.conf18

splunk>

TransUnion and a Time Traveling DeLorean

MTTR Fading Like Marty McFly

Steve Koelpin, TransUnion and Splunk Trust MVP
Andrew Stein, Splunk Principal PM for Machine Learning

Oct 2018

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.

Steve Koelpin

Lead Splunk Engineer
Splunk Trust MVP
New Dad
Winner of the Splunk Answers Karma Contest

Andrew Stein

Splunk Principal Product Manager, Machine Learning

- 18 years creating mathematically modeled solutions as a data scientist
- I spend 80% of time preparing data and 20% of time complaining about the need to prepare data

Agenda

- TransUnion and Splunk
- Why Use Machine Learning?
- TransUnion and ITSI
- TransUnion and ITSI + MLTK
  - How It Works
  - Training the Model
  - Applying the Model
- Challenges in Predictive Analytics
- Pro Tips
- Bring This to Your Organization

TransUnion and Splunk

Information for Good

TransUnion and Splunk

Hundreds of daily users

Several core Splunkers

Casual users to certified consultants

TransUnion

TransUnion and Data

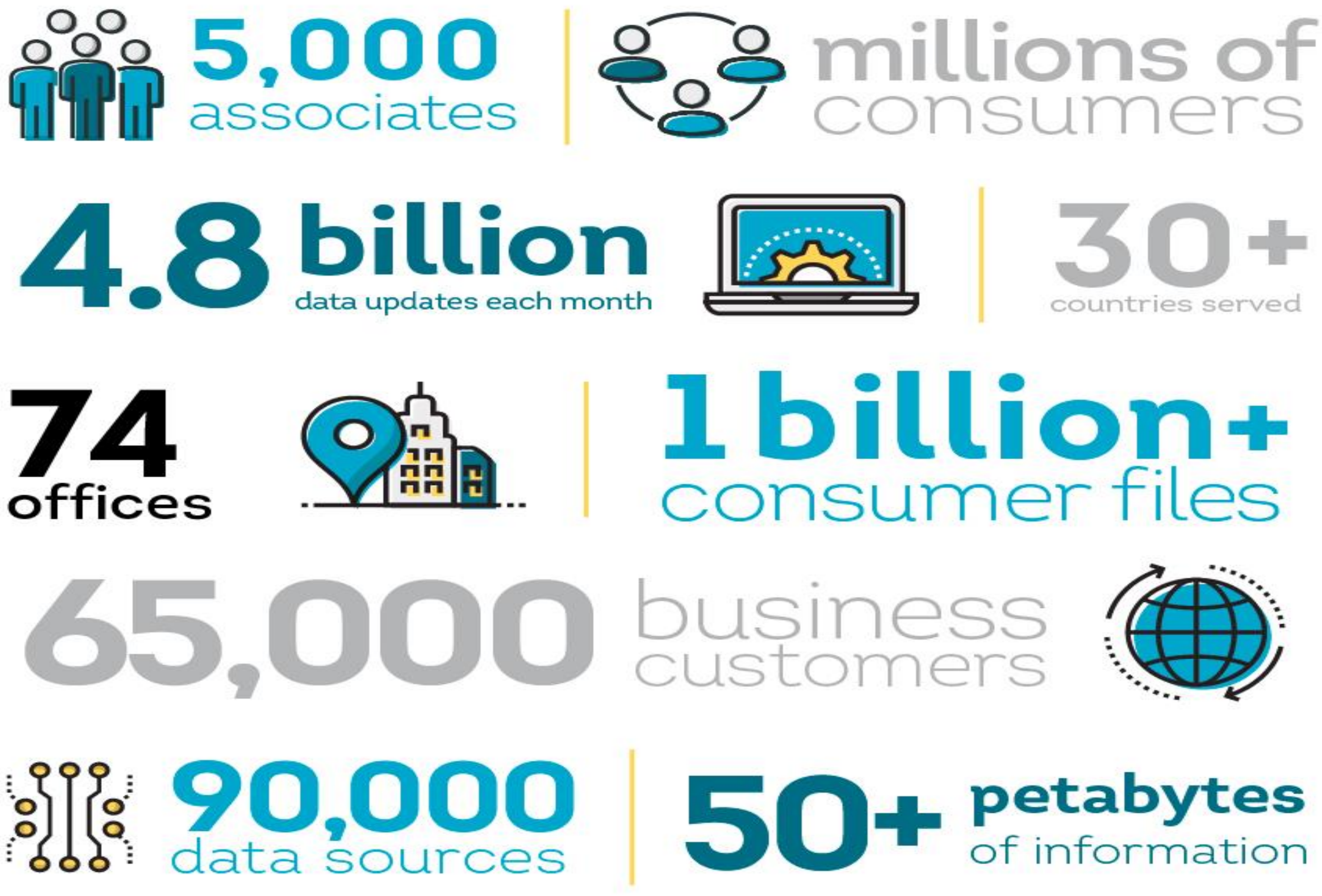
TransUnion is a BIG Data and Information Solutions Company

Founded as a Credit Bureau in 1968

We See Data Differently – Not for What it is – But for What it Can Help People Accomplish

This View – The Individuals for Whom we Steward and Protect Information

We Call this Information For Good



Why Use Machine Learning?

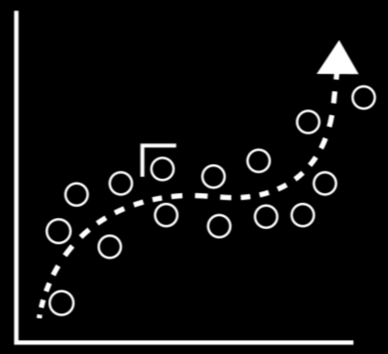
Problems Machine Learning Solves

Getting Answers From Your Data

Anomaly Detection

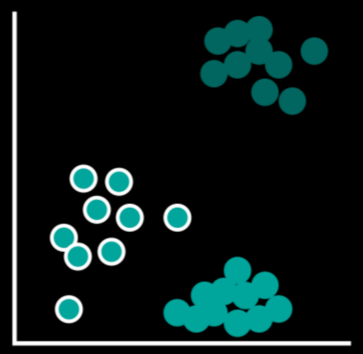
- Deviation from past behavior
- Deviation from peers
- Unusual changes in features
- **ITSI MAD Anomaly Detection**

Predictive Analytics



- Predicting ServiceHealthScore
- 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**

The Cost of an Incident

Line of Revenue

Customer Satisfaction

Brand Reputation

$105,302

= the mean business cost of an IT incident

*According to “Damage Control: The Impact of Critical IT Incidents”

https://www.splunk.com/en_us/form/damage-control-the-impact-of-critical-it-incidents.html

Reduce Your Technical Debt with Machine Learning

Correlate dozens of KPIs against data in the past

No more tribal Knowledge

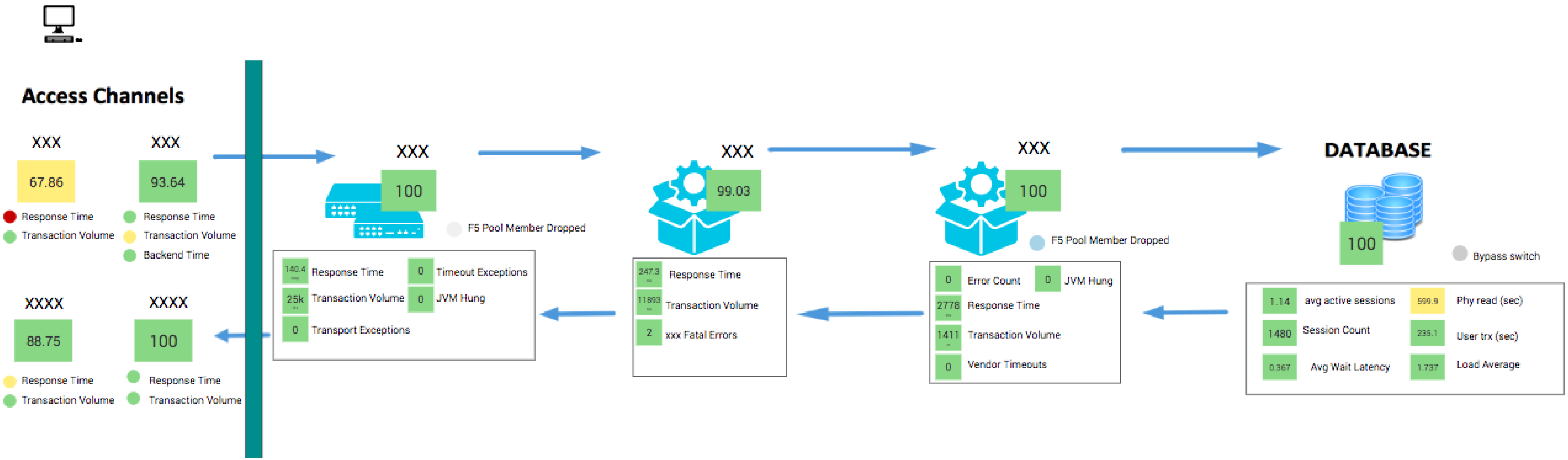
Have machine learning do the leg work

TransUnion and ITSI

IT Service Intelligence

TransUnion and ITSI

Glass Table View of Application Pipeline

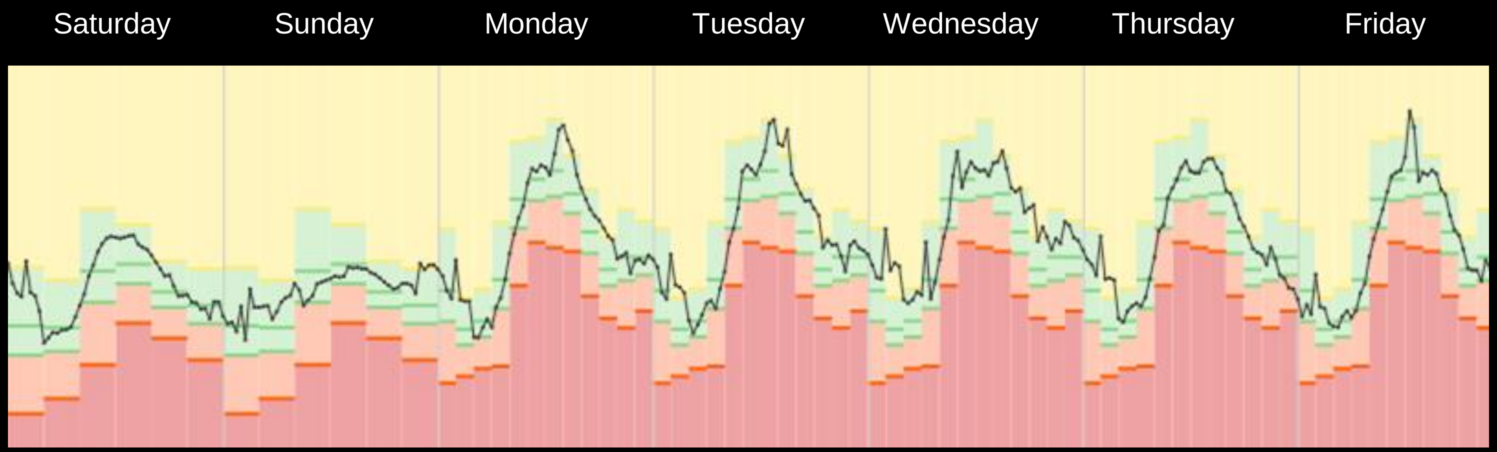


*Updates every 1 minute

What Was the Investment to Build the Solution?

MOST TIME-CONSUMING TASKS

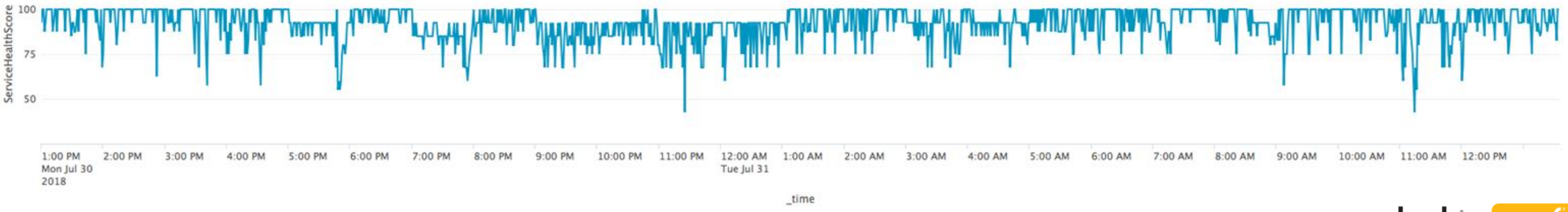
- Understanding effective KPIs
- Getting information from other BUs
- Developing a workflow
- Applying thresholds



How Does ITSI Tie Into Predicting Incidents?

- ITSI gives us the ability to take multiple KPIs and tie them into a single health score
- Apply adaptive thresholding to cyclic-type data patterns
- Faster time to value

```
index=itsi_summary gs_service_id="1ad210ad-329d-4bba-8c31-fc6c878cb608" kpiid="127b54fe58aa59600a64c9d8" OR kpiid="1e75e8ee4a395fc86ced70c3"
    ="b726f6de942dc7a4ce7842eb" OR kpiid="SHKPI-1ad210ad-329d-4bba-8c31-fc6c878cb608"
| eval Errors=if(kpiid="127b54fe58aa59600a64c9d8", 'alert_value',"N/A")
| eval Response_Time=if(kpiid="1e75e8ee4a395fc86ced70c3", 'alert_value',"N/A")
| eval F5_Dropped=if(kpiid="2fa253aaec53c3846492919d", 'alert_value',"N/A")
| eval Volume=if(kpiid="a3c0cc8213e6120c25eca484", 'alert_value',"N/A")
| eval Timeouts=if(kpiid="b726f6de942dc7a4ce7842eb", 'alert_value',"N/A")
| eval ServiceHealthScore=if(kpiid="SHKPI-1ad210ad-329d-4bba-8c31-fc6c878cb608", 'alert_value',"N/A")
| timechart span=1m min(ServiceHealthScore) AS ServiceHealthScore
```

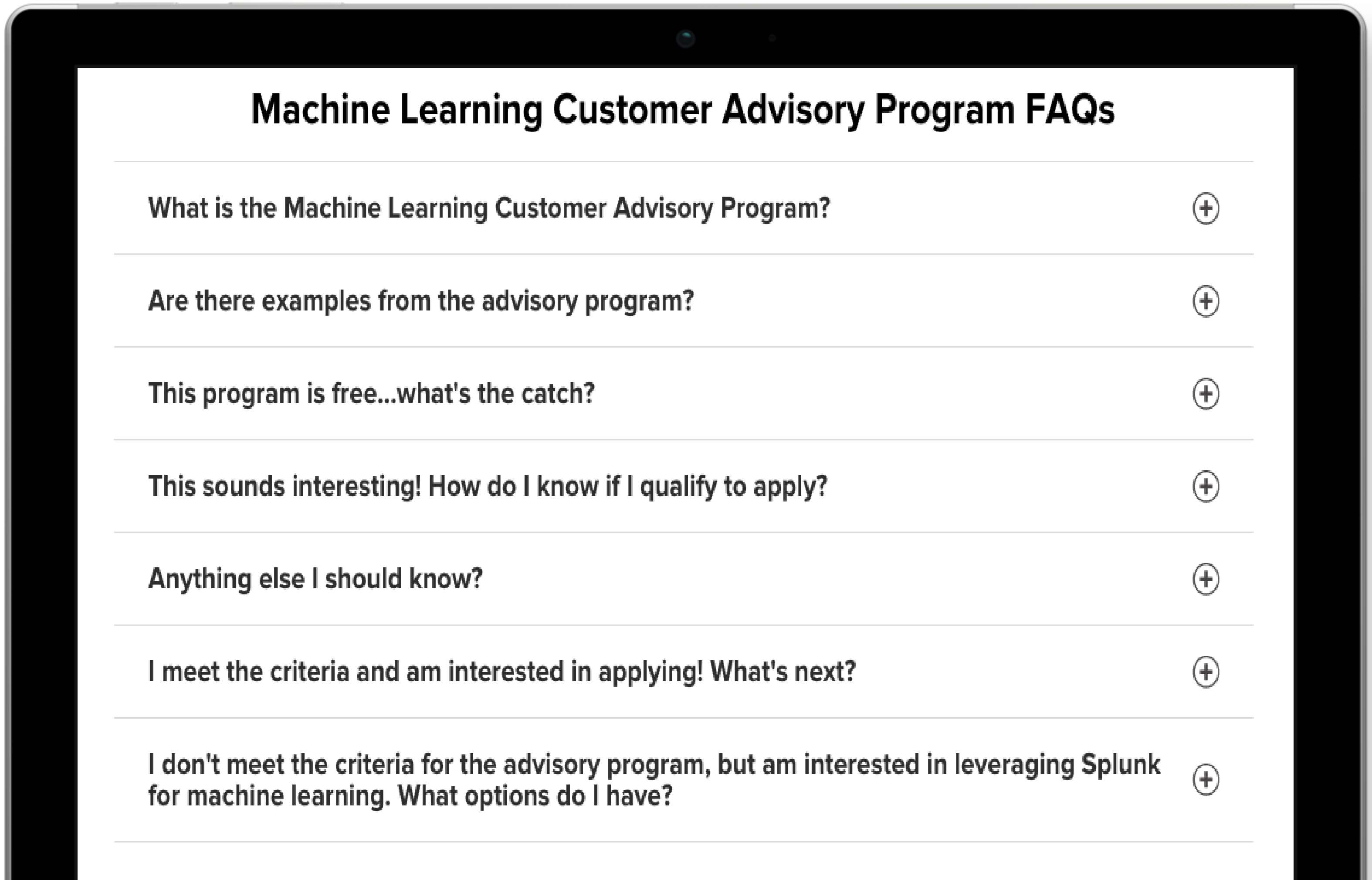


TransUnion and the MLTK

Splunk Advisory Program

What Is the ML Advisory Program?

Provides customers with Splunk data science resources to help operationalize a specific ML use case



Machine Learning Customer Advisory Program FAQs

What is the Machine Learning Customer Advisory Program?

Are there examples from the advisory program?

This program is free...what's the catch?

This sounds interesting! How do I know if I qualify to apply?

Anything else I should know?

I meet the criteria and am interested in applying! What's next?

I don't meet the criteria for the advisory program, but am interested in leveraging Splunk for machine learning. What options do I have?

- Early access to new and enhanced MLTK features
- Opportunity to shape the development of the product
- Assistance in operationalizing a production-quality ML model

ML Advisory Customers

BMW

NTT docomo

HYATT

RECURSION pharmaceuticals

TELUS

TransUnion

UNLV

Zillow

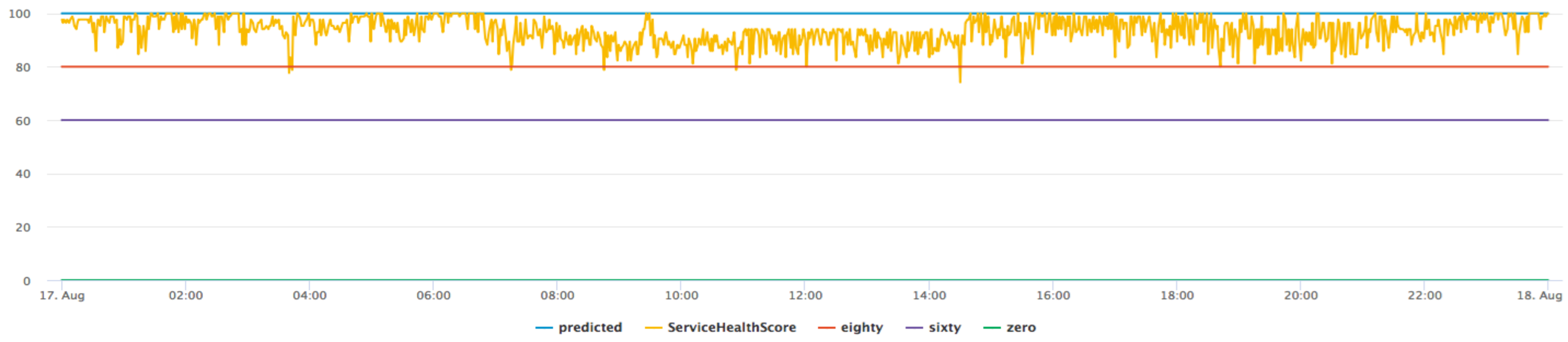
TransUnion and Machine Learning

Anomaly Detection

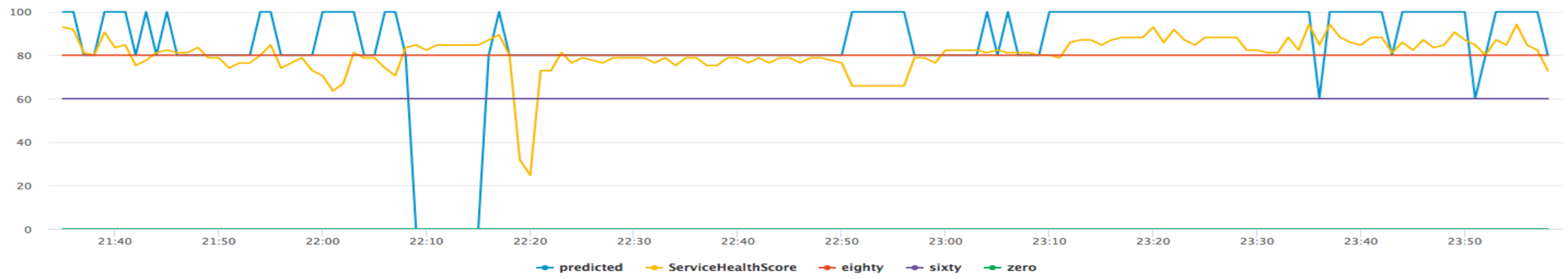
TransUnion and Machine Learning

Predictive Analytics

NORMAL DAY



NON-NORMAL DAY



Investment to Build the Solution

Three months

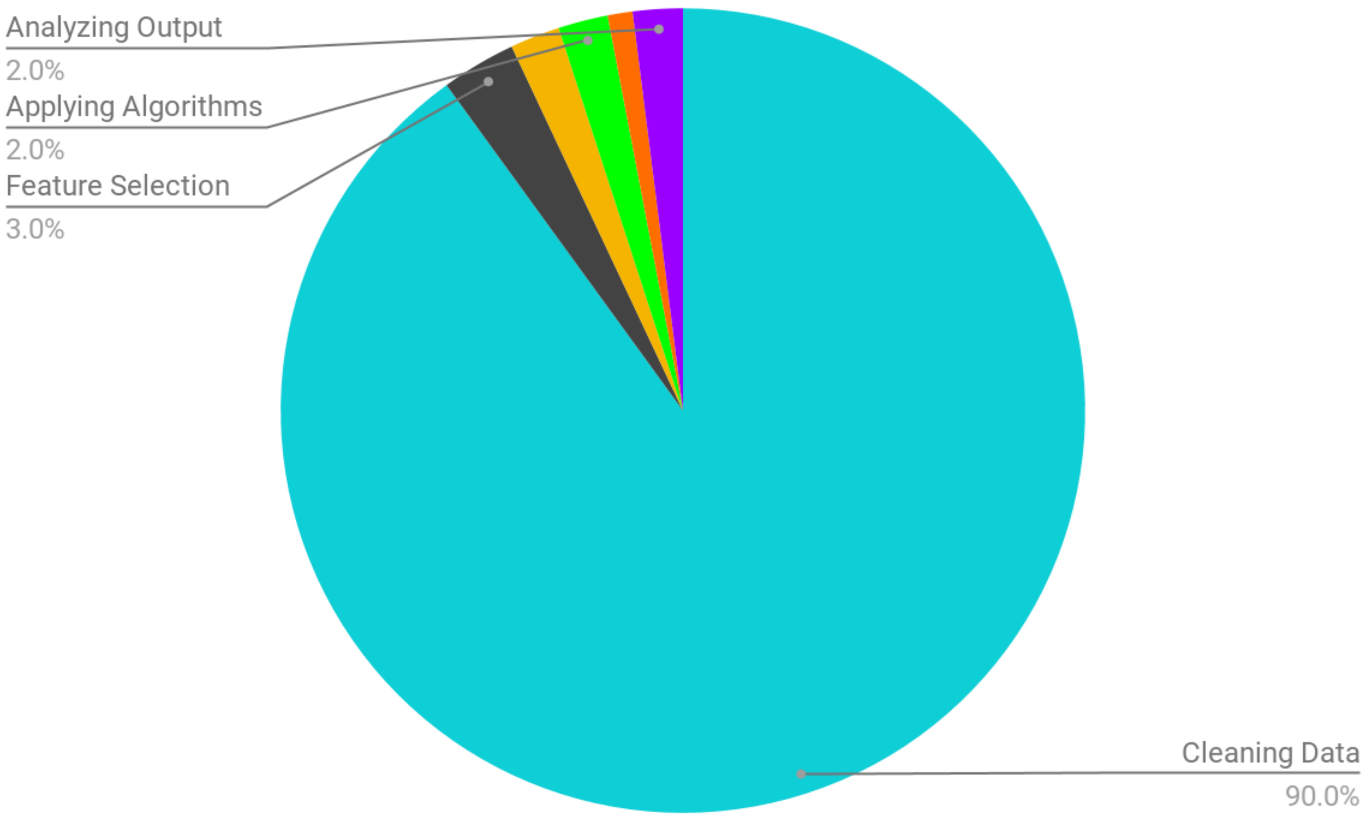
MOST TIME-CONSUMING TASKS:

Obtaining clean quality data

Identifying features

Backfilling service health score

Time Percentage



How Much Effort Does ITSI Save You?

Time + Effort for One Use Case

Just MLTK	ITSI + MLTK	ITSI 4.0
	+ Splunk IT Service Intelligence™	Splunk IT Service Intelligence™
• Two engagements with the Splunk ML Advisory Program • 100+ hours of work over 3 months • 10+ hours of Webex • Multiple business rules	• Leveraged the ITSI and Sophisticated Machine Learning Blog • 30 hours + 1 hour Webex • Everything else was customizing	• ITSI 4.0 now includes this as a turn key feature • Saves a TON of time getting to an outcome

How It Works

Predictive Analytics

Types of Incidents

Two Incident Types

Steady-State Incidents

An Incident Due to a Change

Predictive Analytics Explained

Create a ServiceHealthScoreFromFuture: Read the Blog

```
| bin _time span=1m
| stats min(<FEATURES>) by _time
| eval ServiceHealthScore=(<FEATURES>)/17
| reverse
| streamstats window=10 current=f first(ServiceHealthScore) as ServiceHealthScoreFromFuture
| reverse
| timechart span=1m <FEATURES>
 | eval ServiceHealthFutureState=case(ServiceHealthScoreFromFuture>80,"Green",ServiceHealthScoreFromFuture>60,"Yellow",ServiceHealthScoreFromFuture>40,"Orange"
    ,ServiceHealthScoreFromFuture>0,"Red")
 |fit RandomForestClassifier ServiceHealthFutureState from  <FEATURES>   into Steve_RF_Model_v8
```

https://www.splunk.com/blog/2017/08/28/itsi-and-sophisticated-machine-learning.html

Predictive Analytics Explained

Create a ServiceHealthScore From the Future

- Determine which features have a tight mathematical relationship with the ServiceHealthScore
  - Use the ITSI deep dive view to identify which KPIs started to degrade before the incident occurs
  - Strong leading indicators make excellent features which improve accuracy

Training the Model

Predictive Analytics

Grandfather Paradox: Don’t Use the Future to Predict the Future

Don’t use ServiceHealthScore from the future as your predictor

≠

Applying the Model

The Analysis

Predictive Analytics

The Analysis

Change those string values to numeric for easy visualization

```
| eval predicted=case(Predictive="Green",100, Predictive="Yellow", 80, Predictive="Orange", 60, Predictive="Red", 0)
| fields + _time Predictive predicted
```

✓ 100,606 events (8/12/18 12:00:00.000 AM to 8/13/18 12:00:00.000 AM) No Event Sampling Job Smart Mode

Events Patterns Statistics (1,440) Visualization

100 Per Page Format Preview < Prev 1 2 3 4 5 6 7 8 9 ... Next >

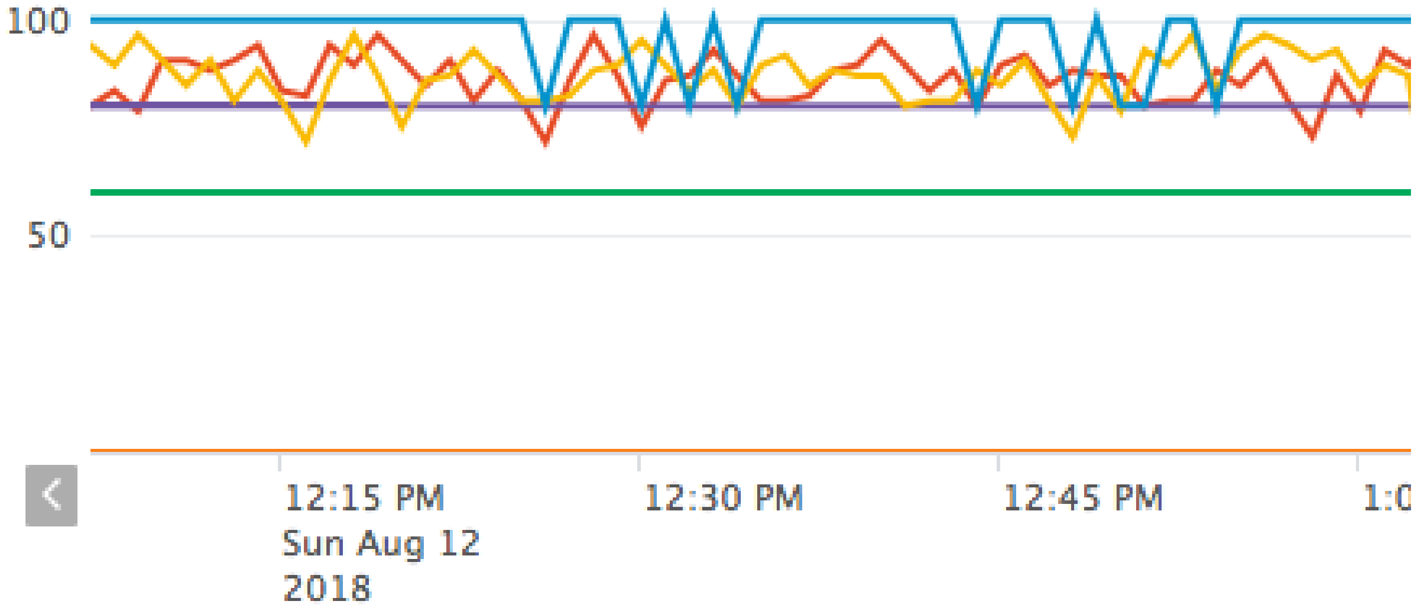
_time	Predictive	predicted
2018-08-12 00:00:00	Green	100
2018-08-12 00:01:00	Green	100

Predictive Analytics

The Analysis

Add boundary lines for easy identification

```
| eval eighty=80
| eval sixty=60
| eval zero=0
```

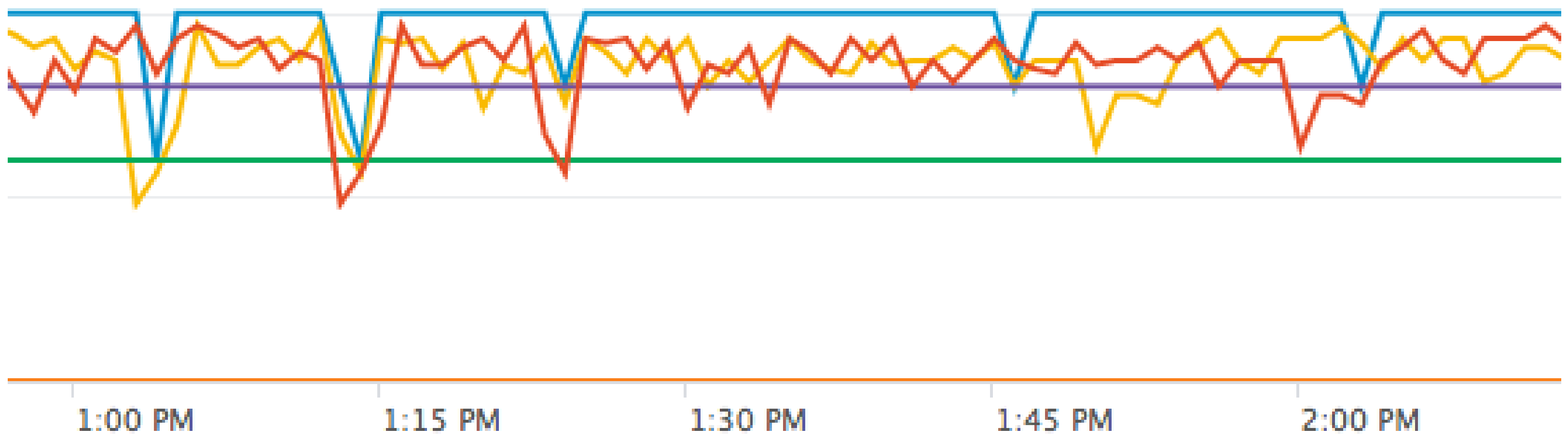


Predictive Analytics

The Analysis

Test against ServiceHealthScoreFromFuture rather than ServiceHealthScore so you don’t have to offset the times in your head

```
53 | timechart span=1m min(predicted) AS predicted  min(ServiceHealthScoreFromFuture) AS ServiceHealthScoreFromFuture
```



Challenges In Predictive Analytics

Challenges

Challenges We Faced

Lots of quality data needed

Slow search speed for large amounts of data

Any minor changes to a KPI requires a new backfill

Dirty data is bad — use adaptive thresholding wisely

Challenges: Accuracy

DIALING IN THE ACCURACY AND FILTERING OUT THE NOISE

- THIS CAN BE SOLVED BY
  - Training on a larger set of data
  - Ensuring clean quality data
  - Visually exploring the data

Challenges

Backfilling the ServiceHealthScore

Any time you add or modify a KPI, it does not retroactively change the ServiceHealthScore

- Change a KPI and you must wait 30 days before having enough quality data to train on

Add a KPI to a service — you must wait to get more runtime until that KPI shows its mathematical relationship with the ServiceHealthScore

- Why not just create a new service with existing/new KPIs and backfill?

ServiceHealthScore Does Not Backfill

Challenges: Custom Predictive Analytics

Backfilling the ServiceHealthScore Through SPL

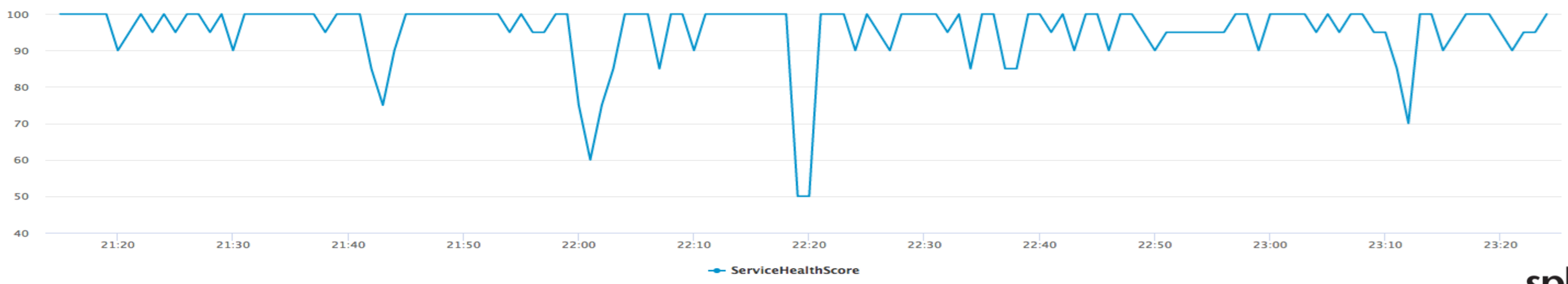
```
index=itsi_summary

| eval Volume=case(kpiid="a3c0cc8213e6120c25eca484" AND serviceid="1ad210ad-329d-4bba-8c31-fc6c878cb608", 'alert_severity')
| eval Response_Time=case(kpiid="1e75e8ee4a395fc86ced70c3" AND serviceid="1ad210ad-329d-4bba-8c31-fc6c878cb608", 'alert_severity')
| eval Errors=case(kpiid="127b54fe58aa59600a64c9d8" AND serviceid="1ad210ad-329d-4bba-8c31-fc6c878cb608", 'alert_severity')
| eval Vendor_Timeouts=case(kpiid="b726f6de942dc7a4ce7842eb" AND serviceid="1ad210ad-329d-4bba-8c31-fc6c878cb608", 'alert_severity')


| eval severity_Errors=case(Errors="normal", 100, Errors="low", 80, Errors="medium", 60, Errors="high", 40, Errors="critical", 0)
| eval severity_Vendor_Timeouts=case(Vendor_Timeouts="normal", 100, Vendor_Timeouts="low", 80, Vendor_Timeouts="medium", 60, Vendor_Timeouts="high", 40,
  Vendor_Timeouts="critical", 0)
| eval severity_Response_Time=case(Response_Time="normal", 100, Response_Time="low", 80, Response_Time="medium", 60, Response_Time="high", 40, Response_Time
  ="critical", 0)
| eval severity_Volume=case(Volume="normal", 100, Volume="low", 80, Volume="medium", 60, Volume="high", 40, Volume="critical", 0)

```

_time	ServiceHealthScore	severity_Errors	severity_Vendor_Timeouts	severity_Response_Time	severity_Volume
2018-08-11 22:15:00	100	100	100	100	100
2018-08-11 22:16:00	100	100	100	100	100
2018-08-11 22:17:00	100	100	100	100	100
2018-08-11 22:18:00	100	100	100	100	100
2018-08-11 22:19:00	50	0	100	0	100
2018-08-11 22:20:00	50	0	100	0	100
2018-08-11 22:21:00	100	100	100	100	100
2018-08-11 22:22:00	100	100	100	100	100
2018-08-11 22:23:00	100	100	100	100	100
2018-08-11 22:24:00	90	100	100	60	100



Pro Tips

Predictive Analytics

Customer ML Tips and Tricks

Pro Tips From the Splunk Trust

- Version each model you create
- Make sure your Service Health Score is aligned with known incidents
- Ensure thresholds are set properly in ITSI
- Validate that regular expressions are capturing correct values
- Make your KPIs as granular as possible

Bring This to Your Organization

Where Do I Start?

splunk> .conf18

How to Get Started With Custom Predictive Analytics

Use ITSI to build a top-level view of your most critical services to understand the input variables needed.

Aggregate indicators into a single Service Health Score.

Use these KPIs to train your models.

Select several KPIs with good runtime and create a backfilled Service Health Score.

Align that Service Health Score against known incidents to test effectiveness.

Train a model and experiment with different algorithms.

Use the MLTK to get feedback about the models you train.

Understand the difference between algorithms.

Use the Service Health Score calculation and search for a score lower than 60%.

Run this over the last six months to pinpoint your larger incidents with day and time.

Create a report so you can use it to go back and identify incidents.

Test your models against known incidents.

Thank You! Questions?

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