

Getting Data In: How to Prevent Data Downtime with ML

Learn how to create an ingest anomaly detection pipeline to proactively prevent data downtime during the GDI process. Leverage the Machine Learning Toolkit (MLTK) app to identify when data volumes coming into Splunk are outside of the expected range.



Anomaly Detection to Help Reduce Data Downtime in Data Ingest Pipelines

Today: **REACTIVE**

Troubleshooting what went wrong **after** your input data feed stops and affects the organization's work.

Data Sources

- Windows
- Linux/Unix
- Databases
- Applications
- Virtual Cloud
- Networking

Ingestion Methods

- Universal Forwarder
- HW Forwarder
- Intermediate Forwarder
- Technical Add-Ons
- Data Manager

Many things can go wrong with many data sources and ingestion mechanisms

With Anomaly Detection: **PROACTIVE**

Find and address abnormalities found during data ingestion **before** your pipeline gets disrupted or broken.

Ingestion Insights

Aggregate ingest volume data, and analyze the data to find anomalies.

Current anomaly score

Anomaly score from last time point



Anomaly History

Warnings

16[↑]₁

Percent Prediction Errors

Alerts

4[↑]₁

Anomaly Detection helps monitor input feeds continuously for insights

How to Prevent Data Downtime with ML Anomaly Detection

Step 1

Download
the MLTK
app

Machine Learning Toolkit (MLTK) app

Download and install the free MLTK app on Splunkbase to get started ([docs](#))

Step 2

Follow
Step-by-Step
Instructions

Open Tutorial Doc

Deep dive: create a data ingest anomaly detection dashboard using ML-SPL commands

Use the documentation to copy the SPL search commands that you'll need for the video tutorial in step 3.

Step 3

Implement the
Anomaly Detection
Technique

Video Tutorial - Follow Along!

Go to the **next slide** and follow along while a Splunk expert shows you how to operationalize the ingest anomaly detection technique...

Video Demo: Follow Along Tutorial

How to set up an automated alerting system that detects unexpected downtimes or spikes in data ingestion volumes.

splunk>

Products ▾

Solutions ▾

Why Splunk? ▾

Resources ▾

Splexicon

Support ▾



Splunk® Machine Learning Toolkit

User Guide

Download manual as PDF

Product

Splunk® Machine Learning Toolkit ▾

Version

5.3.1 (latest release) ▾

≡ Hide Contents ▾

User Guide

➤ Introduction to the Machine Learning Toolkit

➤ MLTK guided workflows

➤ MLTK commands, macros, and visualizations

➤ Algorithms and scoring metrics in the MLTK

➤ Install and upgrade the MLTK

➤ Prepare and preprocess your data

➤ Smart Assistant guided workflows

➤ Experiment Assistant guided workflows

➤ Classic Assistant guided workflows

➤ MLTK models

Additional resources

Share data in the Machine Learning Toolkit

Learn more about the Machine Learning Toolkit

Documentation

Splunk® Machine Learning Toolkit

User Guide

Create a data ingest anomaly detection dashboard using ML-SPL commands

Download topic as PDF

Create a data ingest anomaly detection dashboard using ML-SPL commands

Your data ingest pipelines can be impacted by traffic spikes, network issues, misconfiguration issues, and even bugs. These issues can cause unexpected downtime and negatively impact your organization. One option for accessing real-time insights on data ingestion is through a Splunk dashboard. The data ingest anomaly detection dashboard uses familiar ML-SPL commands like `fit` and `apply` to monitor your data ingest volume.

Prerequisites

- You must have the MLTK app installed. MLTK is a free app available on Splunkbase. For detailed installation steps, see [Install the Machine Learning Toolkit](#).
- You must have domain knowledge, SPL knowledge, and Splunk platform experience, in order to make the required inputs and edits to the SPL queries provided.

Create the dashboard

Perform the following high-level steps to create a data ingest anomaly detection dashboard:

1. Run an SPL search to fit the model.
2. Run an SPL search to apply the model.
3. Run an SPL search to calculate Z-scores.
4. Create the Ingestion Insights dashboard.
5. (Optional) Manually populate dashboard charts with historical data.
6. Set up alerts.

Run an SPL search to fit the model

Perform the following steps to run the SPL search to fit the model:

Create a data ingest anomaly detection dashboard using ML-SPL commands

▮ Prerequisites

▮ Create the dashboard

▮ Run an SPL search to fit the model

▮ Run an SPL search to apply the model

▮ Run an SPL search to calculate Z-scores

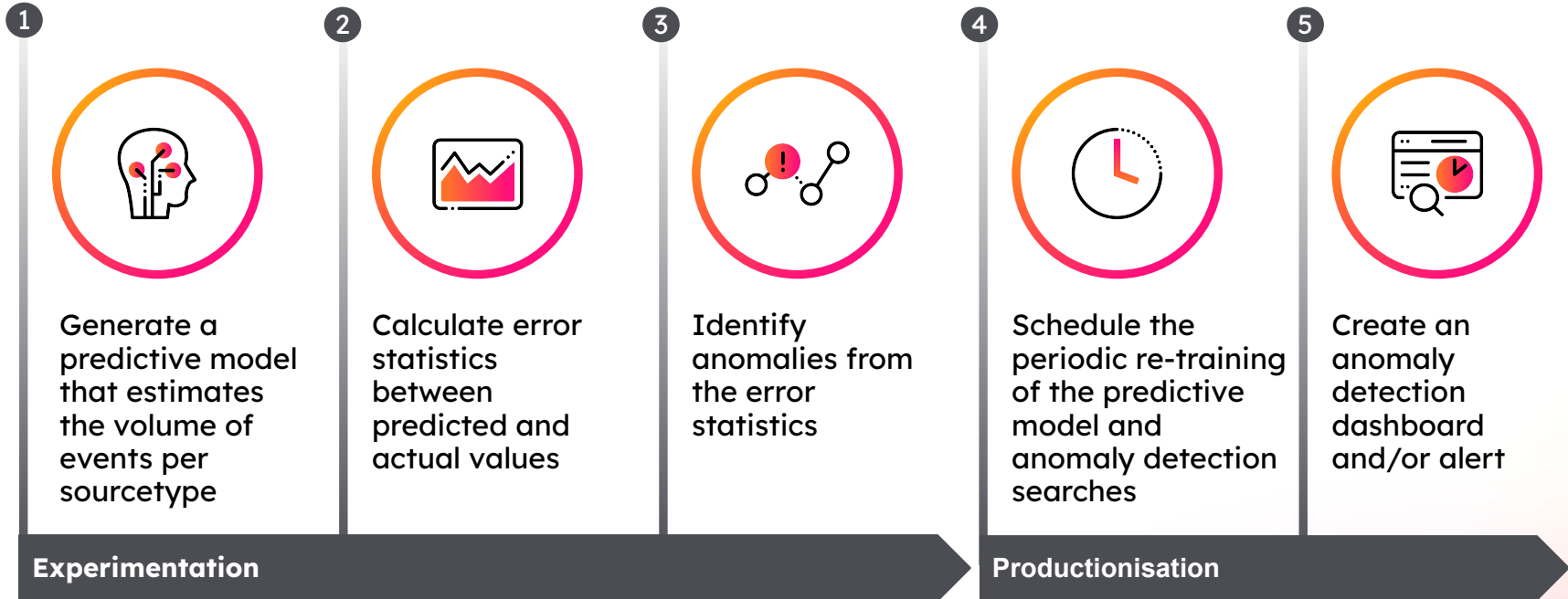
▮ Create the Ingestion Insights dashboard

▮ (Optional) Manually populate dashboard charts with historical data

▮ Set up alerts

▮ Learn more

Recap: Process to create an ingest anomaly detection pipeline to proactively reduce downtime



Additional Resources to Kickstart your ML Journey

- Full webinar: [Prevent Data Downtime with ML](#)
- Blog: [Getting Started with ML at Splunk](#)
- Blog: [Prevent Data Downtime with Anomaly Detection](#)
- Docs: [MLTK - Smart Outlier Detection Assistant](#)
- Docs: [Understand the Fit and Apply Commands](#)
- Splunk Community Answers: [MLTK conversations](#)

