© 2019 SPLUNK INC.

The New Experiment Experience

What's new in Machine Learning Toolkit (MLTK) 5.0

Ryan Oriecuia & Gyanendra Rana

© 2019 SPLUNK INC.

The New Experiment Experience

What's new in Machine Learning Toolkit (MLTK) 5.0

Ryan Oriecuia

Principal Software Developer | Splunk

Gyanendra Rana

Senior Product Manager | Splunk

splunk> .conf19

© 2019 SPLUNK INC.

Forward-Looking Statements

During the course of this presentation, we may make forward-looking statements regarding future events or plans 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 may differ materially. The forward-looking statements made in the this presentation are being made as of the time and date of its live presentation. If reviewed after its live presentation, it may not contain current or accurate information. We do not assume any obligation to update any forward-looking statements made herein.

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 functionalities described or to include any such feature or functionality in a future release.

Splunk, Splunk>, Turn Data Into Doing, The Engine for Machine Data, Splunk Cloud, Splunk Light and SPL are trademarks and registered trademarks of Splunk Inc. in the United States and other countries. All other brand names, product names, or trademarks belong to their respective owners. © 2019 Splunk Inc. All rights reserved.

© 2019 SPLUNK INC.

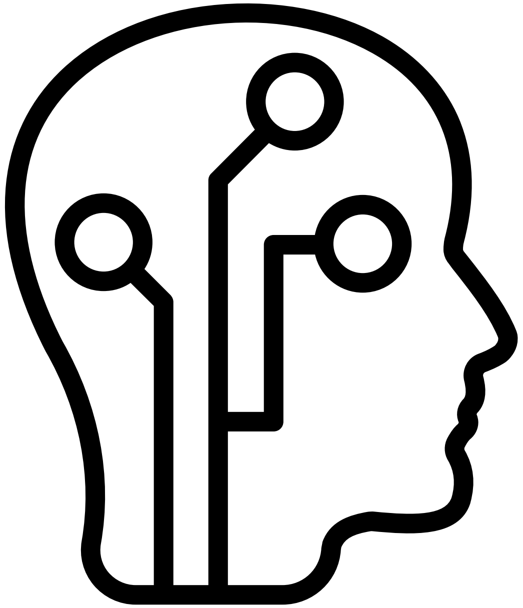
What’s new since MLTK 4.0.0

Showcase redesign

Smart Forecasting

Smart Outlier Detection

…and more



© 2019 SPLUNK INC.

How do I make my machine learn?

Is this even something ML will help with?

Where do I start? What are the steps?

How do I know if this is even working?

How do I tweak things to make it work better?

Now that it’s working… how do I make it go?

© 2019 SPLUNK INC.

You have help

Get by with a little help from your friends

Specialized apps

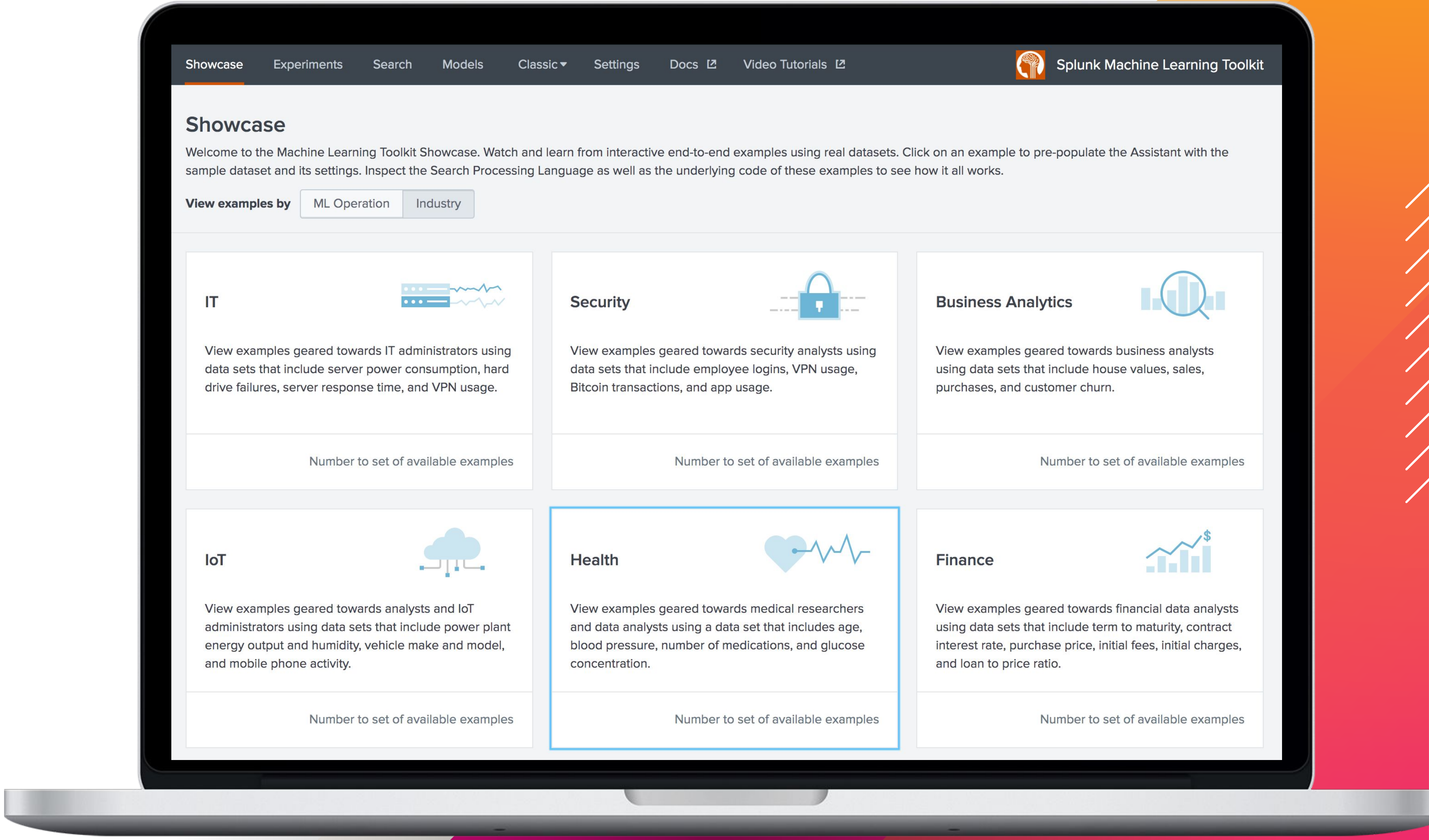
- Splunk User Behavior Analytics
- IT Service Intelligence

Machine Learning Toolkit

- Showcase
- Assistants
- Experiments

Showcase

Examples and ideas



Showcase

Welcome to the Machine Learning Toolkit Showcase. Watch and learn from interactive end-to-end examples using real datasets. Click on an example to pre-populate the Assistant with the sample dataset and its settings. Inspect the Search Processing Language as well as the underlying code of these examples to see how it all works.

View examples by ML Operation | Industry

IT

View examples geared towards IT administrators using data sets that include server power consumption, hard drive failures, server response time, and VPN usage.

Number to set of available examples

Security

View examples geared towards security analysts using data sets that include employee logins, VPN usage, Bitcoin transactions, and app usage.

Number to set of available examples

Business Analytics

View examples geared towards business analysts using data sets that include house values, sales, purchases, and customer churn.

Number to set of available examples

IoT

View examples geared towards analysts and IoT administrators using data sets that include power plant energy output and humidity, vehicle make and model, and mobile phone activity.

Number to set of available examples

Health

View examples geared towards medical researchers and data analysts using a data set that includes age, blood pressure, number of medications, and glucose concentration.

Number to set of available examples

Finance

View examples geared towards financial data analysts using data sets that include term to maturity, contract interest rate, purchase price, initial fees, initial charges, and loan to price ratio.

Number to set of available examples

© 2019 SPLUNK INC.

© 2019 SPLUNK INC.

“Classic” Assistants

Predict Fields, Detect Outliers, Forecast Time Series, Cluster Events

They’re great!

- Freeform search for full Splunk power
- Custom configuration UI
- Custom visualizations

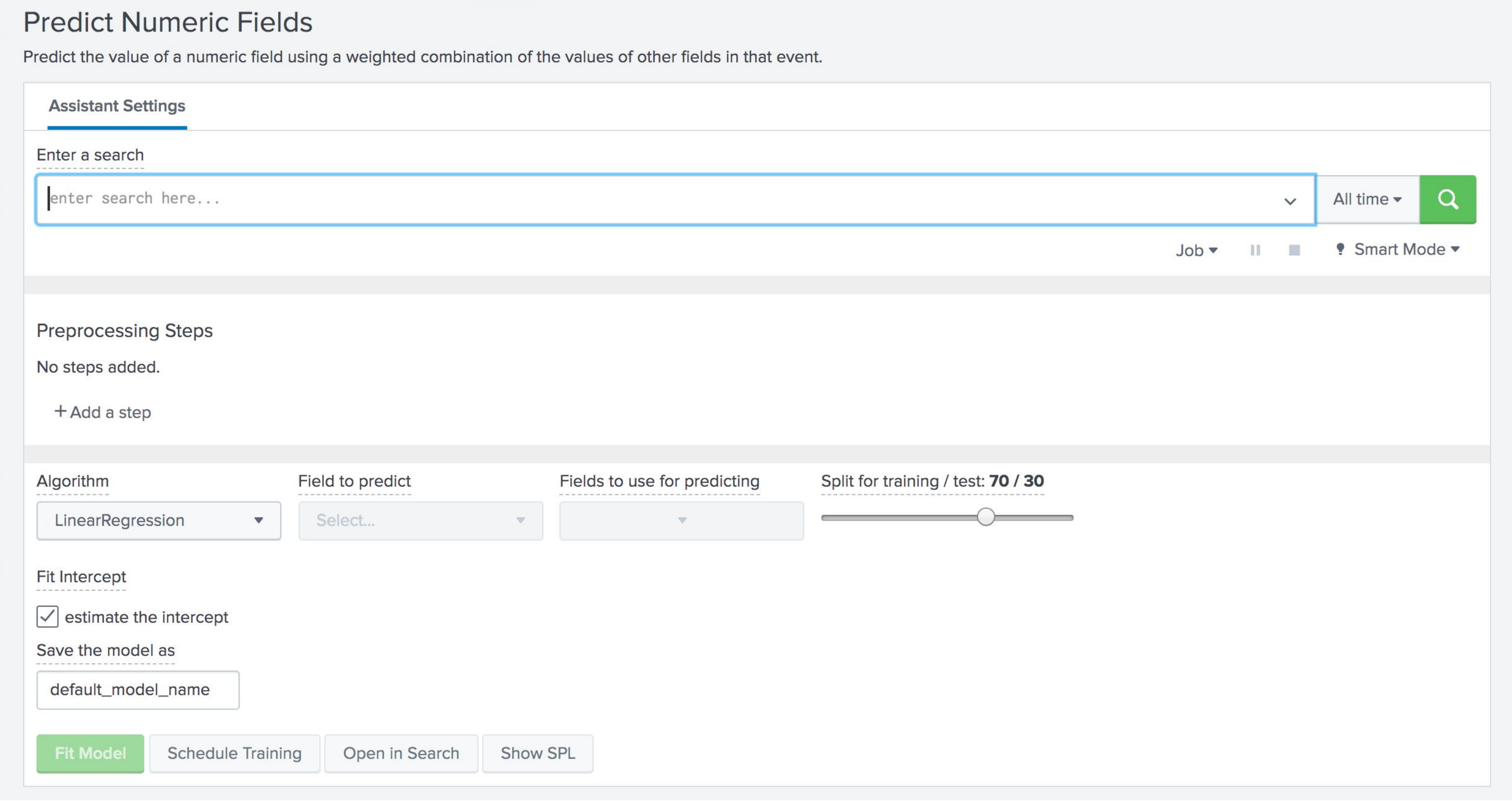
But…

- A lot of UI all at once
- Comparing different configurations is difficult
- No visualization for input data
- No “Save”
  - Export SPL

© 2019 SPLUNK INC.

“Classic” Assistants

Predict Numeric Fields



Predict Numeric Fields

Predict the value of a numeric field using a weighted combination of the values of other fields in that event.

Assistant Settings

Enter a search

enter search here...

All time ▾

Job ▾ Smart Mode ▾

Preprocessing Steps

No steps added.

+ Add a step

Algorithm: LinearRegression ▾

Field to predict: Select... ▾

Fields to use for predicting: ▾

Split for training / test: **70 / 30**

Fit Intercept

☑ estimate the intercept

Save the model as: default_model_name

© 2019 SPLUNK INC.

Experiments

A management layer on top of Assistants

They're great!

- Comparing different configurations is easier (via Experiment History)
- Save / load settings
- Plus everything in the Classic Assistants
  - Freeform search for full Splunk power
  - Custom configuration UI
  - Custom visualizations

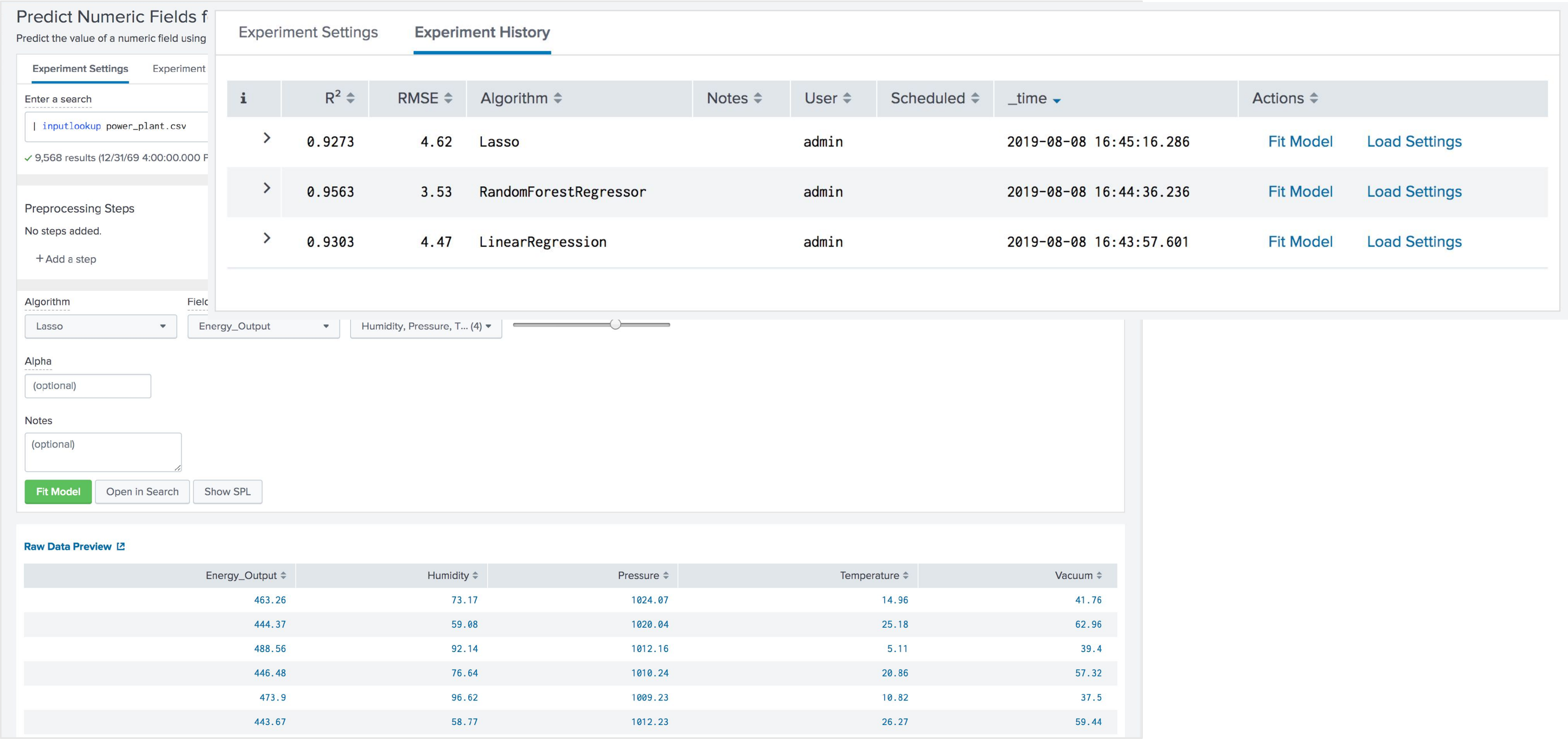
But…

- A lot of UI all at once

© 2019 SPLUNK INC.

Experiments

A management layer on top of Assistants



Predict Numeric Fields f

Predict the value of a numeric field using

Enter a search

```
| inputlookup power_plant.csv
```

✓ 9,568 results (12/31/69 4:00:00.000 F

Preprocessing Steps

No steps added.

Algorithm: Lasso

Fielc: Energy_Output

Humidity, Pressure, T... (4)

Alpha: (optional)

Notes: (optional)

i	R^2	RMSE	Algorithm	Notes	User	Scheduled	_time	Actions	
>	0.9273	4.62	Lasso		admin		2019-08-08 16:45:16.286	Fit Model	Load Settings
>	0.9563	3.53	RandomForestRegressor		admin		2019-08-08 16:44:36.236	Fit Model	Load Settings
>	0.9303	4.47	LinearRegression		admin		2019-08-08 16:43:57.601	Fit Model	Load Settings

Raw Data Preview

Energy_Output	Humidity	Pressure	Temperature	Vacuum
463.26	73.17	1024.07	14.96	41.76
444.37	59.08	1020.04	25.18	62.96
488.56	92.14	1012.16	5.11	39.4
446.48	76.64	1010.24	20.86	57.32
473.9	96.62	1009.23	10.82	37.5
443.67	58.77	1012.23	26.27	59.44

© 2019 SPLUNK INC.

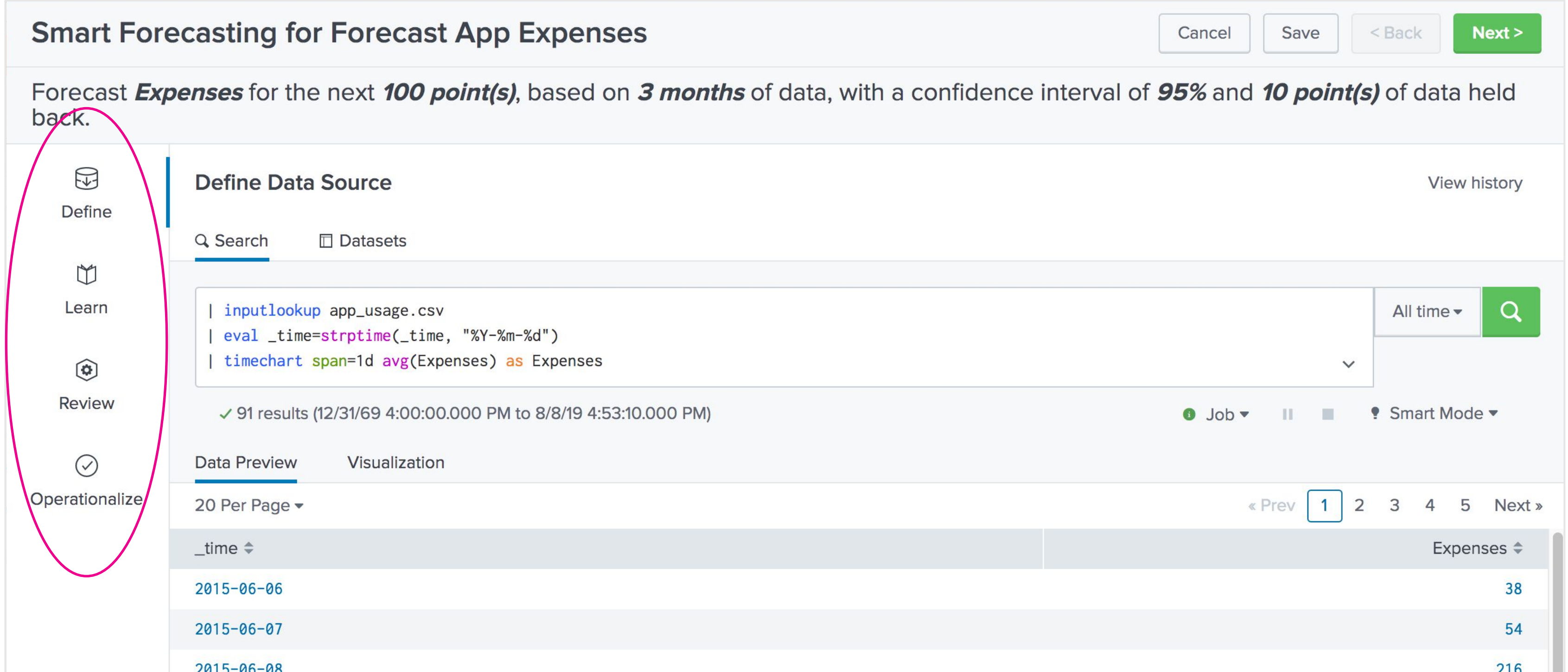
New Smart Assistants

Smart Forecasting, Smart Outlier Detection

Designed around the machine learning workflow

Each workflow step has its own place

Only via Experiments



Smart Forecasting for Forecast App Expenses

Forecast ***Expenses*** for the next ***100 point(s)***, based on ***3 months*** of data, with a confidence interval of ***95%*** and ***10 point(s)*** of data held back.

Define Data Source

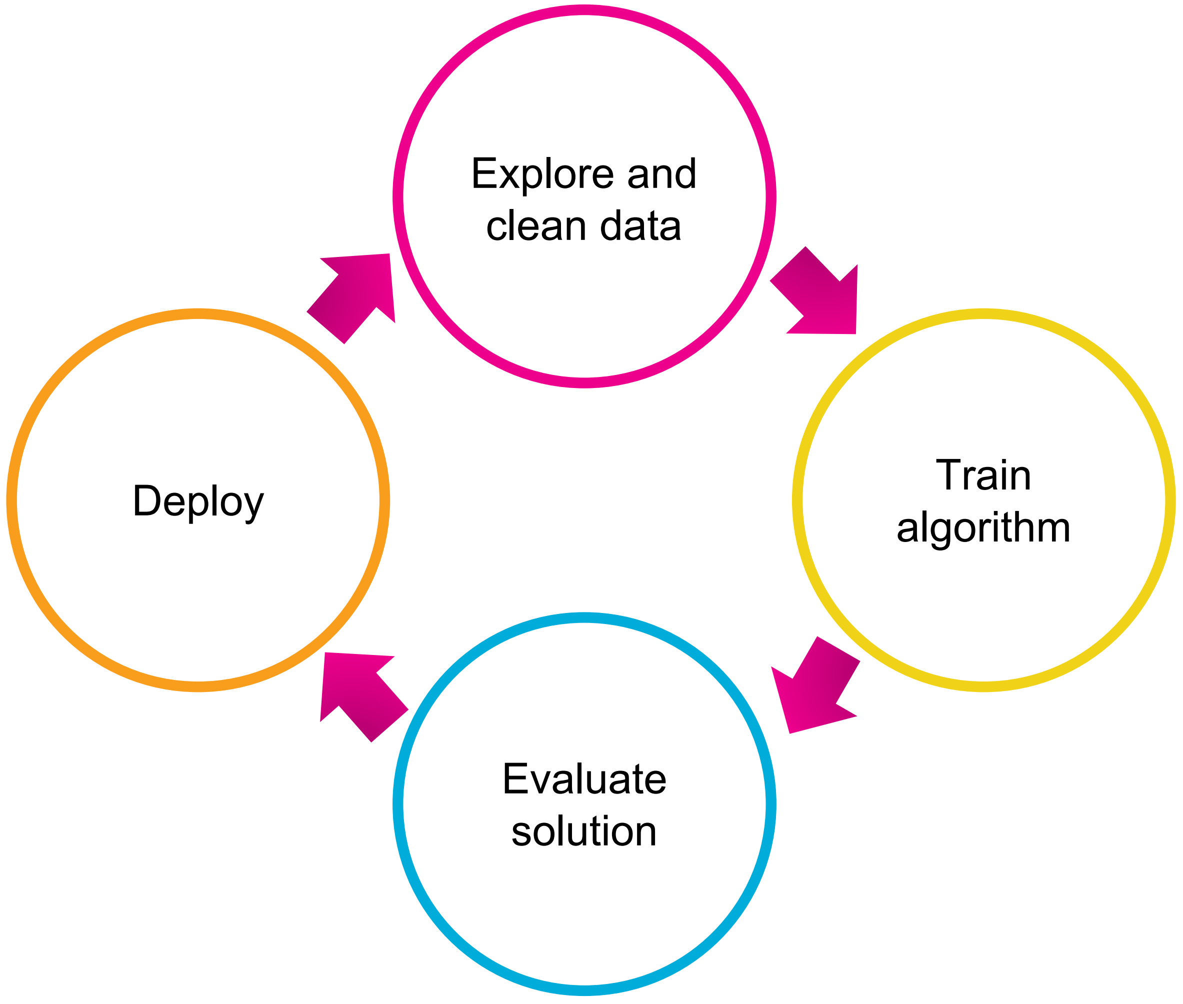
```
| inputlookup app_usage.csv
| eval _time=strptime(_time, "%Y-%m-%d")
| timechart span=1d avg(Expenses) as Expenses
```

✓ 91 results (12/31/69 4:00:00.000 PM to 8/8/19 4:53:10.000 PM)

_time	Expenses
2015-06-06	38
2015-06-07	54
2015-06-08	216

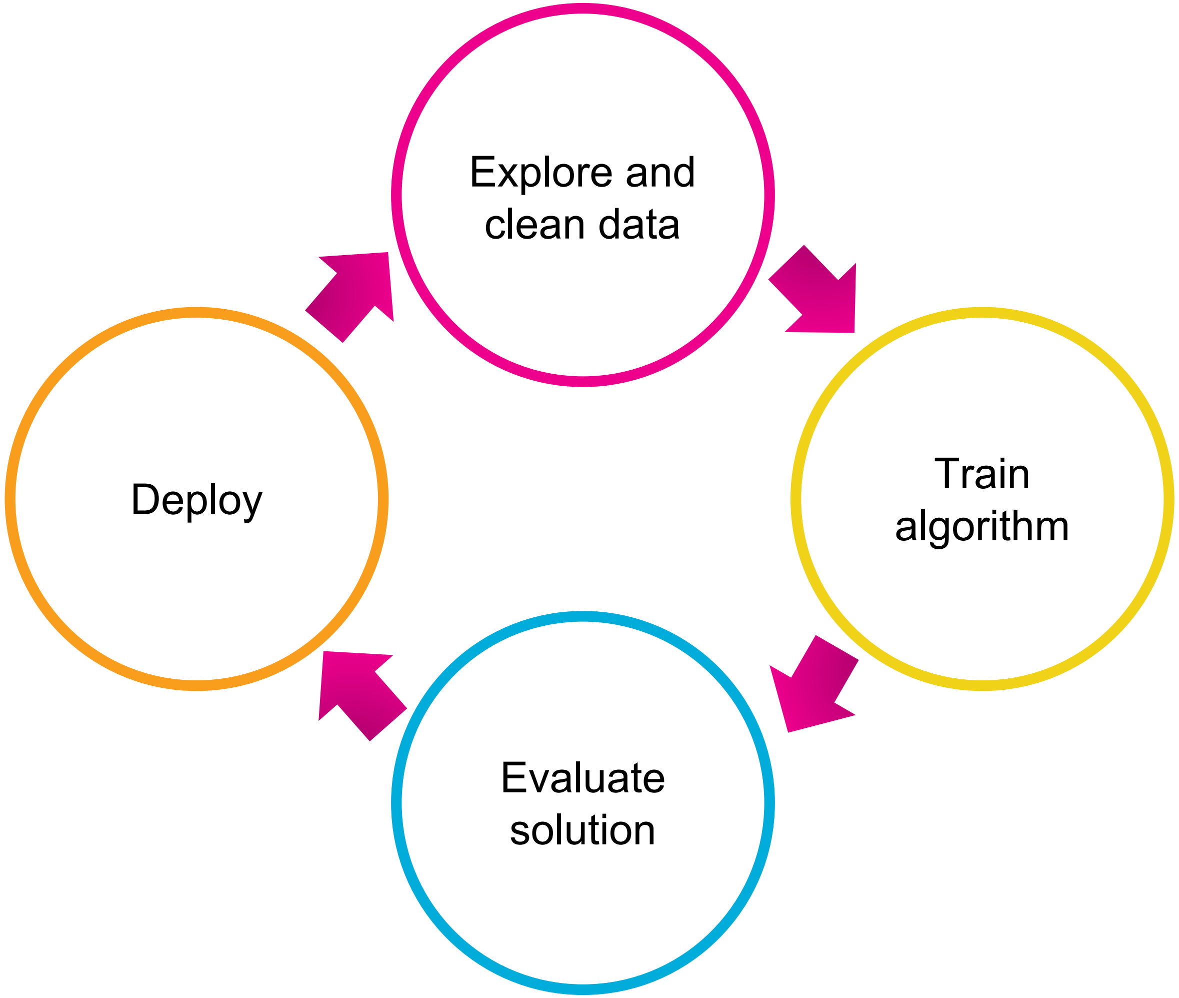
© 2019 SPLUNK INC.

The machine learning workflow



© 2019 SPLUNK INC.

The machine learning workflow



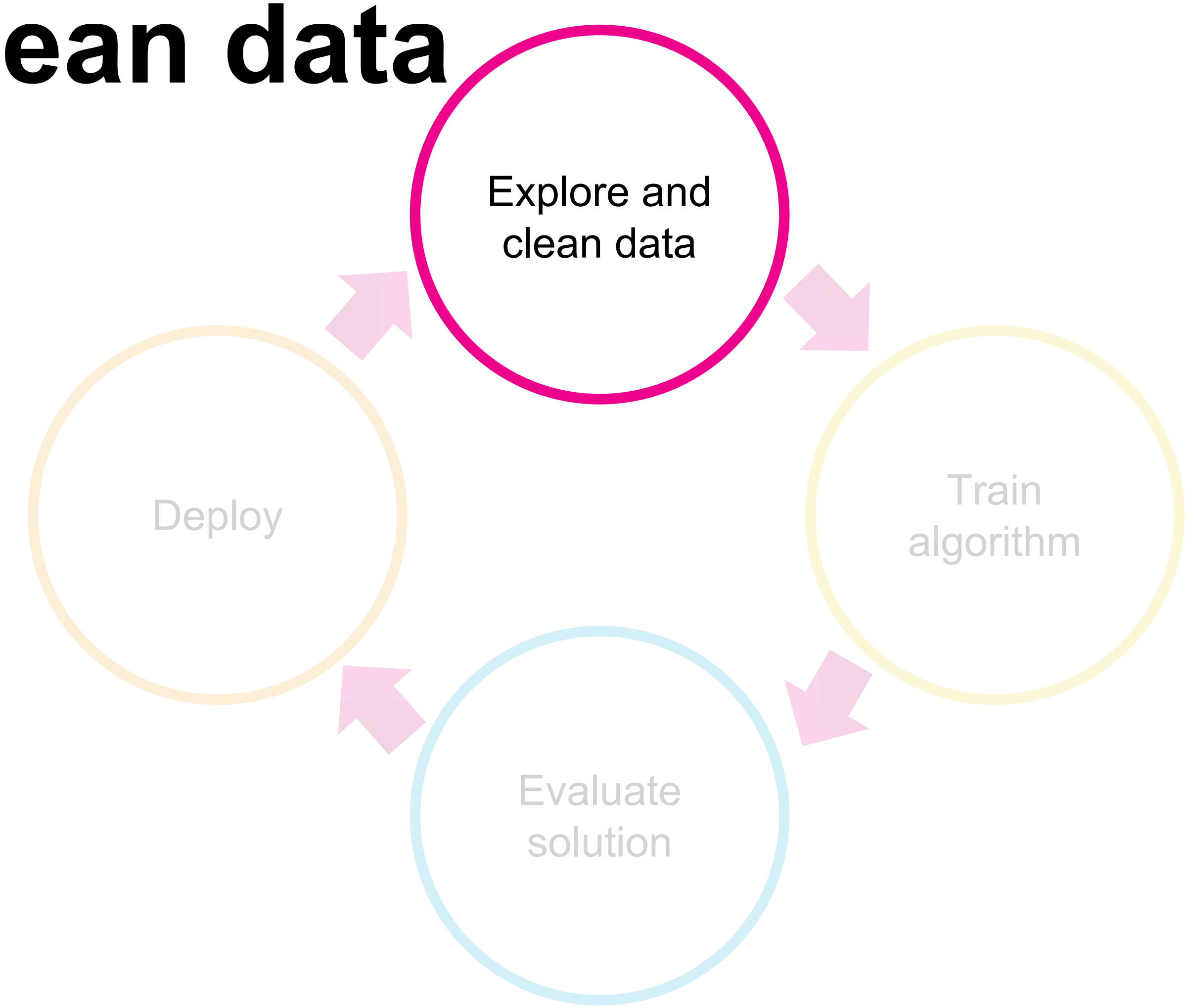
© 2019 SPLUNK INC.

Define: Explore and clean data

Full Splunk search

- Data tables and visualization to aid exploration

Datasets support



© 2019 SPLUNK INC.

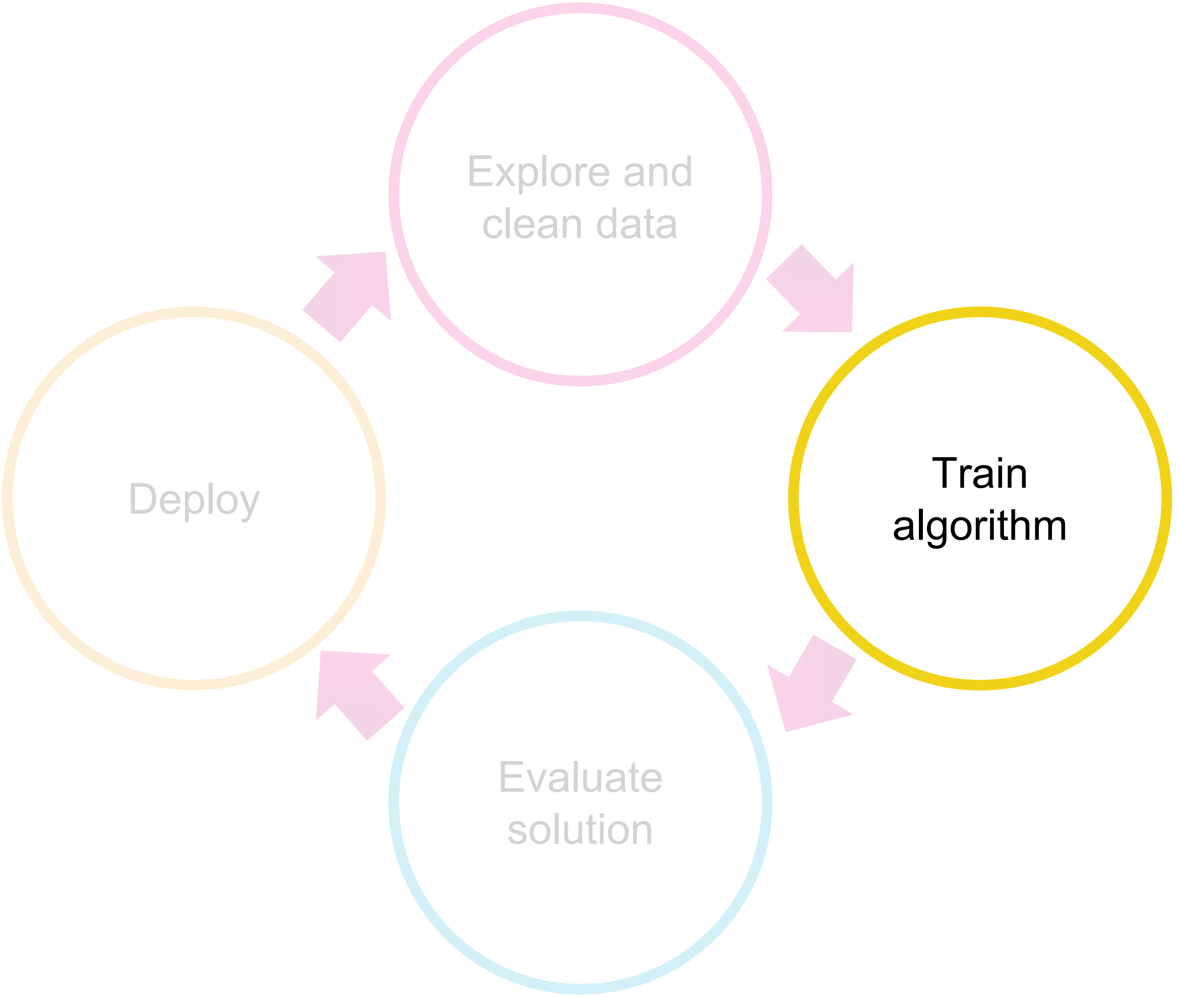
Learn: Train algorithm

Define hyperparameters

New Smart Assistants demand less configuration

- Forecast and Outlier Detection available now
- Clustering coming soon

Preprocessing and fit steps separated



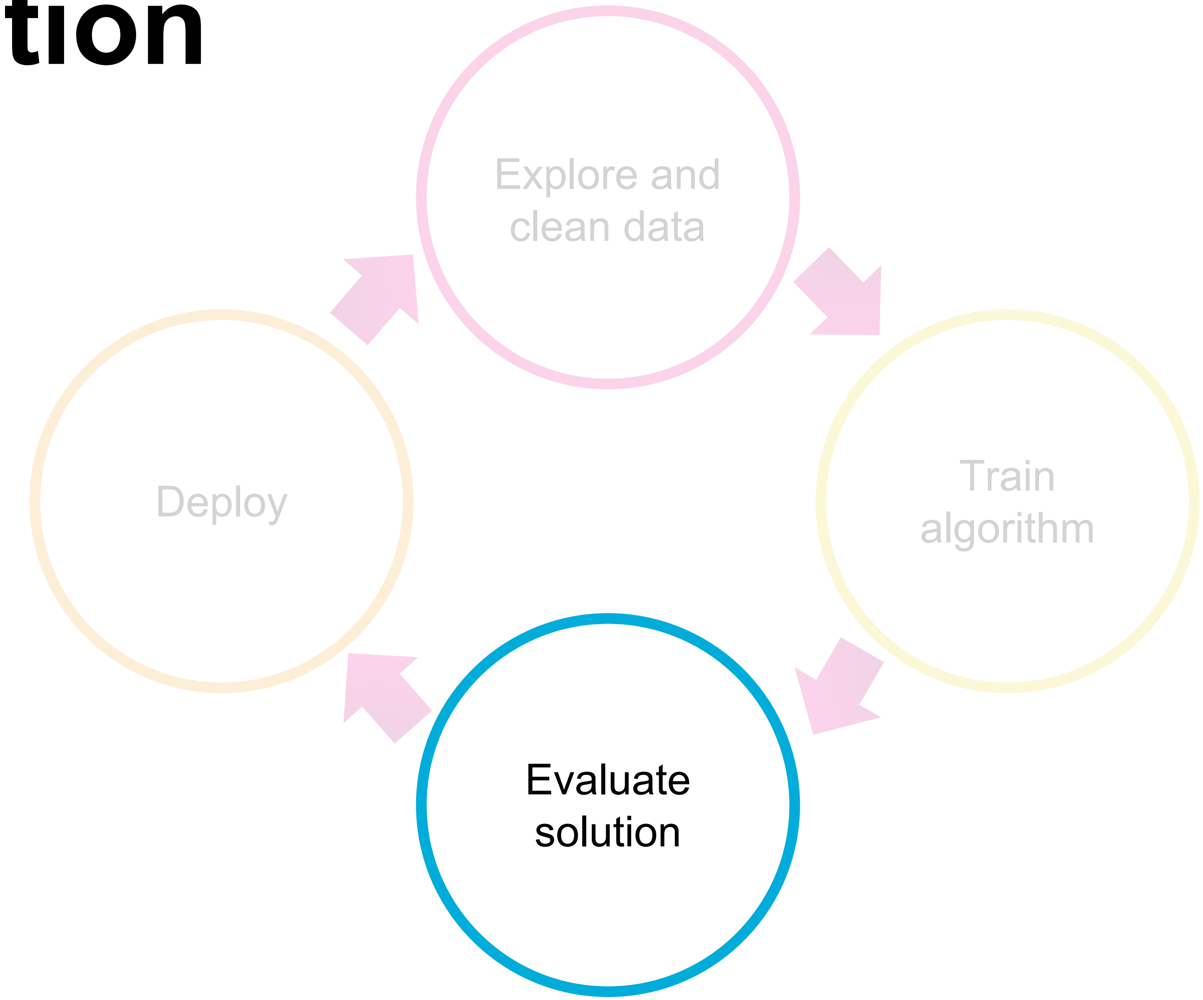
© 2019 SPLUNK INC.

Review: Evaluate solution

Algorithm-specific visualizations

Summary stats

- Common statistics (e.g., R^2, RMSE)
- Algorithm-specific measures



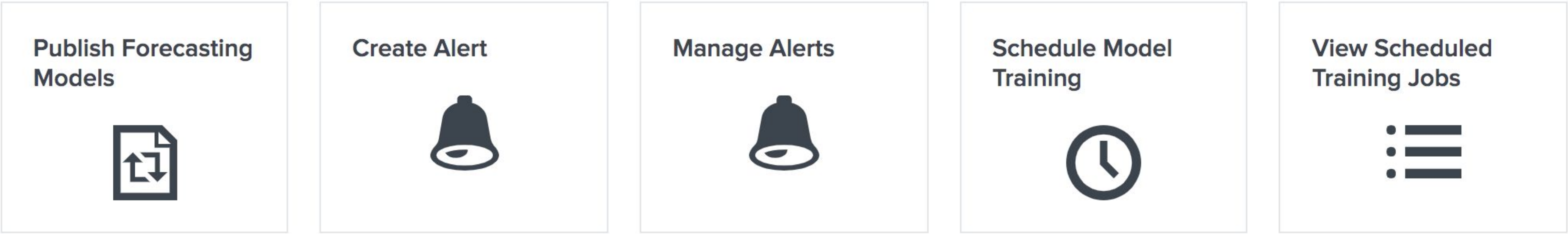
© 2019 SPLUNK INC.

Operationalize: Deploy

All the options from other Experiments

- Publish to other apps
- Create and manage alerts
- Create and manage scheduled learning

Explore and clean data
Train algorithm
Deploy



© 2019 SPLUNK INC.

Iterate

Tweak to your heart’s content

- Compare current performance vs. historical runs
- Refit your existing model with new data
- Select a winner, load/fit, and save/publish the model

© 2019 SPLUNK INC.

Why would I use forecasting?

Typically used for planning

- Based on past trends, what do we expect next week/month/quarter/year to look like?
- Capacity planning (hard drive, operating temperature)

Forecasting is not a crystal ball, but it gives you a quantitative estimate on future values

- Getting a picture of what the future **might** look like.

© 2019 SPLUNK INC.

Using the old way for forecasting

There’s nothing wrong with the old way, it’s just often improperly used

You have to be an expert at the math

- You have to specify the algorithm mode to use for the **predict** command
- You have to know how to optimize on P, D, and Q parameters for ARIMA

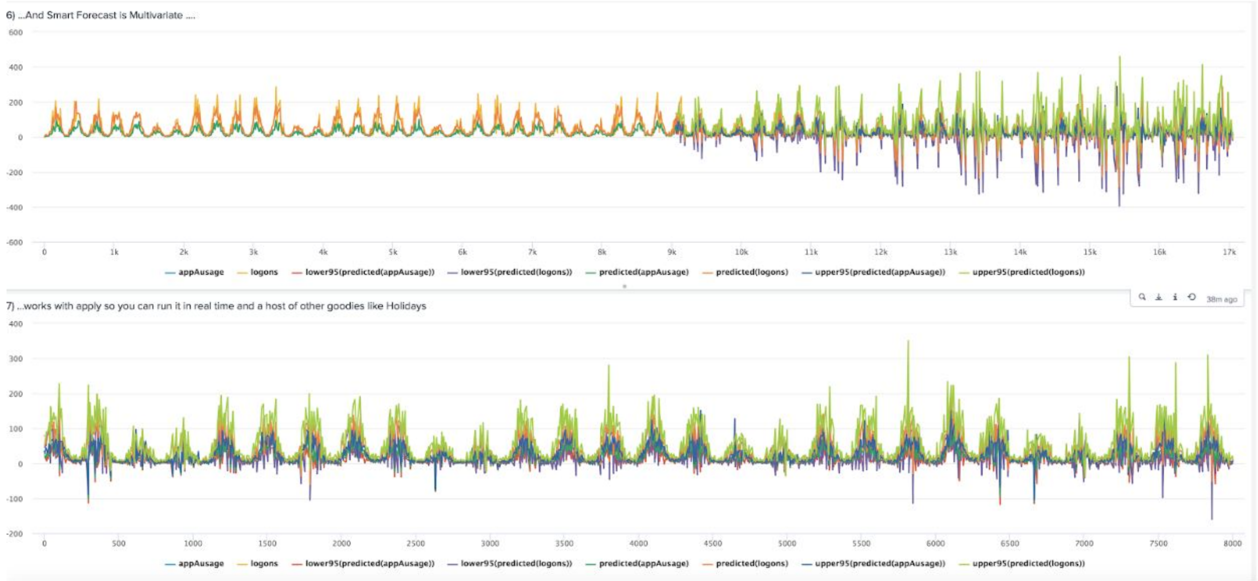
There is no model file created, which means you can’t “apply” your model to future data

Doesn’t consider special days (holidays)

© 2019 SPLUNK INC.

StateSpace Algorithm for Smart Forecasting

- Persists models with the fit command, use with apply
- Forecast multiple time series data into the future together as a unified system
- Add Special Days to improve forecasts by accounting for days which should be treated differently, like Calendar Holidays, Black Friday sales or IP traffic on July 4th
- Automatically imputes missing values
- No need to choose parameters or mode



6) ...And Smart Forecast is Multivariate ....

7) ...works with apply so you can run it in real time and a host of other goodies like Holidays

© 2019 SPLUNK INC.

Demo

© 2019 SPLUNK INC.

Splunk Machine Learning Advisory Program

Get help from the Splunk Data Scientists to solve your business use case with Machine Learning Toolkit

1. Get help from the Splunk Data Scientists to solve your business use case with Machine Learning Toolkit
2. Complimentary support with your Enterprise or Cloud license
3. Early access to new Machine Learning features
4. Results in opportunity to tell your success story with Splunk
5. Contact mlprogram@splunk.com for more information

© 2019 SPLUNK INC.

Review: How do I make my machine learn?

Is this even something ML will help with?

- Check the MLTK Showcase for inspiration

Where do I start? What are the steps?

- Use a “Smart” assistants and move through the steps in order
- Don’t forget the “classic” assistants

How do I know if this is even working?

- Use Learn’s Evaluate tab and Review’s scoring metrics

How do I tweak things to make it work better?

- Explore different settings and use History to load the one that works best

Now that it’s working… how do I make it go?

- Operationalize!

© 2019 SPLUNK INC.

Operationalize this session

Install / update the Machine Learning Toolkit (it’s free!)

Try out the new Assistants

- Leverage the Machine Learning Customer Advisory Program
- Ask questions at answers.splunk.com

Let us know what you think

- Send feedback to mlprogram@splunk.com

© 2019 SPLUNK INC.

Q&A

Ryan Oriecuia | Principal Software Developer
Gyanendra Rana | Senior Product Manager

© 2019 SPLUNK INC.

.conf19

splunk>

Thank You!

Go to the .conf19 mobile app to

RATE THIS SESSION