulate the gradual increase/decrease of traffic throughout the day
- Sudden spikes, plateaus or missing/zero data

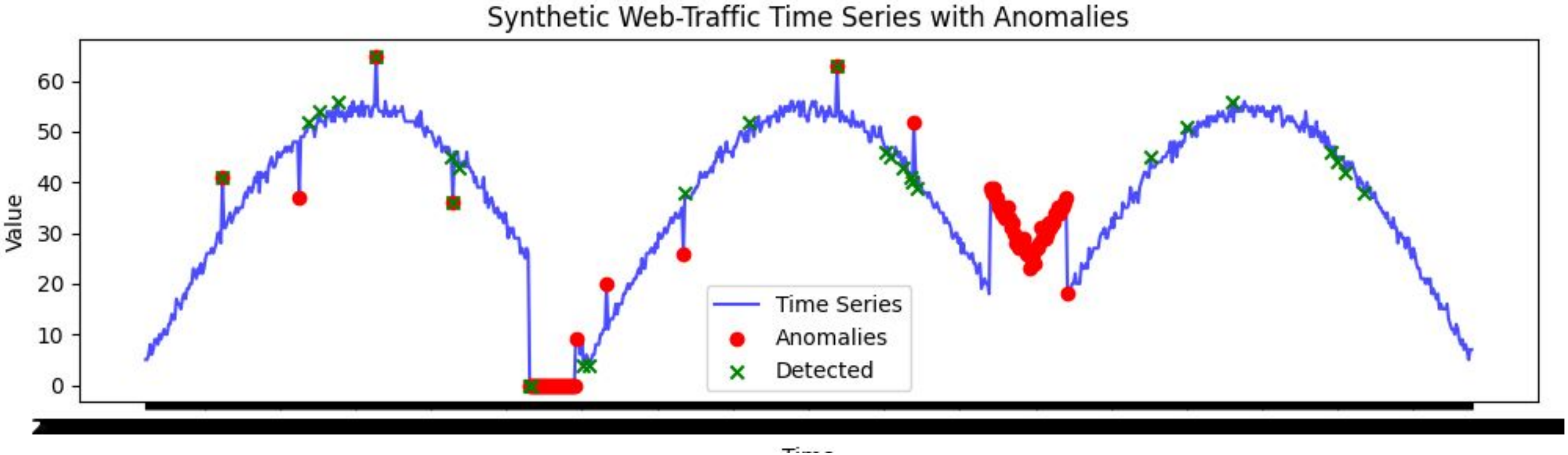
Unfortunately, we had to generate synthetic data to best represent experienced data flow but we evaluated some publicly available datasets.



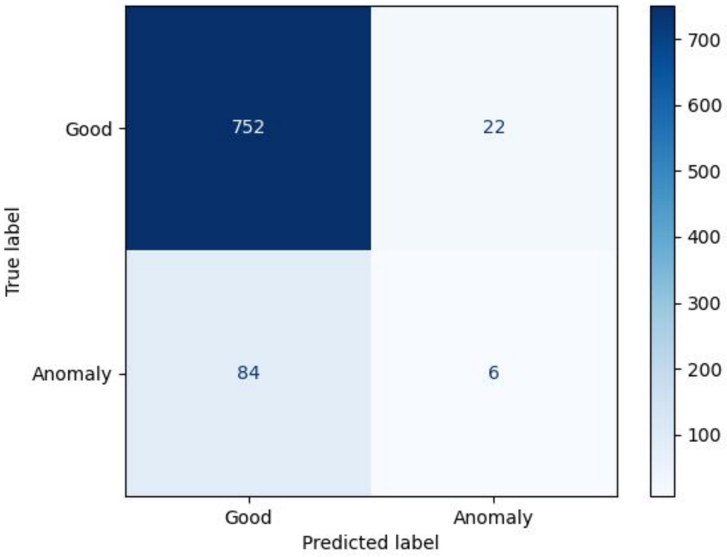
Generated DataSet



RRCF Performance

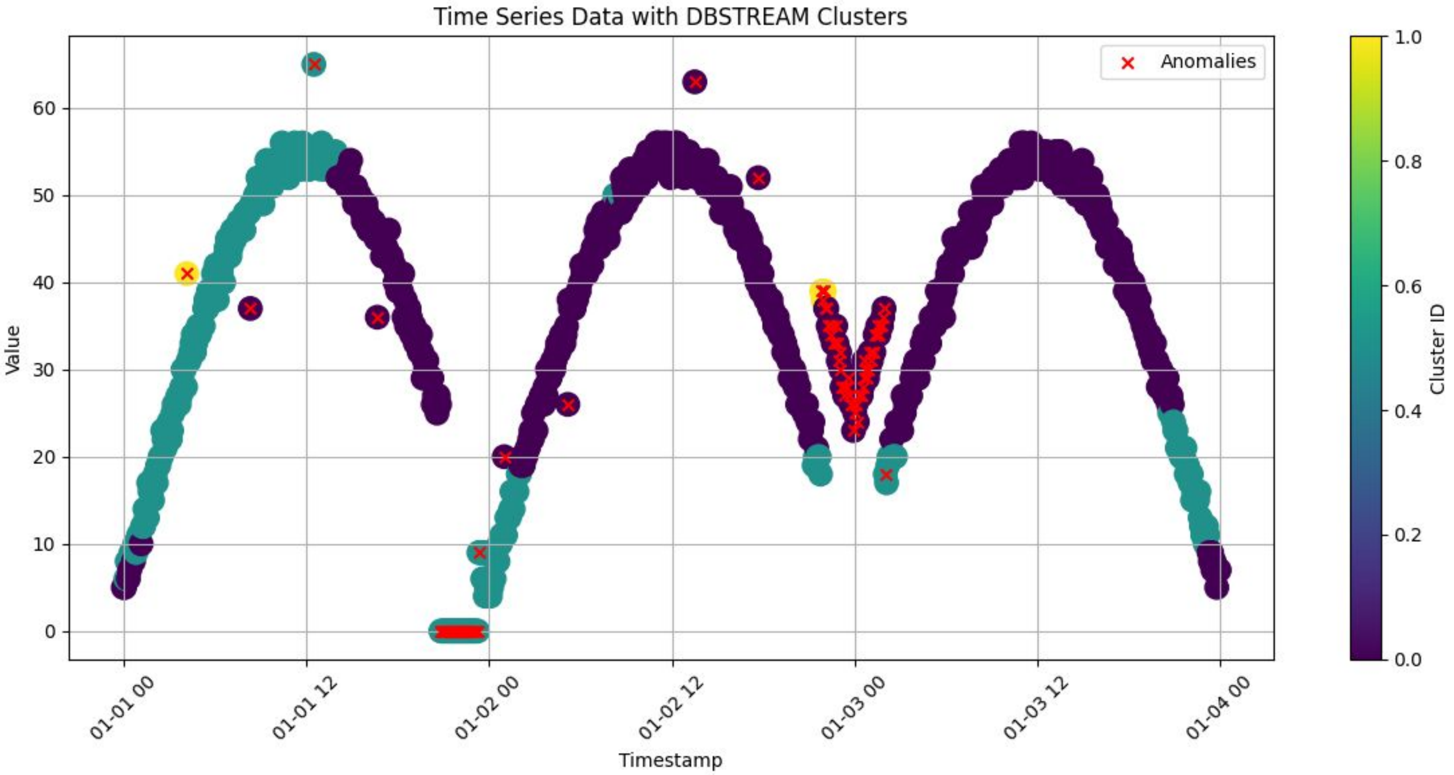


RRCF Performance



	precision	recall	f1-score	support
0	0.90	0.97	0.93	774
1	0.21	0.07	0.10	90
accucary			0.88	864
macro avg	0.56	0.52	0.52	864
weighted avg	0.83	0.88	0.85	864

DBSCAN Performance



DBSCAN Performance

1. Density-Based Grouping, Not Scoring

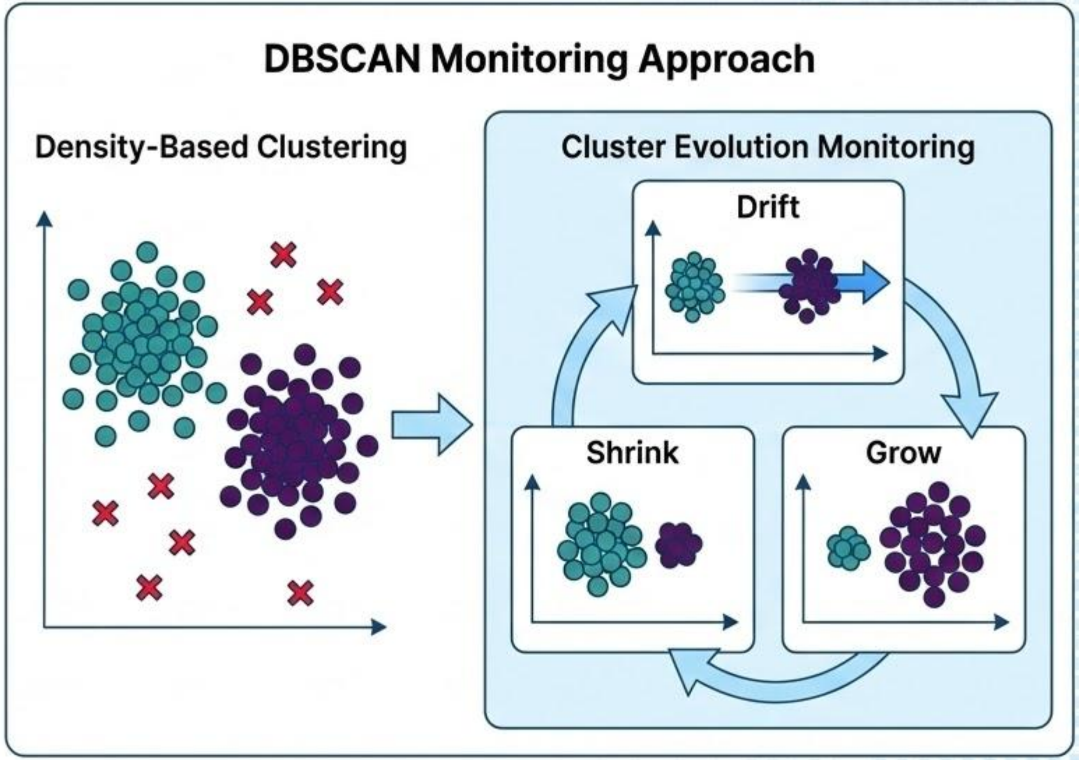
DBSCAN groups data based on density, rather than providing a direct anomaly "score".

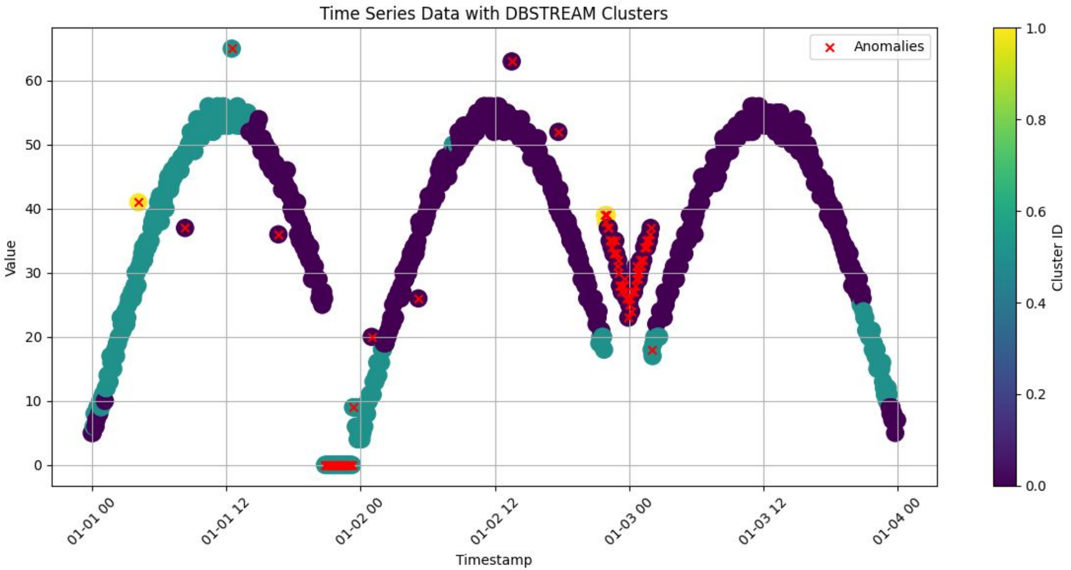
2. Track Cluster Changes

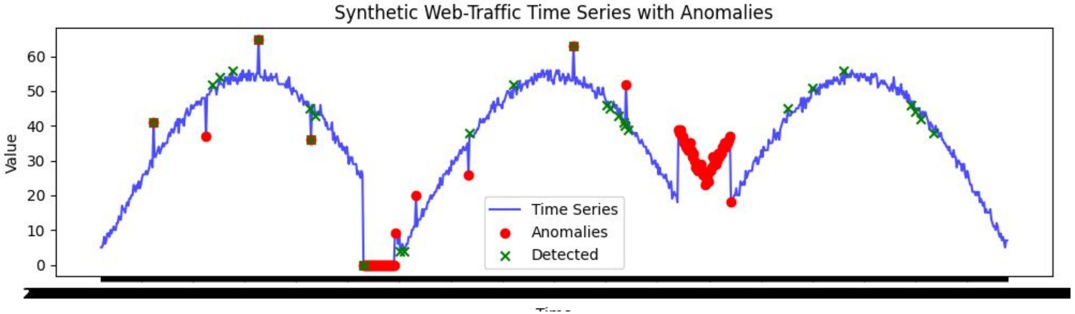
Move beyond simple outlier detection by monitoring how clusters evolve over time.

3. Monitor Structural Fluctuations

Observe cluster drift, shrinkage, or growth to detect significant system changes.




Time Series Data with DBSTREAM Clusters
Anomalies
Value
Timestamp
Cluster ID
0
10
20
30
40
50
60
01-01 00
01-01 12
01-02 00
01-02 12
01-03 00
01-03 12
01-04 00
0.0
0.2
0.4
0.6
0.8
1.0


Synthetic Web-Traffic Time Series with Anomalies
Value
Time Series
Anomalies
Detected
0
10
20
30
40
50
60

Recommendation

1. Proof-of-Concept Scope

Our proof-of-concept is just a fraction of what the tools is advertising.

2. Soft-Transition Pilot

We recommend to pilot a soft-transition to the new tool.