An Introductory Use Case Guide

Splunk Artificial Intelligence for Observability

splunk>

Table of contents

Introduction

This e-book is designed to help readers looking for ways to get value from implementing artificial intelligence (AI) or machine learning (ML) in Splunk, outlining example use cases that have been successfully implemented elsewhere. For each use case, the business challenge, approach to using Splunk and value will be presented. Additionally, each use case includes links to supporting information such as customer case studies or documentation to help reproduce the use case in the Splunk environment.

What are artificial intelligence (AI) and machine learning (ML)?

The term "machine learning" is often used interchangeably with the term "artificial intelligence," but ML is a subfield of AI. ML is a field of computer science that develops computer systems that can autonomously learn from experience by processing the data they receive and improving the performance of specific tasks.

Artificial intelligence is the ability of a system to handle representations, both explicit and implicit, and to perform tasks that would be considered intelligent if performed by a human.

Machine learning is the ability for computer systems to use algorithms and statistical models to continuously improve the performance of specific tasks.

Deep learning is a specialized type of ML algorithm designed to mimic a human brain's neural network, allowing machines to use massive amounts of data to learn from their own actions and improve future outcomes. An example of Deep Learning includes Large Language Models (LLMs) such as those used by ChatGPT.

Why do organizations invest in artificial intelligence?

The past few years have seen organizations have to cope with disruption on a global scale, with business resilience being tested like never before. As noted in our Digital Resilience Pays Off report, being able to prepare for change is a key factor in building resilience and the ability to thrive during uncertain times.

One subject that is often associated with change and innovation is ML. Specifically for observability, the ability to predict and prevent incidents before they occur is one of the key areas for driving value with ML; companies that can prevent downtime have much greater resilience than those who are reactive to downtime. Organizations that adopt ML and auto remediation across all their products and services are twice as likely (66%) to be prepared for the demands of a recession, compared to those that do not (34%).

Splunk AI portfolio

Splunk provides a number of ways of utilizing AI/ML across the product portfolio. Broadly there are two ways of using AI/ML: using out-of-the-box features that are deeply integrated into existing product workflows, or through customization.

ML is embedded into the Splunk platform within Splunk Cloud Platform and Splunk Enterprise, allowing users to:

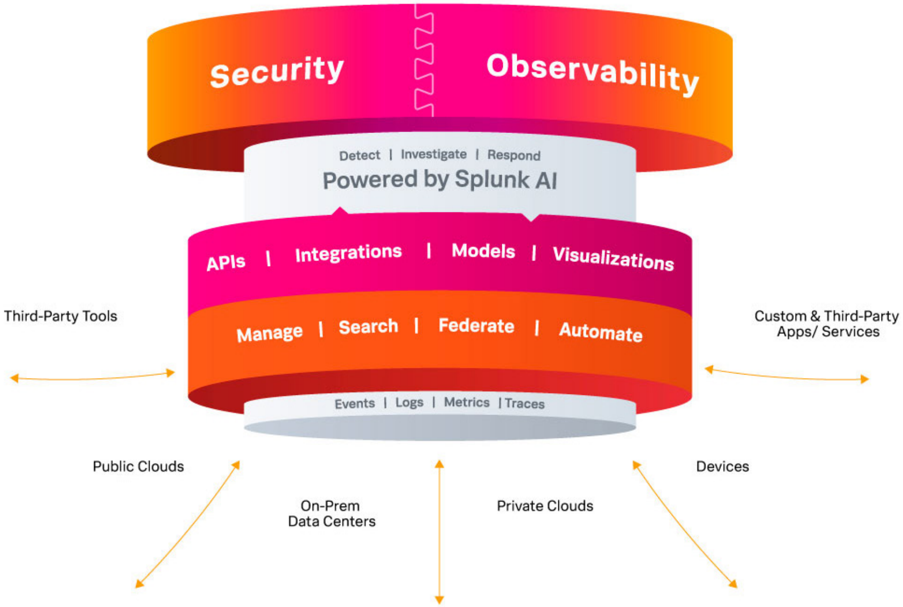
- **Detect anomalies**, such as identifying outliers in the number of application errors.
- **Generate forecasts**, for example forecasting resource utilization.
- **Make predictions**, like predicting potential outages.
- **Cluster data into groups**, for instance, clustering network activity to detect potentially misconfigured services.

These techniques can be applied via assistants that guide the user through a series of steps to train, assess and operationalize ML models. Alternatively, ML-based analytics can be created directly using Splunk's search language — Search Processing Language (SPL) — with a number of ML search commands embedded into core search and reporting, such as predict. The patterns tab in the search and reporting app also presents embedded machine learning to help identify groups of similar events in search results. In addition to the core platform, Splunk also provides ML-powered experiences in the following products:

- Out-of-the-box ML analytics in Enterprise Security, a market leading Security Incident and Event Management (SIEM) platform
- Pre-defined threat detection modeling in User Behaviour Analytics, designed to identify advanced persistent threats and insider threats
- Workflows in IT Service Intelligence — an AIOps solution — for creating adaptive thresholds for key metrics, as well as predicting potential outages
- Assistive wizards in Splunk Infrastructure Monitoring to detect outliers in metrics or predict when resource utilization thresholds will be crossed

We also offer assistive intelligence experiences to provide personalized guidance. The Splunk AI Assistant uses generative AI to provide a chat experience that helps users author and learn SPL by interacting in plain English and providing query suggestions, explanations and detailed breakdowns. The Splunk App for Anomaly Detection enables Splunk users to detect anomalies in their time series data sets and metrics using powerful machine learning algorithms in just a few clicks, while providing an end-to-end operationalization workflow to streamline creating and running anomaly detection jobs.

The Splunk platform can be extended with add-ons that are designed specifically for running ML workloads, namely the Machine Learning Toolkit and the Splunk App for Data Science and Deep Learning. The Splunk Machine Learning Toolkit (MLTK) provides SPL commands, custom visualizations, assistants, and examples to explore a variety of ML concepts all inside the Splunk platform. Extending beyond MLTK, the Splunk App for Data Science and Deep Learning (DSDL) provides the ability to integrate advanced custom ML and deep learning systems with the Splunk platform.


Security
Observability
Detect | Investigate | Respond
Powered by Splunk AI
APIs | Integrations | Models | Visualizations
Manage | Search | Federate | Automate
Events | Logs | Metrics | Traces
Third-Party Tools
Custom & Third-Party Apps/ Services
Public Clouds
On-Prem Data Centers
Private Clouds
Devices

Foundational elements for observability

Observability is a modern approach to monitoring that provides complete visibility and context across the full stack of infrastructure, applications and the customer experience. In addition to Splunk receiving numerous TrustRadius awards for Event Analytics, GigaOm named Splunk a leader in cloud observability and Gartner listed Splunk as a leader in application performance monitoring (APM) and observability. Underpinning this success is the core data platform, which allows querying and visualizing log, metric and trace data in near real time.

Observability's goal of providing complete visibility and context across the full stack is underpinned by Splunk's data schema, the Common Information Model (CIM). With the CIM machine generated data can be normalized across different data sources to provide a holistic view of activity. Additionally, data can be enriched with contextual information such as business unit information, asset data or configuration details. Getting data foundations right with the CIM and contextual enrichment is critical for advancing operational monitoring.

AIOps (Artificial Intelligence for IT Operations) is a mature approach to Observability, incident management and response. Splunk's AIOps solution — IT Service Intelligence (ITSI) — has several areas that use ML: Adaptive Thresholding, used to generate baselines that describe typical key performance indicator (KPI) behavior; Anomaly Detection, used to identify unusual KPI readings; Predictive Analytics, used to predict future service health scores; and Smart Mode, used to automatically group related alerts.

Monitoring complex systems beyond the scope and ability of a human is another strong use case of ML and one that is provided in Splunk Infrastructure Monitoring. Splunk Infrastructure Monitoring is designed to address real-time cloud monitoring requirements at scale, focusing on monitoring metrics and traces. A number of features use ML such as anomaly detection assistants, which provide a guided workflow to configure alerting based on metrics that have anomalous values. Additionally, alerts can be set up via an assistant for predicting when resources, such as disk space, will cross thresholds.

Considerations at the start of an AI journey

Before starting a new AI project, Splunk always recommends running a brief impact assessment to help prepare for success. At a minimum, the impact assessment scope should consider objectives, risk and execution capacity. Below are some main things to consider.

Assess the objectives

87% of data science projects fail to make it to production,[1] highlighting the importance of defining clear outcomes to make an ML project successful. At a high level, these outcomes often fall into the following categories:

- Increasing detection efficiency
- Reducing manual processing or minimizing human error
- Identifying previously unknown scenarios

The most successful ML projects are often tied to granular outcomes, such as increasing detection accuracy by 70% for alerts related to application errors or reducing manual triage time for Network Operations Center (NOC) analysts by 50% when assessing alert storms.

[1] https://venturebeat.com/ai/why-do-87-of-data-science-projects-never-make-it-into-production/

When creating an ML Project, developing business outcomes and success metrics are important considerations. Typical questions to consider when determining an objective for an ML project are:

Increasing detection efficiency

a. Are current true positive and true negative rates known for existing detections? If so, are there understood business benefits from improving these rates, for example by improving detection accuracy analysts will not need to spend as much time triaging false positives.

b. Are there certain alerts that trigger frequently, generating a lot of noise for the NOC? If so, improving the accuracy of these alerts could improve overall NOC efficiency.

c. Can target benchmarks be set to reduce false positives or improve alert accuracy?

Reducing manual processing or minimizing human error

a. Is there case management data available detailing the amount of time NOC analysts spend triaging alerts? This information can help identify alerts that could benefit from reduced triage times.

b. Are there alerts that are ignored or closed in high volumes without triage? This may indicate situations where potential threats are being missed due to noisy alerts.

c. Are there business objectives set for the amount of time analysts should spend improving NOC capabilities? Automating routine tasks to free up time for higher value activity is one mechanism for achieving this.

Identifying previously unknown scenarios

a. Have there been historic service incidents that were not identified by existing triggering alerts?

b. Are there emerging technologies being adopted that are not well understood by the business?

Assess the risk

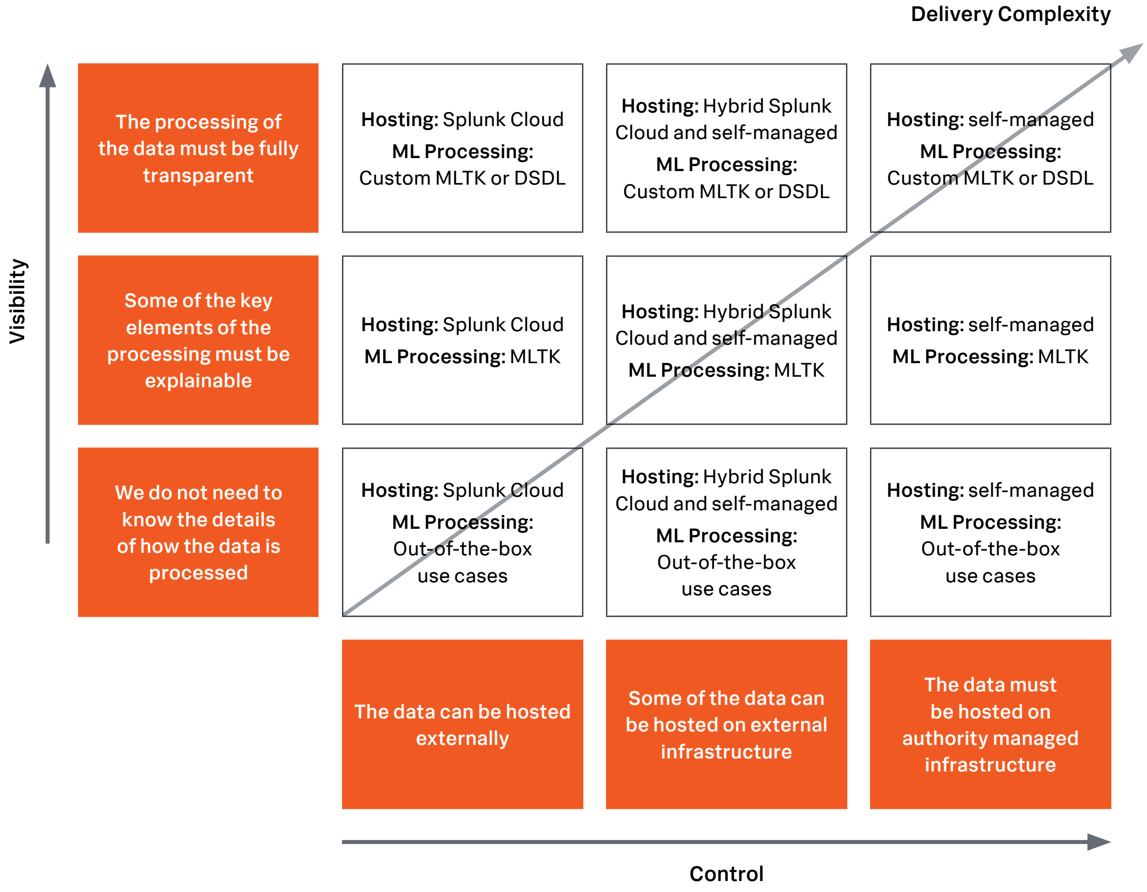
Guidance published by the World Economic Forum[2] notes that risks can be associated with using AI/ML techniques. Often, these risks are related to the difficulty of explaining the processing of data. Users of AI systems are often unsure how a particular output has been generated, which makes it difficult to determine the correct action to take.

Three areas that are important when considering risk of an AI system are:

- **Visibility:** How much detail is required on how the data is being processed by an AI system? ML algorithms use complex mathematics when processing input data to generate an output, which can often make it difficult to understand why a particular output has been generated.
- **Control:** What requirements for the hosting of data are in place in a given organization? In the public sector and in highly regulated industries many organizations have requirements to host certain types of data on authority infrastructure, making use of cloud services challenging.
- **Failure Tolerance:** What will be the consequence of a failed AI project? Investment in AI is no guarantee for success, therefore consider how much flexibility there is in the business objectives and the downstream impact on the NOC team while working on the AI project.

[2] https://www.weforum.org/reports/ai-procurement-in-a-box/

The matrix below illustrates how some of the considerations about visibility and control of data might impact the choice of Splunk products (please refer to section A of AI Procurement in a Box Workbook[3] for further questions to assess risk against visibility and control). Note that this is suggested guidance, and where appropriate, MLTK or an out-of-the-box use case may be preferable depending on risk appetite or the need for customization.



	The data can be hosted externally	Some of the data can be hosted on external infrastructure	The data must be hosted on authority managed infrastructure
The processing of the data must be fully transparent	**Hosting:** Splunk Cloud **ML Processing:** Custom MLTK or DSDL	**Hosting:** Hybrid Splunk Cloud and self-managed **ML Processing:** Custom MLTK or DSDL	**Hosting:** self-managed **ML Processing:** Custom MLTK or DSDL
Some of the key elements of the processing must be explainable	**Hosting:** Splunk Cloud **ML Processing:** MLTK	**Hosting:** Hybrid Splunk Cloud and self-managed **ML Processing:** MLTK	**Hosting:** self-managed **ML Processing:** MLTK
We do not need to know the details of how the data is processed	**Hosting:** Splunk Cloud **ML Processing:** Out-of-the-box use cases	**Hosting:** Hybrid Splunk Cloud and self-managed **ML Processing:** Out-of-the-box use cases	**Hosting:** self-managed **ML Processing:** Out-of-the-box use cases

[3] https://www3.weforum.org/docs/WEF_AI_Procurement_in_a_Box_Workbook_2020.pdf

Assess execution capacity

Provided a clear objective can be set for an ML project, the requirements for executing the project should also be evaluated. The list below is not exhaustive or prescriptive, but presents some of the areas to think about before embarking on an experience with ML.

People:

a. Are subject matter experts on the data available to provide guidance on the meaning and quality of the data?

b. What ML expertise is available to help develop and support these use cases in the future?

c. If there is no ML expertise, can partners like Splunk provide the right resources to upskill or guide the project?

d. Is there capacity in the NOC to handle additional workloads from the ML project?

Process:

a. Who will use the analysis and how will they use it in their daily functions?

b. How will the use case be operationalized?

c. How will false positives be handled?

d. Are clear next steps identified if the ML analytic identifies a true positive?

Information:

a. Is the data required for analysis already indexed in Splunk or are there plans in place for getting it indexed?

b. Are there any special handling requirements for the data, for example does personal data need to be obfuscated?

c. Are there blogs and content available about the use case on which the project is based?

d. Does the output of the ML processing need to be explainable to a non-technical end user?

Technology:

a. Is ML necessary to solve the use case or can a correlation rule or basic statistics suffice? To determine this performance analysis comparing workloads to accuracy using different methodologies may be required, also looking at how users will consume the outputs. (Simpler methods are often more explainable.)

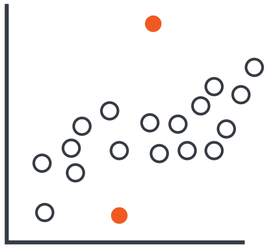
b. Will the current Splunk architecture need updating to handle special ML processing, such as introducing a dedicated search head for training ML models?

Use cases

This section outlines introductory use cases for using ML for observability across information technology (IT) and operational technology (OT) domains. Each use case details the business challenge, Splunk's approach to solving it and the value that can be realized from implementing the use case. Where available, links are included to case studies or additional information, such as MLTK docs pages, that can help with getting started implementing the use case in a Splunk environment.

Although this document focuses on use cases, typically ML-based analytics in Splunk use one of three common techniques: anomaly detection, predictive analytics and clustering.

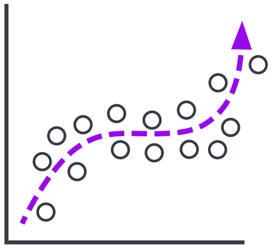
Anomaly Detection



Deviation from past behaviors

Resource utilization
Error rate deviation
Access pattern baselining

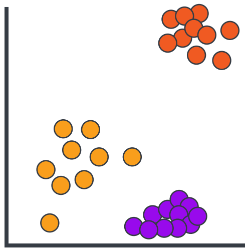
Predictive Analytics



Future state prediction

Predict storage requirements
Identify patterns leading to failure

Clustering



Grouping based on behaviors

Identify traffic
Identify common behaviors

The table below shows how the following use cases map to the different techniques.

Use Case	Anomaly Detection	Predictive Analytics	Clustering
1. Forecasting Resource Utilization		✔	
2. Detecting Service Performance Issues	✔	✔	
3. User Experience Monitoring	✔	✔	
4. Noise Reduction			✔
5. Predicting Data Downtime in Splunk	✔	✔	
6. Predictive Maintenance	✔	✔	✔
7. Cell Tower Monitoring	✔		
8. Geohazards Monitoring	✔		

1. Forecasting resource utilization

Business challenge

According to some reports,[4] roughly a quarter of incidents of downtime are caused by inadequate server hardware, compounded by half of incidents being affected by understaffed IT departments and the complexity of provisioning new services. Getting ahead of these types of incidents and reducing the workload on overburdened IT teams is an important factor in reducing downtime, but predicting resource utilization of complex IT systems is a challenge.

Splunk's approach

With Splunk's MLTK, users can create analytics for forecasting or predicting resource utilization using a set of guided workflows or directly from the search bar.

Additionally, alerts can be set in Splunk Infrastructure Monitoring for conditions like resources running out, based on predictions that use double exponential smoothing to forecast utilization. These predictions are generated based on a guided workflow, without the need for data science expertise to create the predictions.

Value

Increase detection efficiency: Being able to predict the overutilization of resources can help avoid potential outages.

Reduce manual processing: Furthermore, helping IT teams get ahead of situations where resources will run too hot can improve efficiency, alerting teams to degradation ahead of labor intensive outages.

Case studies and further information

TransUnion
As a trailblazer in predictive analytics in Splunk ITSI, TransUnion has been pushing the boundaries of what is possible when it comes to predicting resource utilization.

TransUnion commonly faces disk space issues, with 30% of all database administrator (DBA) incidents related to disk space. In the past, the triage workflow took 20 minutes for a DBA to respond to a ticket, assess the issue, and then assign a ticket requesting a specific amount of additional space to the server storage if required. Using MLTK, they have reduced the triage time by 65%, presenting the DBA team with a set of servers that are projected to run out of space within a month — the lead time for getting increased storage space provisioned.

Read more about TransUnion's experience forecasting resource utilization here, including a detailed set of searches for recreating their analytics.

IKEA
IKEA is another customer who has invested time using MLTK to determine potential database outages based on looking at the amount of free disk space.

Initially creating experiments in MLTK, IKEA incorporated a successful technique into their Splunk ITSI workflows to simplify the experience for their end user analysts. The ability to augment Splunk ITSI KPIs, which operate at an aggregated entity level with per entity analytics using Splunk Enterprise, provided the granularity IKEA needed for certain use cases like predicting free disk space.

Find out more about IKEA's use of MLTK and Splunk ITSI here.

[4] https://itic-corp.com/security-data-breaches-top-cause-of-downtime-in-2022/

2. Detecting service performance issues

Business challenge

Service outages can be costly, with our Digital Resilience Pays Off report finding that each hour of downtime can cost $365,000. Avoiding outages is therefore important for maintaining revenue, as well as for avoiding reputational damage. For companies who have complex technology stacks, run many thousands of customer-facing applications, or have very high volumes of customers using their platforms, it can often be difficult to proactively spot outage conditions.

Splunk's approach

Indicators of a service outage can often come from metrics like the number of application error messages or network response times. Using Splunk's Common Information Model these types of metrics can be normalized to provide consistency across different infrastructure, application, and cloud stacks.

With Splunk ITSI, users can set up adaptive thresholds for KPIs like application error rates out-of-the-box by simply clicking through the Splunk ITSI user interface (UI) to set up baseline thresholds for their applications. Splunk ITSI also has assistants that can run anomaly detection against KPIs like application error rates that will identify unusual error rates for a given application by looking at historic trends and patterns.

In addition to monitoring single KPIs, Splunk ITSI also has features to train predictive models on services to predict future health scores. This allows the prediction of potential outages before they occur.

Splunk Infrastructure Monitoring also provides a range of ML techniques for identifying unusual rates of metrics like session durations without having to be a data scientist.

Moreover, customers can create their own bespoke analytics for detecting unusual error message rates using MLTK, which contains multiple UI based assistants and algorithms for creating ML models.

Value

Increase detection efficiency: Uncovering application performance issues allows the user to get ahead of potential outages and prevent downtime. Furthermore, spotting periods of time where applications are showing signs of poor performance or degradation can prevent poor customer experiences — either by proactively reporting these issues to affected customers or addressing the cause of the issues before customers notice.

Reduce manual processing: Programmatically identifying service or application behavior changes can reduce manual processing by operational analysts, minimizing the time they need to spend reviewing dashboards or running searches to spot a drift in behavior.

Identify previously unknown scenarios: Creating baselines for service or application behavior can help uncover previously unknown service performance issues.

Case studies and further information

IG Group

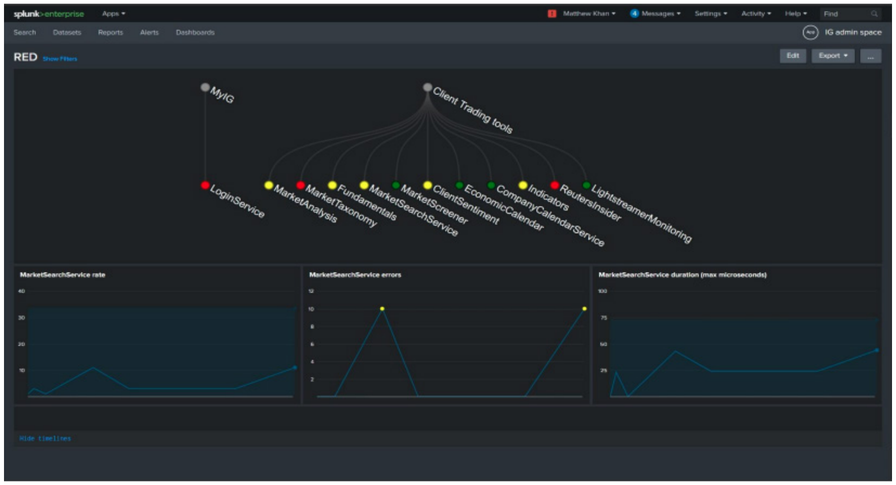
The IG Group is a world leader in online trading, with access to over 17,000 markets and with over 200,000 active clients. To enable their online trading, they use over 1,500 web applications. Understanding the performance of these web applications to ensure minimal downtime or service disruption is important for maintaining their reputation as a world leader as well as to reduce any financial damage.

Historically, IG Group monitored these web applications using static thresholds, focussing mainly on three key RED metrics: rate, error rate, and duration. This method required manual assessment to determine the appropriate thresholds for these metrics, creating risks where setting thresholds too high would lead to missed outages or where setting thresholds too low would lead to alert fatigue from too much noise. In addition, the RED metrics would vary depending on the time of day or day of the week depending on the application being monitored, meaning static thresholds were not fit for round-the-clock monitoring.

By using Splunk's MLTK, IG Group was able to set up dynamic thresholds for its 1,500 web applications, becoming more effective at spotting potential outages. This took trial and error from IG Group to assess the best ML algorithm for providing scale and accuracy, a process supported by partnering with Splunk and the assistive workflows in MLTK.

IG Group landed on a system of centralizing and normalizing the application RED metrics, doing some basic outlier detection using percentiles before training a model that describes the baseline behavior of their applications. Applying that model to the normalized data as it is coming into Splunk allowed IG Group to build a set of alerts and reports for quickly identifying and triaging the anomalous applications, providing better accuracy and ultimately better service performance of their web apps.

Read more about IG Groups' use case here.



StubHub

Another customer who has deployed a machine learning solution successfully for identifying unusual error rates is StubHub, a world leading online ticket marketplace for live events. Their application stack is incredibly complex, with many thousands of servers supporting distributed microservices with over 1000 endpoints. Troubleshooting this environment is incredibly complex, and they developed an 'exception sniffer' using Splunk's MLTK to identify points in time when their stack was experiencing unusual volumes of errors.

By normalizing their error data into a metrics index, StubHub created an optimized index for searching across their application error volumes. From this index, they developed a set of ML models to baseline the expected number of errors, using this model to identify in real-time any unusual volumes of errors. This reduced their exception logging by 50% and also uncovered several previously unknown application issues.

Read more about StubHub's use of machine learning here.

TransUnion

TransUnion is a consumer credit reporting agency, serving millions of consumers across the globe. In order to serve their customers, they run a number of big data platforms, with tens of thousands of data sources and multiple petabytes of data. The ability to understand how their environment is performing and avoid IT incidents is challenging in the face of such large volumes of data.

TransUnion partnered with Splunk to implement a solution for predicting potential outages in Splunk ITSI. After working originally just with MLTK, they then delivered a solution that works with Splunk ITSI, which got incorporated in the product itself as ITSI Predictive Analytics!

Read more about TransUnion's experience with MLTK and Splunk ITSI here.

Further information

Online travel agency Priceline used ML to automatically detect application issues, reducing the manual triage in their NOC, with more information available here. Additionally, global service provider Accenture incorporated ML into their insights as a service platform for identifying issues with services, which can be read about here.

Additionally, there is an MLTK deep dive designed to help users develop analytics for identifying unusual volumes of error rates here. Another deep dive related to finding outliers in server response times can be found here. Provided all the correct apps are installed (namely MLTK, Python for Scientific Computing plus any technical add-ons required to structure your application data) and relevant data is in a Splunk instance, these how-to guides should take a few hours each to implement and test.

3. User experience monitoring

Business challenge

Investment in user experience can produce impressive returns; Forrester estimates that every $1 invested on improving user experience can realize a $100 return.[5] Identifying the elements of the user journey that need the most improvement can be difficult with digital experiences often enabled by complex application stacks. The ability to baseline and understand these application stacks and their behaviors can provide valuable insights into the areas of the stack that could benefit from performance improvements.

Splunk's approach

Splunk's MLTK can be used to better understand user experience, for example, by identifying outliers in transaction times or by clustering clients based on their attributes to identify unusual patterns.

Value

Improve detection efficiency: Using ML to baseline user experiences can help identify areas of application stacks that would benefit the most from improvement.

Identify previously unknown scenarios: Furthermore, the ability to identify bottlenecks and issues in user experiences can improve the customer experience by identifying and responding to poor user experiences before customers report issues.

[5] https://www.forrester.com/report/The-Six-Steps-For-Justifying-Better-UX/RES117708

Case studies and further information

Paychex

Paychex is a provider of human capital management solutions serving hundreds of thousands of customers across the globe, processing the payments of one in twelve American private sector employees in 2019. Paychex was keen to ensure that the user experience provided to their customers is as seamless as possible, and used MLTK to gain insight into areas of their application stack that could benefit the most from improvement.

Paychex used Splunk as the standard platform for correlating and searching across transaction data, and thus decided to use Splunk as the platform for running more advanced analytics on this type of data. By stitching together their transaction data in Splunk, Paychex was able to gain visibility into transaction durations from the UI all the way down to individual SQL queries. From these transaction durations, they used MLTK to help generate predictive models to determine which user demographics had the most impact on transaction times, detect outliers in transaction times, and identify unique clients.

Read more about the user experience monitoring use cases at Paychex here.

BankID

BankID provides secure online identification and signing for financial transactions in Norway. Serving the majority of the population and running millions of transactions a day, BankID needed a way to aggregate and gain insight into billing transactions, particularly for newly introduced services.

Deciding that Splunk was the platform to help them deliver this insight, BankID used MLTK to identify unusually long transaction times in their customers' journeys, proactively identifying potential issues. To enable this analytic, they first used Splunk to generate a set of enriched transaction events, correlating across a number of data sources by transaction ID to retain a number of key fields, which were then enriched with contextual data from a lookup before being pushed into a summary index. From this summary index, BankID could easily visualize the steps in their transactions and automate the identification of outliers using MLTK.

Find out more about BankID's experience with MLTK here.

4. Noise reduction

Business challenge

With a quarter of organizations reporting understaffed IT departments,[6] ensuring that the workloads assigned to those teams are manageable is important. Monitoring IT systems can be a highly manual task, with some services often generating high volumes of alerts that can be overwhelming for analysts to triage, leading to missed signals amongst the noisy alerts. The ability to reduce the volumes of alerts and prioritize the most critical events can help reduce the burden on already overworked IT staff.

Splunk's approach

Reducing noisy alerts can happen at many states of the event analytics pipeline, from suppressing alert generation by using smarter analytics (as described in many situations in the Detecting Service Performance Issues section) through to aggregating alerts into meaningful groups. With Splunk ITSI, customers can enable Smart Mode, which will automatically identify patterns in alerts to generate groupings, creating episodes of alerts that can be triaged as a single incident rather than going from alert to alert.

Value

Reduce manual processing: Reducing the number of alerts that an analyst has to triage can dramatically improve efficiency. Furthermore, IT operations teams can spot true incidents more easily when the risk of missing an incident is reduced by minimizing the human error associated with sifting through massive volumes of alerts.

Case studies and further information

Further information

Read about how Smart Mode works in Splunk ITSI here.

[6] https://itic-corp.com/security-data-breaches-top-cause-of-downtime-in-2022/

5. Predicting data downtime in Splunk

Business challenge

Maintaining operational resilience becomes challenging when data feeds into monitoring tools such as Splunk are interrupted, reducing visibility into how services are performing. Data source owners do not generally grant full access to Splunk administrators, so identifying interrupted data feeds into Splunk is problematic. Furthermore, understanding how event feeds operate can be difficult when many thousands of hosts and many hundreds of data source types are being monitored, with each host and source type combination potentially having different behavior patterns.

Splunk's approach

Splunk can collect data from most systems using forwarders, database connectors, or by using Data Manager (DM), for example. Once data has been collected for a period of time, ML models can be created to describe the expected number of events for a given host and source combination. The models continuously monitor the data feeds and detect anomalies when data streams start to deviate from the expected throughput.

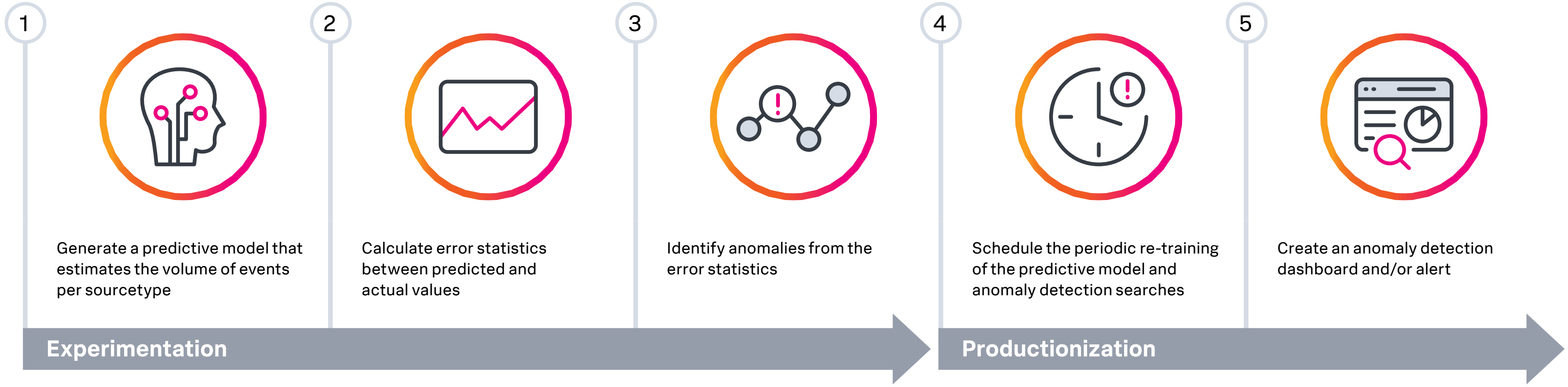
Value

Improve detection efficiency: By finding and addressing abnormalities during data ingestion (before an ingest pipeline gets disrupted or broken), data uptime can be maintained in Splunk. As Splunk is used for monitoring critical systems, maintaining data uptime will provide continued visibility of service issues or service degradation, helping to maintain operational resilience.

Case studies and further information

Further information

There are a number of ways that anomaly detection can be used to monitor data input feeds into Splunk, but for a prescriptive approach to setting up these monitoring capabilities, Splunk has an in-depth webinar that walks through how to implement data feed anomaly detection here. Furthermore, Splunk also has a step-by-step deep dive on how this can be implemented, including sample xml for creating an anomaly detection dashboard here. This deep dive should take a few hours to implement from end-to-end.



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

Productionization

6. Predictive maintenance

Business challenge

The cost of downtime in manufacturing can be high, with some estimates putting the cost of downtime at over $250,000 an hour[7] with additional costs associated with the cost of poor quality outputs. Therefore, the ability to predict when to run maintenance on critical production equipment before it breaks can avoid downtime and avoid producing bad quality products. Generating baseline behavior for manufacturing systems can be difficult due to equipment lifespans of over 10 years, or equipment being complex, and often bespoke.

Splunk's approach

Splunk's MLTK allows users to develop predictive models on OT data through a series of guided workflows or directly using search. Additionally, OT metrics can be baselined using anomaly detection algorithms to help understand what normal looks like.

DSDL can also be used to develop more sophisticated analytics on top of OT data, taking advantage of cutting edge ML libraries such as TensorFlow or PyTorch to better identify patterns in complex OT data feeds.

Value

Improve detection efficiency: Predictive analytics can help reduce unplanned maintenance activities by prompting engineers to fine tune systems before outages occur. Moreover, using predictive analytics can contribute to increased availability of production systems and reduced operating costs. If using Splunk for predictive maintenance, much of the OT data collected can help improve security posture as well.

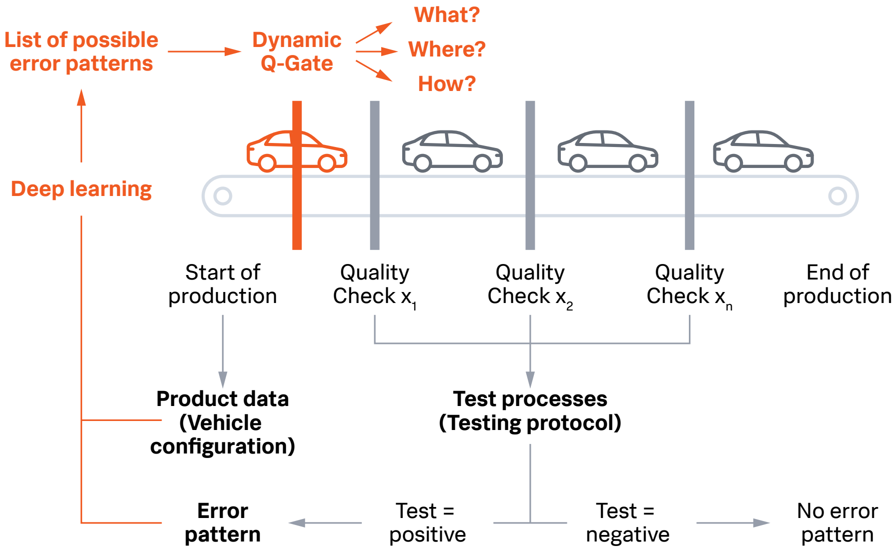
Additionally, predictive analytics on the products that are being manufactured can improve the quality of the output, reducing manufacturing costs and potential recall costs.

[7] https://www.iiot-world.com/predictive-analytics/predictive-maintenance/the-actual-cost-of-downtime-in-the-manufacturing-industry/

Case studies and further information

BMW

BMW already applied a high degree of planned and reactive testing to vehicles during the production process, but by building a predictive testing model using DSDL, BMW has been able to predict potential issues that might have been missed by a human operator.



BMW manufacturing engineers ensure high production quality using insights from Splunk, which provides a holistic view of vehicle data during production. Using DSDL, the BMW Innovation Lab team identified benchmarks for different vehicle types and configurations using Uniform Manifold Approximation and Projection (UMAP), helping BMW generate baselines for expected behaviors and error patterns. Using the dataset of expected baselines and error patterns, BMW trained a dense neural network to predict potential issues with test results, flagging vehicles that might require additional checks. Using these analytics, engineers on the factory floor can troubleshoot vehicles during production.

Read about BMW's experience with the Splunk App for Data Science and Deep Learning here. In addition, listen to the BMW team and Splunk discuss their use of ML here.

Volkswagen

Volkswagen was keen to gain better visibility into their manufacturing processes to improve their security posture. The data required for gaining better visibility also led to work on use cases that would reduce unplanned maintenance.

With hundreds of meters of production lines within a single plant and dozens of production cells performing different tasks, Volkswagen implemented a passive collection architecture to provide visibility into their production processes. From this architecture, Volkswagen can obtain an overview of metrics and status messages for each production cell, such as Programmable Logic Controller (PLC) changes, protocol and data volumes, and logical communications. The data allowed Volkswagen to generate baseline behaviors for normal production cycles, as well as the ability to flag deviations from historic baselines.

Read more about Volkswagen's use of ML here.

Honda
Honda has hundreds of manufacturing plants worldwide, and their plant in Alabama runs on a just-in-time delivery model. Therefore, any downtime at this plant can impact the downstream supply chain and ultimately customers. A critical component of the factory is the regenerative thermal oxidiser used to remove environmentally damaging volatile organic compounds during the production process. If the oxidiser goes down, the plant has to halt production.

Using Splunk's MLTK, Honda was able to create a predictive model for potential failures with the oxidiser, with engineers having taken preventative action based on anomalies observed with the motor.

Some of their detailed use cases included predicting chamber temperatures 24 hours in advance using the StateSpaceForecast algorithm, using DensityFunction to identify outliers in motor readings, and creating a glass table to present all of the predictive indicators to equipment engineers.

Read about Honda's experience here.

Further information
Stimson Lumber Co. used Splunk and MLTK to autonomously detect outliers and assess their impact on production processes, enabling maintenance teams to react faster and more accurately to potential problems. Read about Stimson Lumber Co.'s experience with MLTK here.

The Battelle Memorial Institute is one of the partners who work at the Pueblo Chemical Destruction Agent Pilot Plant, a facility designed to destroy chemical weapons. The Battelle Memorial Institute used ML to identify the impacts that readings were having on other metrics, helping to inform maintenance of certain systems. Read more about the work at the Pueblo Chemical Destruction Agent Pilot Plant here.

ESE GmbH is a services provider operating out of Germany specializing in rail operations, automotive production, and the manufacturing industry. While working for a customer, ESE GmbH was engaged to help with the operations of a sugar processing plant and used MLTK to uncover insights that supported increasing sugar output, increasing the runs per day, and avoiding machine failures. Find out more about ESE GmbH's work with MLTK here.

7. Cell tower monitoring

Business challenge

Communication service providers (CSPs) provide a world class experience to customers by avoiding issues like traffic congestion. Therefore, being able to identify and prevent potential issues with network infrastructure is a critical part of providing a great customer experience. With most CSPs running a huge number of connected devices and network components across a broad geographic area, understanding and baselining normal traffic patterns can be a challenge to effective monitoring.

Splunk's approach

With Splunk's MLTK, users can train and apply ML models that can detect anomalies in network traffic data, forecast or predict traffic patterns, or cluster entities into cohorts. These ML techniques can be applied via guided workflows or directly as a search.

Value

Improve detection efficiency: ML offers numerous end user benefits for monitoring cell tower data, such as reduction in cell tower congestion, improvements in data speeds for subscribers, reduction in customer complaints, and ultimately a better user experience.

Reduce manual processing: Furthermore, by automating the detection of network issues, engineering teams can gain efficiency, and network resources can be utilized more efficiently.

Case studies and further information

T-Mobile
T-Mobile has been partnering with Splunk for many years to apply ML to their cell tower network data.

By using ML across a range of different network use cases, T-Mobile predicts potential outages, optimizes 5G networks for better throughput, and uses anomaly detection to help remediate potential messaging issues.

T-Mobile began by analyzing cell tower congestion in 2019. Using the StateSpaceForecast algorithm to forecast congestion, T-Mobile was able to predict points in time when there would be congestion. In addition, T-Mobile implemented a dashboard for operators to mark predicted congestion as actual or not, helping continuously improve their detections.

Read more about this use case from T-Mobile here.

Operating America's largest 5G network, T-Mobile's radio frequency engineers often have to balance capacity and coverage to ensure the best user experience for their customers. With thousands of cell towers and hundreds of configurable network parameters, this can be a challenging task, often requiring multiple manual iterations by engineers in order to achieve the best results. Using configuration and capacity data in Splunk, the team at T-Mobile developed an anomaly detection model using the DensityFunction algorithm to identify cell towers that require additional configuration, reducing the time required by engineers to manually tune the network. In one particular location in New York, tuning based on the anomaly detection model improved data speeds by over 80%.

Read more about this use case here.

Messaging services are extremely complicated, often traversing multiple nodes on the network, which are underpinned by many more services. With thousands of metrics generated by messaging services, understanding what a healthy system looks like can be difficult. Through use of an anomaly detection system, T-Mobile was able to reduce P1 and P2 incidents, reduce customer reported issues, and reduce the mean time to detect potential network issues.

Read about this use case from T-Mobile here.

8. Geohazards monitoring

Spotlight case study

Norwegian Water Resources and Energy Directorate (NVE)
The final use case is observability for the physical world–how NVE uses Splunk to support their early warning system for potential rockslides.

Rockslides in Norway can have devastating consequences, with whole villages washed away by tsunamis caused by sudden rockslides. Having an early warning system in place to warn citizens of potential rockslides is particularly important in remote parts of northern Norway, where evacuations often can take more planning than in more populated regions.

Taking data from dozens of different types of sensor devices, NVE uses Splunk to provide visibility into environmental situations, with many sensors used to provide information about rock degradation. With sensors placed in hazardous locations, often exposed to extreme weather conditions, getting accurate readings from the sensors can be difficult.

Using Splunk's MLTK, NVE in partnership with Sopra Steria are able to remove outliers from their sensor readings, helping to provide accurate views on the risk of potential rockslides. By using ML to improve the data quality, NVE can improve decision making.

Hear about NVE's use of ML here.

Get started today

Download MLTK and DSDL to get started with AI/ML! (But first you need Splunk Cloud Platform or Splunk Enterprise! If you don't have one of these yet, you can try a free trial.)

Check out our MLTK Deep Dives for detailed implementation guides for some popular uses of AI in Splunk, including a video overview and detailed walkthrough.

Explore solutions like IT Service Intelligence or Splunk Infrastructure Monitoring for out-of-the-box AI use cases.

Review Splunk Blogs to find out more about given techniques and approaches to AI in Splunk (check out some of these in the table below).

For current users, your Splunk account team will also be able to help you explore support that is available from Splunk for implementing ML use cases.

Explore more resources

Anomaly Detection	Predictive Analytics	Clustering
Cyclical Statistical Forecasts and Anomalies part 1 Cyclical Statistical Forecasts and Anomalies part 4 Cyclical Statistical Forecasts and Anomalies part 5 Cyclical Statistical Forecasts and Anomalies part 6 A Splunk Approach to Baselines, Statistics and Likelihoods on Big Data Anomalies Are Like a Gallon of Neapolitan Ice Cream part 1 Anomalies Are Like a Gallon of Neapolitan Ice Cream part 2	Cyclical Statistical Forecasts and Anomalies part 2 Cyclical Statistical Forecasts and Anomalies part 3 Anomalies Are Like a Gallon of Neapolitan Ice Cream part 1 Anomalies Are Like a Gallon of Neapolitan Ice Cream part 2 ITSI and Sophisticated Machine Learning Predicting Resource Exhaustion with Double Exponential Smoothing	Anomalies Are Like a Gallon of Neapolitan Ice Cream part 2

23-253850-Splunk-Splunk Artificial Intelligence for Observability-EB-106