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

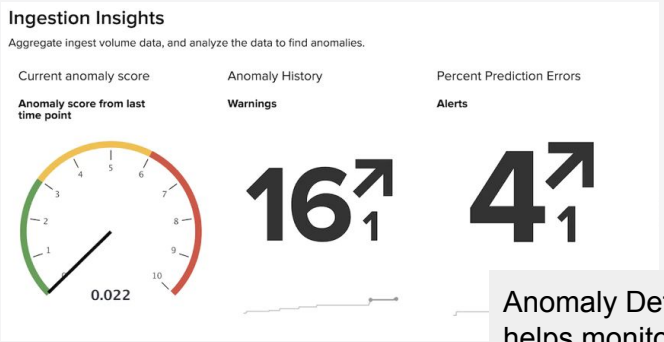
Ingestion Methods

- 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.



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

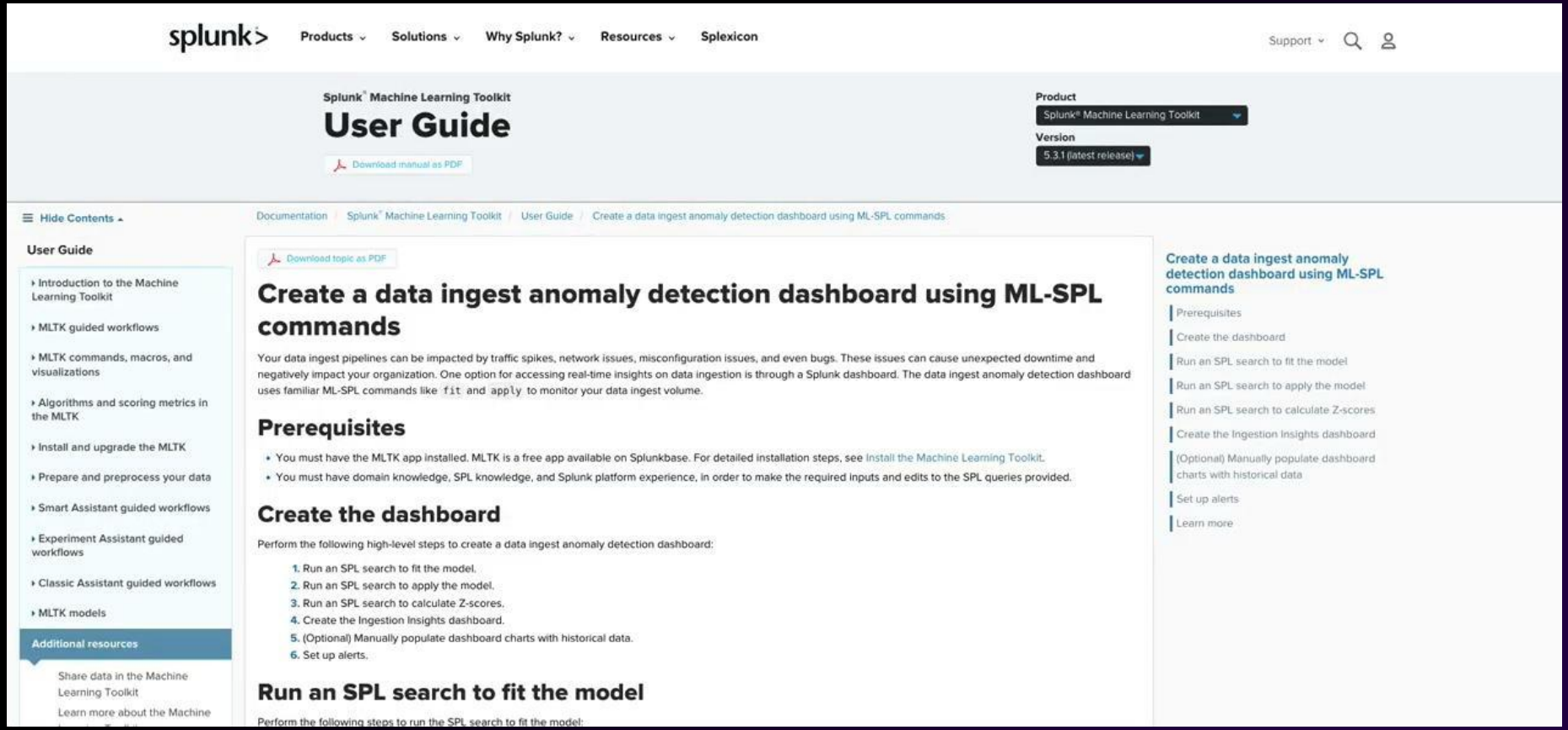
Implement the Anomaly Detection Technique

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.



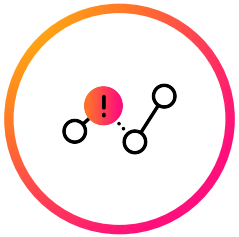
Recap: Process to create an ingest anomaly detection pipeline to proactively reduce downtime

1. 

   Generate a predictive model that estimates the volume of events per sourcetype

2. 

   Calculate error statistics between predicted and actual values

3. 

   Identify anomalies from the error statistics

4. 

   Schedule the periodic re-training of the predictive model and anomaly detection searches

5. 

   Create an anomaly detection dashboard and/or alert

Experimentation | **Productionisation**

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

splunk>