See the Problem Before It Becomes a Problem: Predict and Prevent the Problem with Splunk IT Service Intelligence (ITSI)

Bill Babilon | Splunk Global Solutions Architect

October 2018 | Version 1.0

Forward-Looking Statements

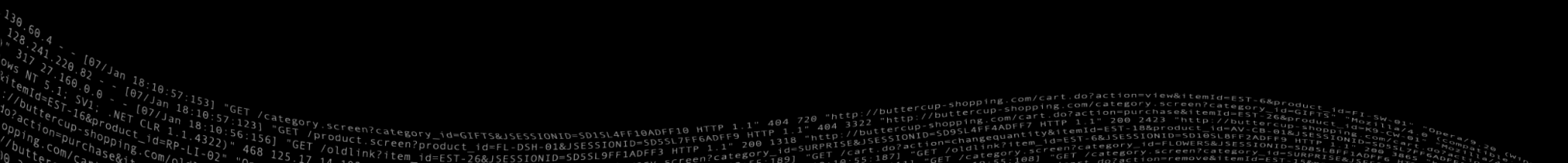
During the course of this presentation, we may make forward-looking statements regarding future events or the expected performance of the company. We caution you that such statements reflect our current expectations and estimates based on factors currently known to us and that actual events or results could differ materially. For important factors that may cause actual results to differ from those contained in our forward-looking statements, please review our filings with the SEC.

The forward-looking statements made in this presentation are being made as of the time and date of its live presentation. If reviewed after its live presentation, this presentation may not contain current or accurate information. We do not assume any obligation to update any forward-looking statements we may make. In addition, any information about our roadmap outlines our general product direction and is subject to change at any time without notice. It is for informational purposes only and shall not be incorporated into any contract or other commitment. Splunk undertakes no obligation either to develop the features or functionality described or to include any such feature or functionality in a future release.

splunk> .conf18

Agenda

- Speaker Introduction
- Role of IT Today
- What is ML
- ML in ITSI
- Case Study 1 – SI's Financial Application
- Case Study 2 – Gov't Agency, Security Infrastructure
- What We Learned



splunk> .conf18

BILL BABILON

Global ITOA Solutions Architect, Splunk

splunk> .conf18

Problem Statement

How did we get started?

What is the Role of IT Today?

What is the Goal of IT Today?

“Negative MTTR”

Giving You a ‘Heads Up’ to an Operational Impacting Issue BEFORE it Occurs

splunk> .conf18

Alerts

Finding the Known Knowns

- **Alerts,** by their very nature, are **reactive** – we have already crossed a threshold
- **Alerts** are for ‘known’ problems that have happened in the past
- **Finding the ‘known knowns’**
  - Page response time is too long!
  - CPU on Server 3 is too high!
  - Low disk space on storage array 3!

Metrics

Finding the Unknowns

- **Metrics** are for finding issues BEFORE the become a problem
- **Finding the 'Unknowns'**
  - Server3 CPU=99%
  - SAN Array fill ration=95%
  - MQ3 QueueDepth=25
  - PS UserCount=4682

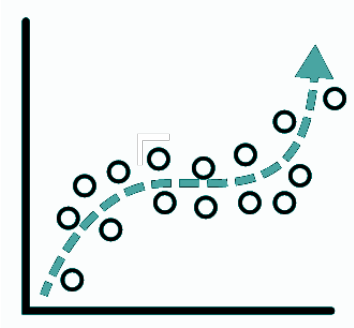
Machine Learning 101

Getting Answers From Your Data!

Anomaly Detection

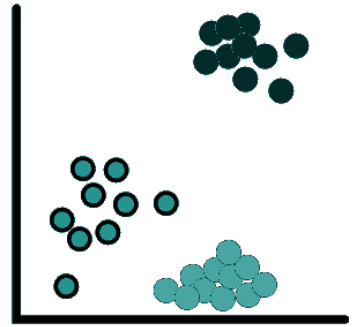
- Deviation from past behavior
- Deviation from peers
- Unusual changes in features
- **ITSI MAD Anomaly Detection**

Predictive Analytics



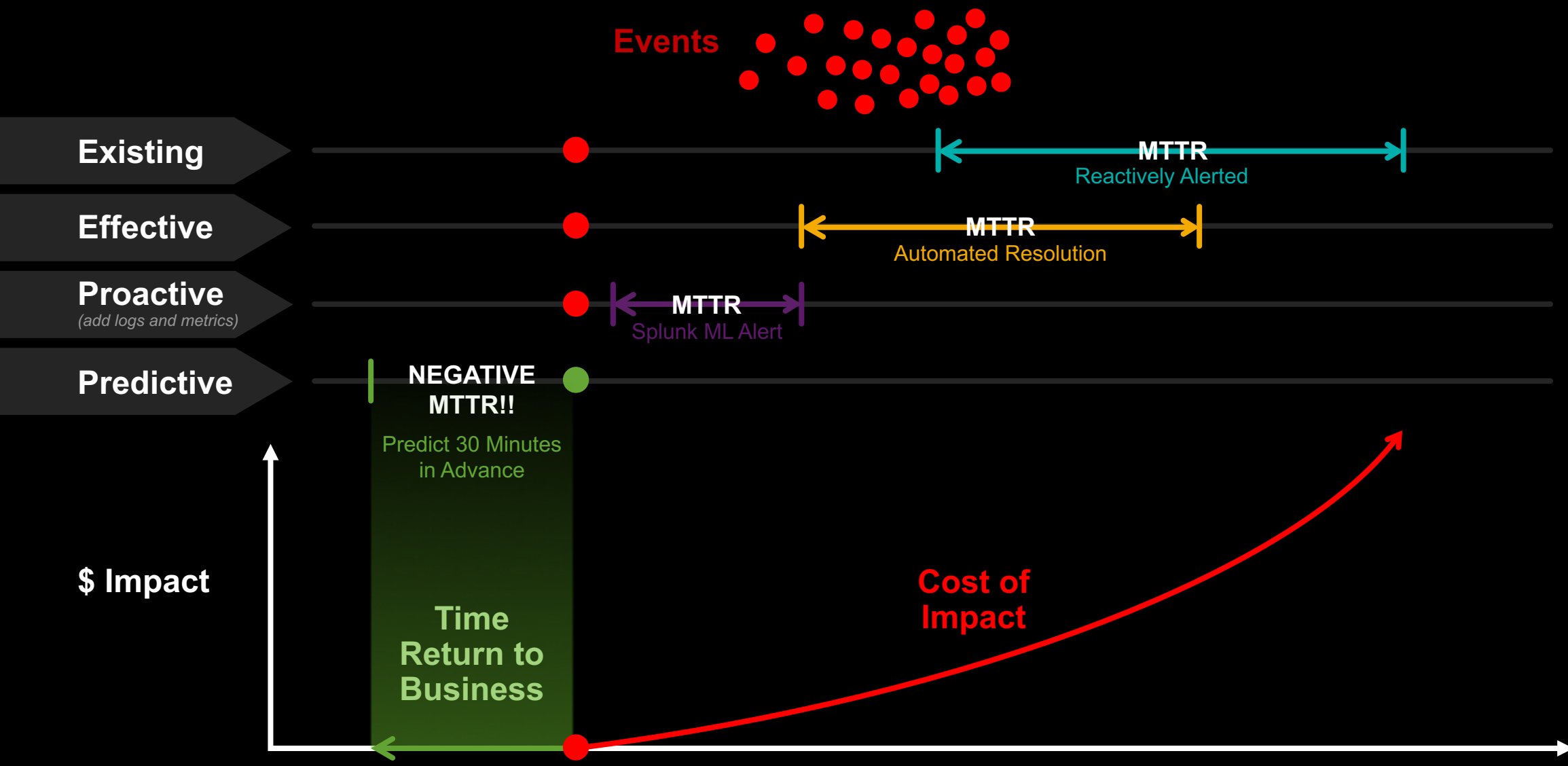
- Predict Service Health Score
- Predicting churn
- Predicting events
- Trend forecasting
- Detecting influencing entities
- Imminent outage prediction
- **ITSI Predictive Analytics**

Clustering



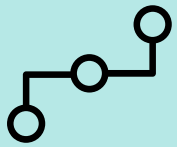
- Identify peer groups
- Event correlation
- Reduce alert noise
- Behavioral analytics
- **ITSI Event Analytics**

Predict and Prevent Operational Issues with AI



Splunk's AI/ML offerings

AIOps



IT Service Intelligence

- **Tailored for IT use cases**
- **Out of the box**

Analytics-driven Security

User Behavior Analytics

- **Tailored for Security use cases**
- **Out of the box**

IT and Security Practitioners

Machine Learning

Machine Learning Toolkit

- Custom ML for any use case
- Requires Data Science and Splunk expertise
- Integration with open source algorithms
- Works inside any splunk search pipeline

Citizen Data Scientist

splunk> .conf18

Splunk IT Service Intelligence™

Artificial Intelligence for IT Operations

Powered by machine learning and analytics for real-time service insights, simplified operations and root-cause isolation

splunk> .conf18

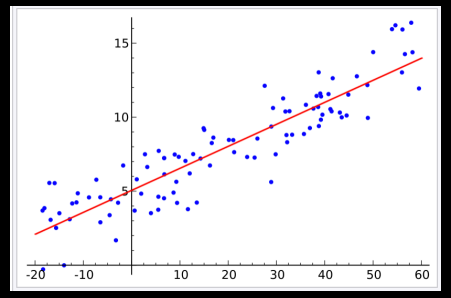
Predictive Analytics in ITSI

The ‘Easy Button’ for ML

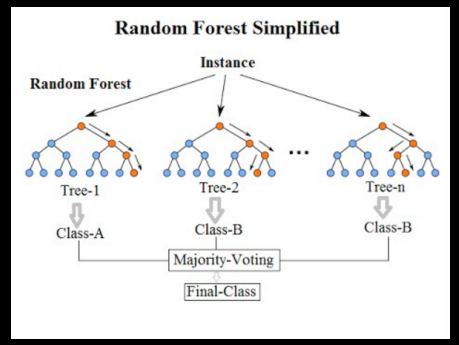
splunk> .conf18

Predictive Analytics Algorithm

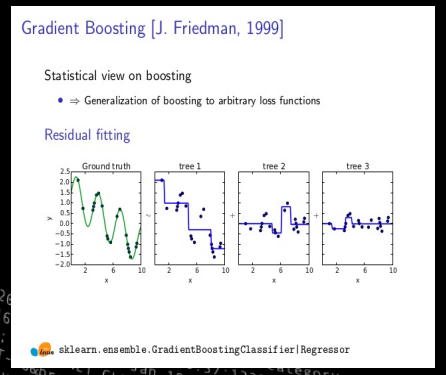
Imminent Outage Prediction Dashboard



The **Linear Regression** algorithm fits a linear line to your data, using each input as an additional dimension. It assumes that your data is normal and is highly scalable



The **Random Forest Regressor** algorithm takes the inputs (KPIs and historical service health scores) and forms a random decision tree (a "forest") to determine the output value. This model makes no assumptions about the normalcy of your data, but requires more processing power and takes longer to run



The **Gradient Boosting Regressor** algorithm uses a loss function to fit a line to your data, a decision tree, and an additive model to predict the service health score value. Think of this as a combination of the Random Forest Regressor and Linear Regression. This algorithm can continuously learn, but in this dashboard it runs only once.

splunk> .conf18

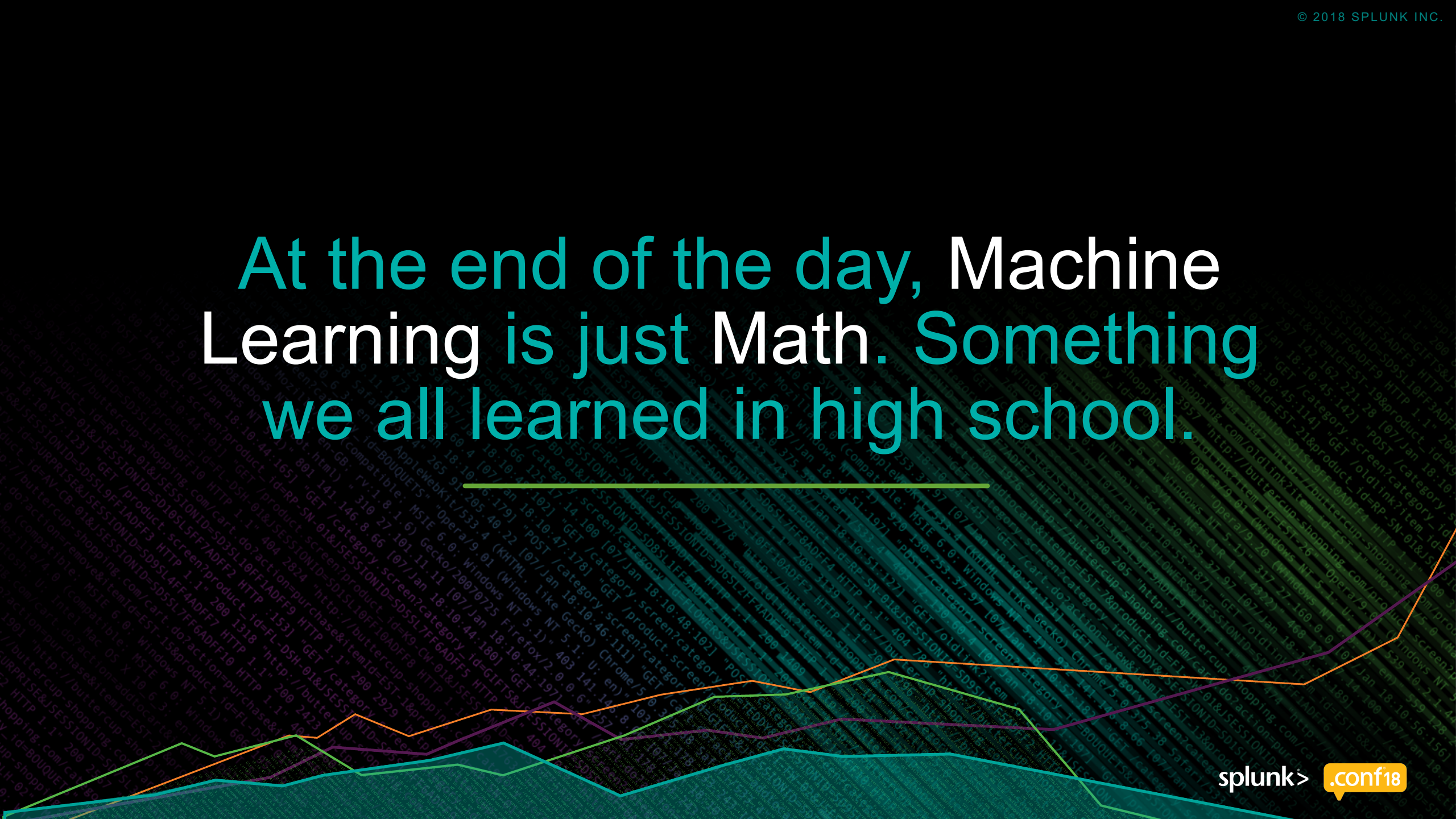
Predictive Analytics Algorithm

Imminent Outage Prediction Dashboard

How does the Algorithm work?

- We use historical KPI data and the service health score at a point in time as input to our model that predicts service health score in 30 minutes
- We train our model to understand the health score 30 minutes later when the combination of input KPI's and current health score is a specific value
- Once the model is trained, the real time KPI values and current health score are used to predict the health score in 30 minutes

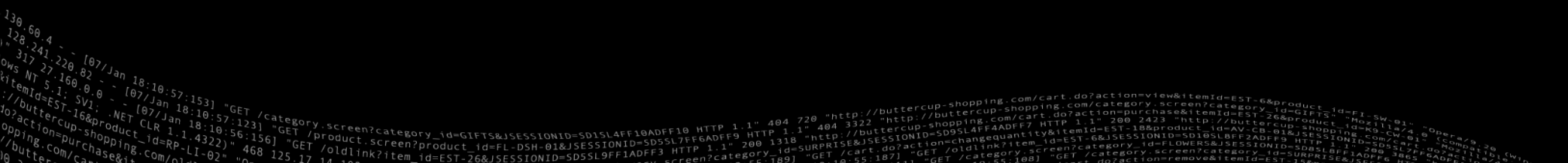
At the end of the day, Machine Learning is just Math. Something we all learned in high school.



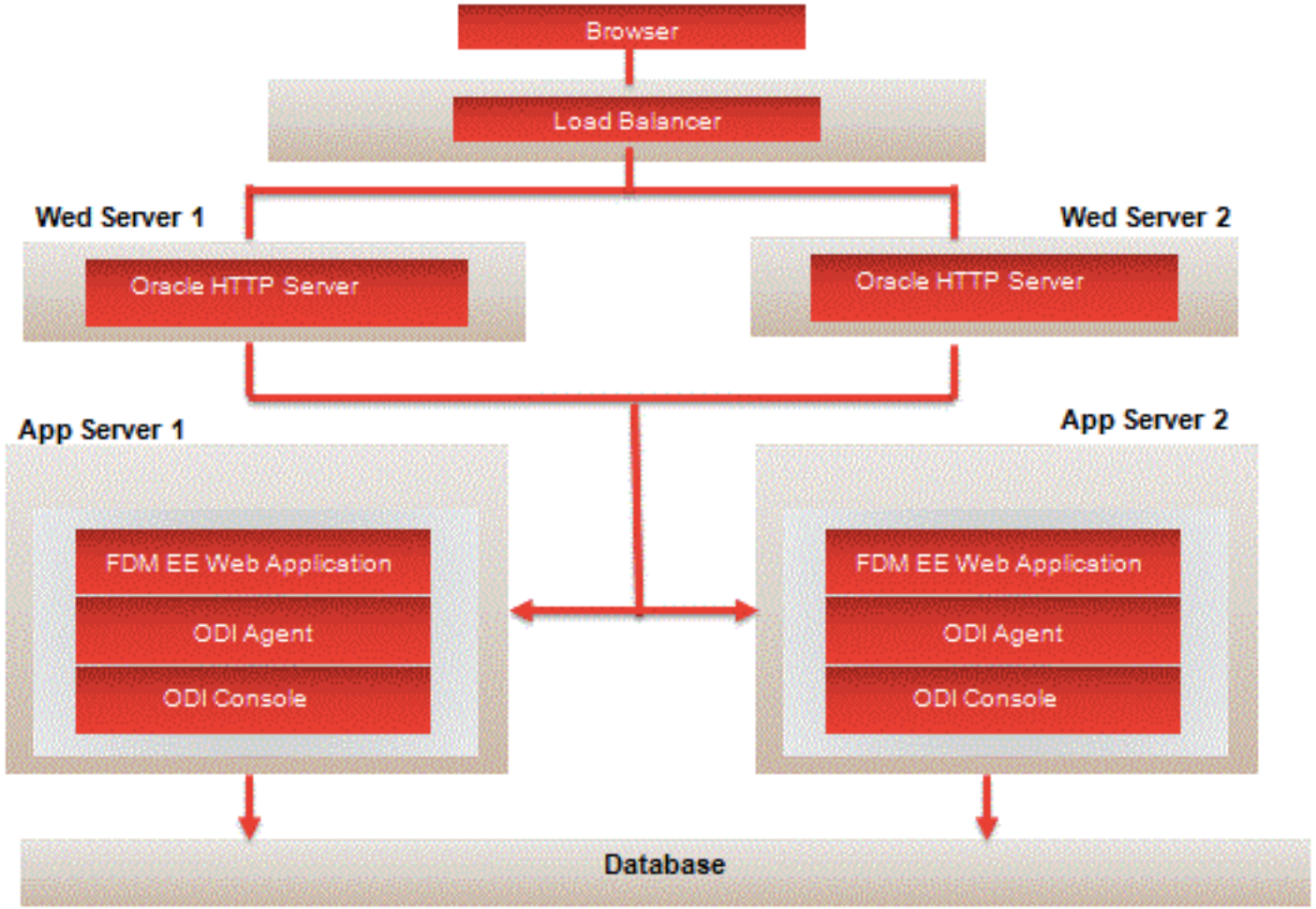
splunk> .conf18

Case Study 1: Financial Planning Application – Government Systems Integrator

- Hyperion
  - Financial Planning and Management
  - Mission critical for month/quarter/year close
  - Handles over $7B in annual billing
  - Tends to ‘work’ but when it doesn’t it has huge impact to the business
  - Has lots of complaints of being ‘slow’

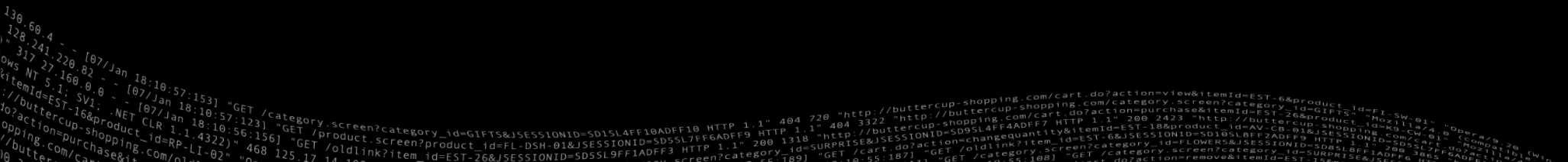


Hyperion Architecture

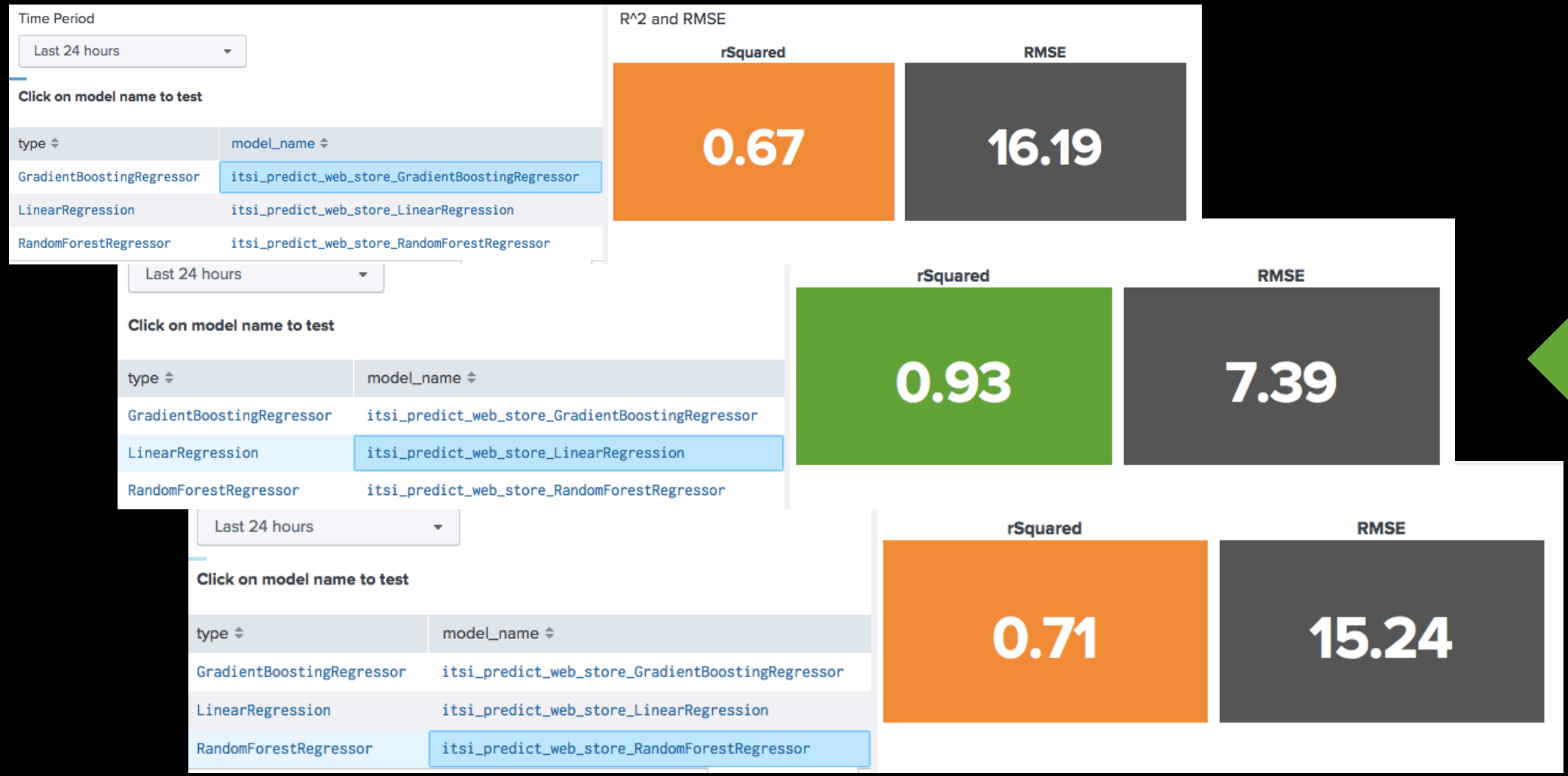


Hyperion Metrics

- OS Metrics – web/app/DB
  - CPU load, memory, network
- Web Tier Specific
  - User sessions
  - Page build time
  - HTTP status (200/500)
- App Tier Specific
  - User sessions
  - JVM ‘Stuck’ Threads
  - JVM Garbage Collection Times
  - HTTP status codes (200/500)
- DB Tier
  - DB threads in use
  - Error message
  - Job runtime
  - Number of jobs



Predictive Analytics on the Service Health Score

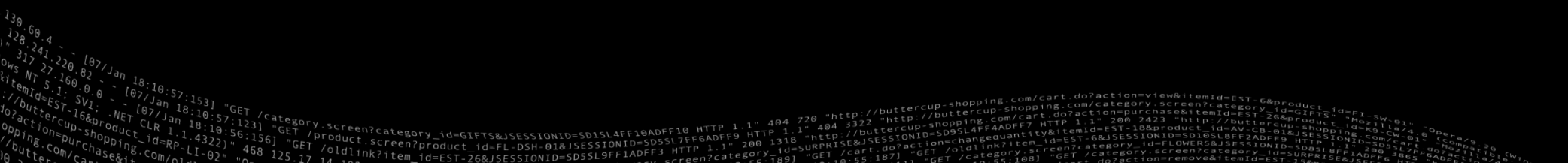


Best Choice

splunk> .conf18

What We Learned

- User Behavior
  - Go around the global load balancer
  - Users submitted most of their ad hoc reports in the evening
- Critical Resources
  - Database worker threads were a key metric to track
- Predictive Analytics
  - Saw the drop in the HS when the DB thread pool saturated
  - Eventually RCA'd to a table lock issue of untimely backups
  - ***When occurring, ITSI gave a 20 minute alert to running out of DB threads***

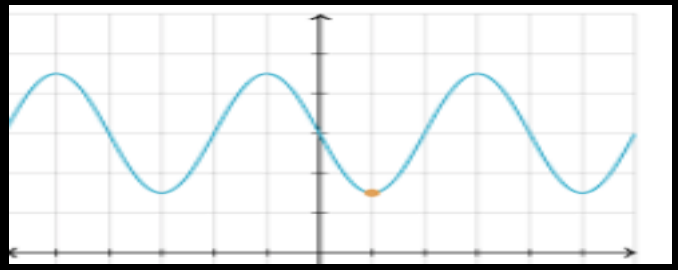


Case Study 2: Security Infrastructure – Government Agency

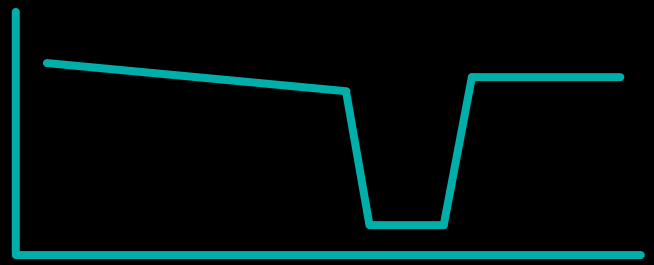
- Vulnerability scans and analysis taking excessive amount of time
  - Multiple scanning tools in play
  - Mixture of COTS and GOTS tools
  - Looking at vulnerabilities as multiple levels – server/workstation/mobile devices
  - Initially believed to be a simple four (4) step process – scan, data aggregation, analysis and reporting
  - It was anything but simple – there was ‘home grown’ orchestration engine handling the data aggregation and analysis

What We Learned…

- Modeled the Security Service in ITSI
- Initially thought the service health score would be sinusoidal in nature



- The Health Score was actually:



- Traditional Linear Regression based models don't work well in this scenario

MLTK to the Rescue

splunk>enterprise App: Splunk Machine Learning Toolkit ▾ | Administrator ▾ | 4 Messages ▾ | Settings ▾ | Activity ▾ | Help ▾ | Find

Showcase | Experiments | Search | Models | Classic ▾ | Docs | Video Tutorials | Splunk Machine Learning Toolkit

Your Splunk Machine Learning Toolkit has been upgraded to 3.4.0. What's new?

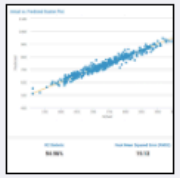
Showcase

Showcase

Welcome to the Showcase, which exhibits some of the analytics enabled by this app. Click on one of the examples to see that Assistant applied to a real dataset. Please see the video tutorials for more information.

Select which examples to show

All Examples

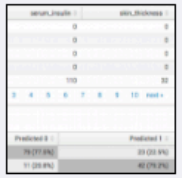


Predict Numeric Fields

Predict the value of a numeric field using a weighted combination of the values of other fields in that event. A common use of these predictions is to identify anomalies: predictions that differ significantly from the actual value may be considered anomalous.

Examples

- Predict Server Power Consumption
- Predict VPN Usage
- Predict Median House Value
- Predict Power Plant Energy Output

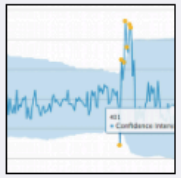


Predict Categorical Fields

Predict the value of a categorical field using the values of other fields in that event. A common use of these predictions is to identify anomalies: predictions that differ significantly from the actual value may be considered anomalous.

Examples

- Predict Hard Drive Failure
- Predict the Presence of Malware
- Predict Telecom Customer Churn
- Predict the Presence of Diabetes
- Predict Vehicle Make and Model



Detect Numeric Outliers

Find values that differ significantly from previous values.

Examples

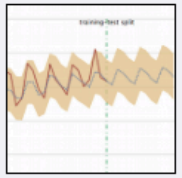
- Detect Outliers in Server Response Time
- Detect Outliers in Number of Logins (vs. Predicted Value)
- Detect Outliers in Supermarket Purchases
- Detect Outliers in Power Plant Humidity

Detect Categorical Outliers

Find events that contain unusual combinations of values.

Examples

- Detect Outliers in Disk Failures
- Detect Outliers in Bitcoin Transactions
- Detect Outliers in Supermarket Purchases
- Detect Outliers in Mortgage Contracts
- Detect Outliers in Diabetes Patient Records
- Detect Outliers in Mobile Phone Activity

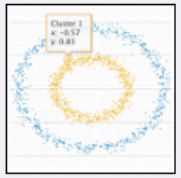


Forecast Time Series

Forecast future values given past values of a metric (numeric time series).

Examples

- Forecast Internet Traffic
- Forecast the Number of Employee Logins
- Forecast Monthly Sales
- Forecast the Number of Bluetooth Devices
- Forecast Exchange Rate TWI using ARIMA



Cluster Numeric Events

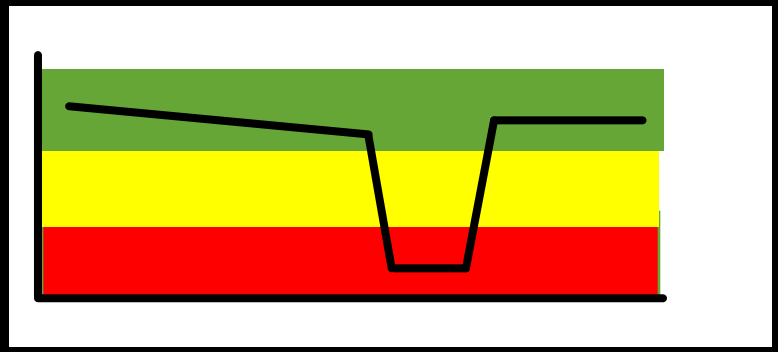
Partition events with multiple numeric fields into clusters.

Examples

- Cluster Hard Drives by SMART Metrics
- Cluster Behavior by App Usage
- Cluster Neighborhoods by Properties
- Cluster Vehicles by Onboard Metrics
- Cluster Power Plant Operating Regimes

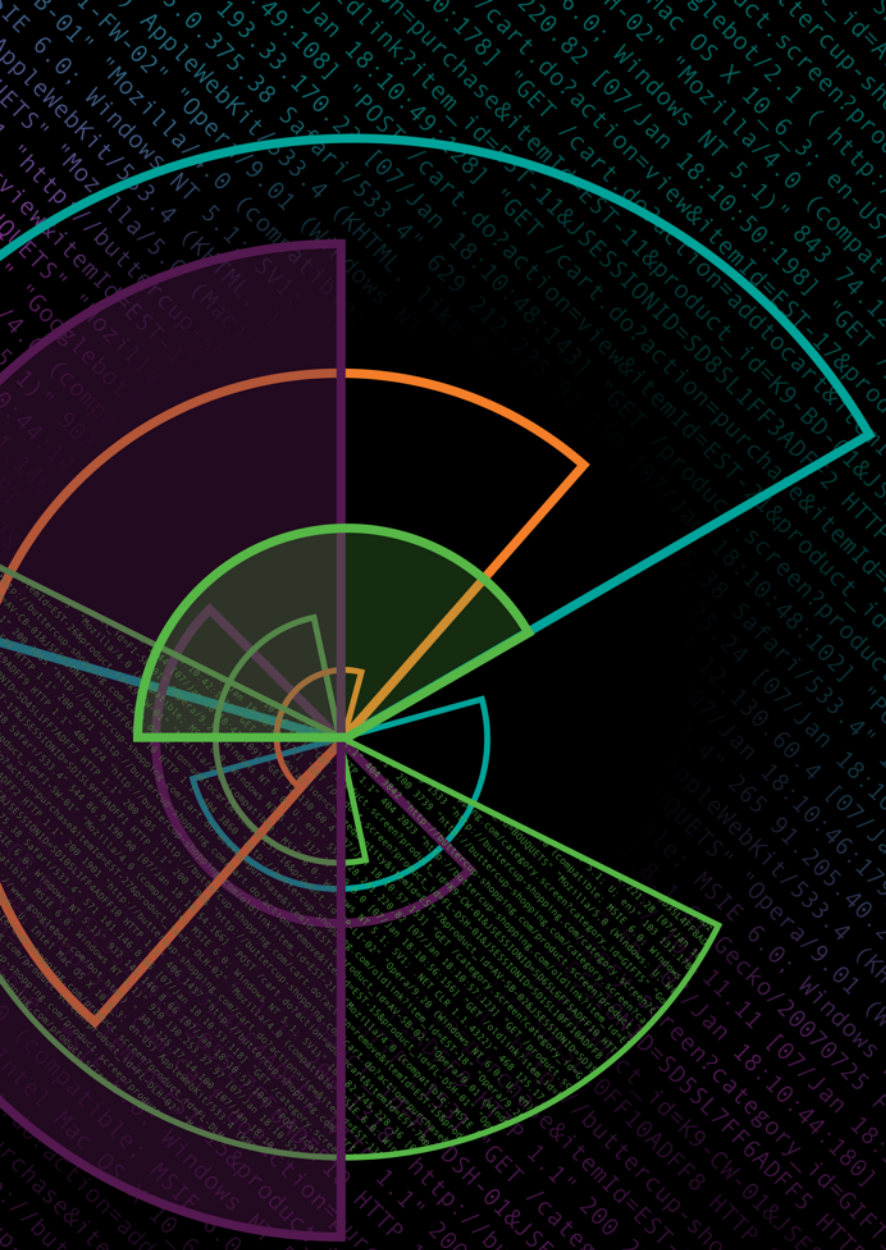
MLTK to the Rescue
Logistical Regression Model

- ITSI and Sophisticated Machine Learning
  - https://www.splunk.com/blog/2017/08/28/itsi-and-sophisticated-machine-learning.html
- Integrates well with ITSI
- Includes over 200 models
- Categorization models work really well
  - Focus on the threshold of the ITSI Health Score – Green/Yellow/Red
- **Was able to give a 90 minute warning on likely failure of aggregation and analysis jobs**



Key Takeaways

1. Machine Learning for Predictive Analytics WORKS for improving IT Operations!!!
2. Machine Learning is NOT limited to PhD's and wanting to 'learning about our data'
3. Not every failure mode of a service can be predicted
4. ITSI remains to be the 'easy button' for ML when it comes to monitoring applications
5. **Being Pro-Active can require a culture change in an IT organization**



Q&A

splunk> .conf18

Thank You

Don't forget to rate this session in the .conf18 mobile app