Metric departure on the Server Machine Dataset

28 production servers, 38 metrics each, five weeks per machine, with anomalies marked by the operators who ran them.

0.7747 pooled window ranking AUC over 176,950 windows, with 23 of 28 machines at 0.70 or better

Results

Measure	AUC
Pooled, 176,950 windows	**0.7747**
Median machine	0.8289
Best machine	0.9766
Weakest machine	0.5287

Bars run 0 to 1 on the ranking scale. The light tick marks 0.5, which is chance.

How it was measured

Unit	The machine
Atom	The change in each metric, not its level
Rule	A machine is scored by its worst metric
Learning	Normal operation only

Conditions

Machines at 0.70 or better	23 of 28
Naming the responsible metric	1.24 times better than chance

The pass mark, fixed before the run

> Set before the run: pooled AUC at or above 0.75, at least 19 of 28 machines at 0.70 or better, and attribution better than random. Verdict: PASS.

Limits of this result

Detection is stronger than explanation: the named metric is a starting point for an engineer, not a diagnosis. Five of the 28 machines fall below 0.70, and the weakest is listed rather than dropped.

Where it is recorded

simplex/validate/smd/PREREG.md, simplex/validate/BENCHMARKING.md. The data behind this sheet is published beside it at senuaai.com/evidence/datasheets/.