cyber & data

"From bits to information"

Introduction to Splunk and Machine Learning – Part 2

Outline

- Introduction to Machine Learning.
- ML Experiments.

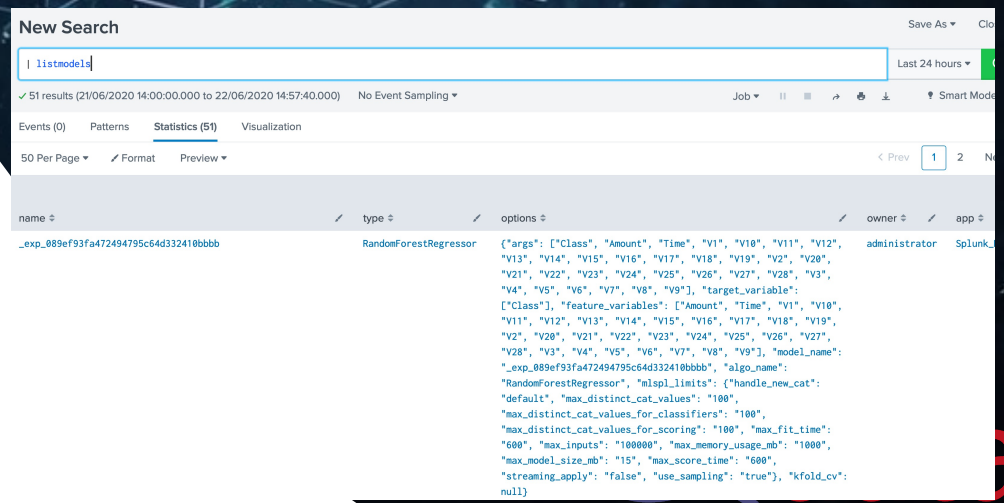
Splunk and Machine Learning

- fit. Fit a model
- apply. Apply a model run by the fit command.
- summary. Show summary of model.
- listmodels. List the models.
- deletemodel. Delete a model.
- score. Show scores for tests.

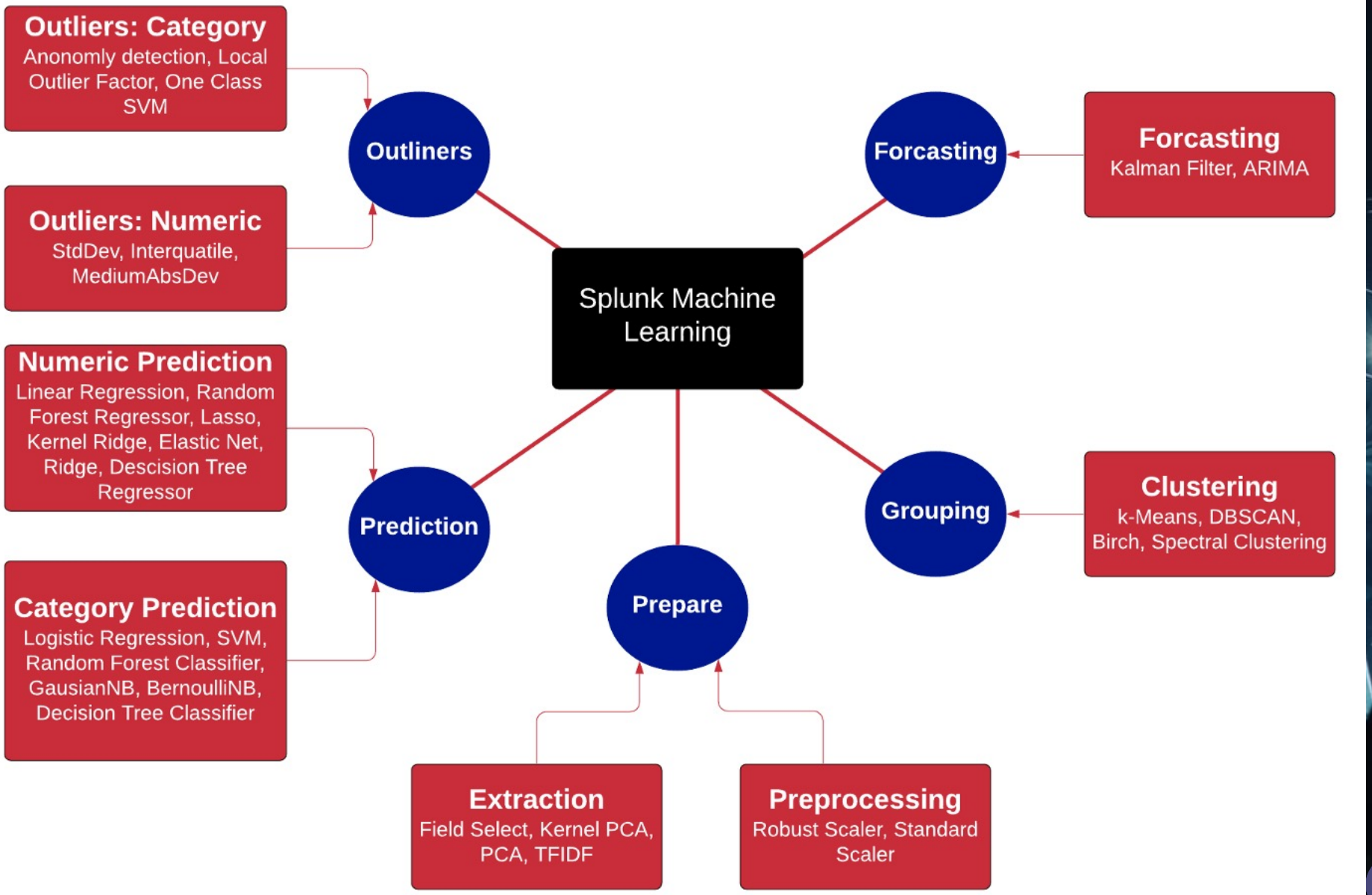
| inputlookup iris.csv
| fit GaussianNB petal_length from * into myModel
| apply myModel as new_petal link.

| summary myModel link

| listmodels link



Machine Learning

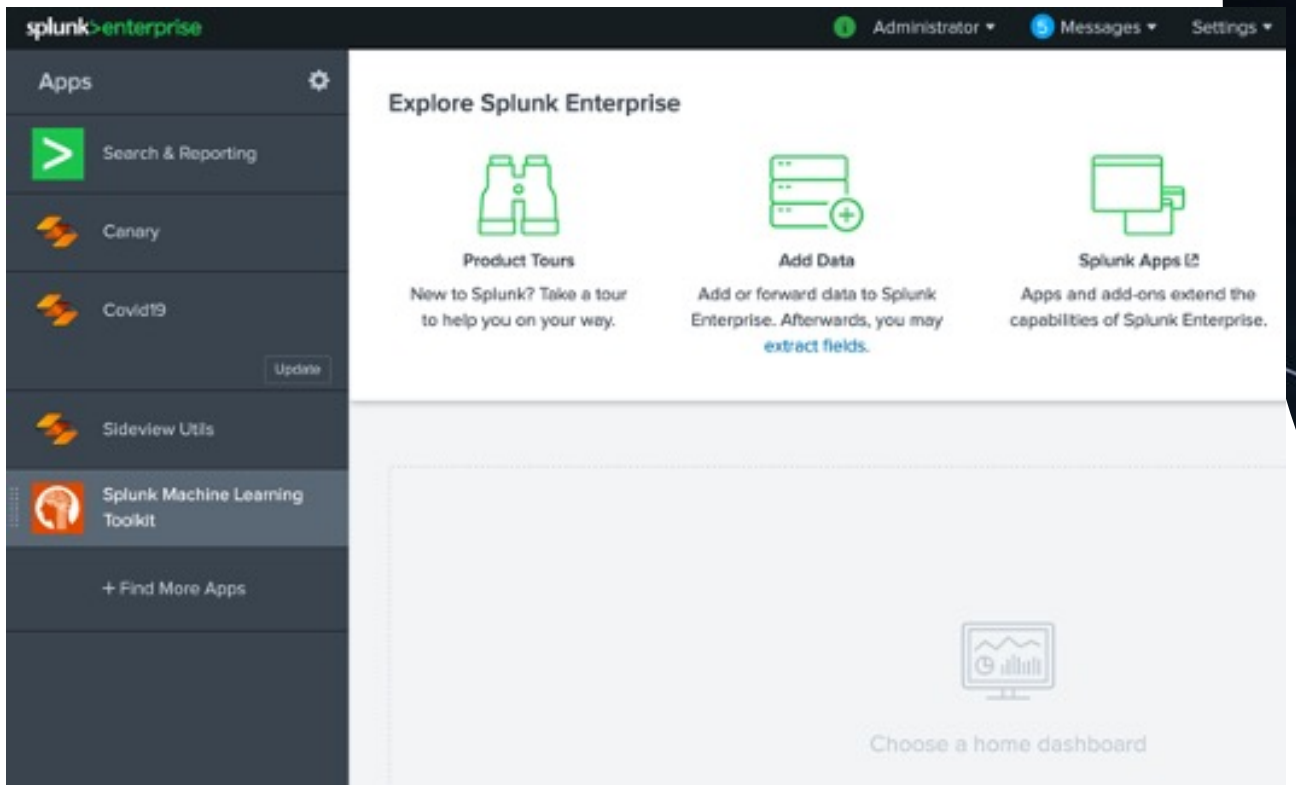


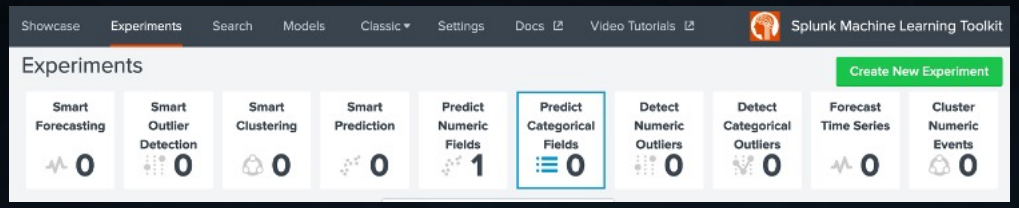
cyber & data

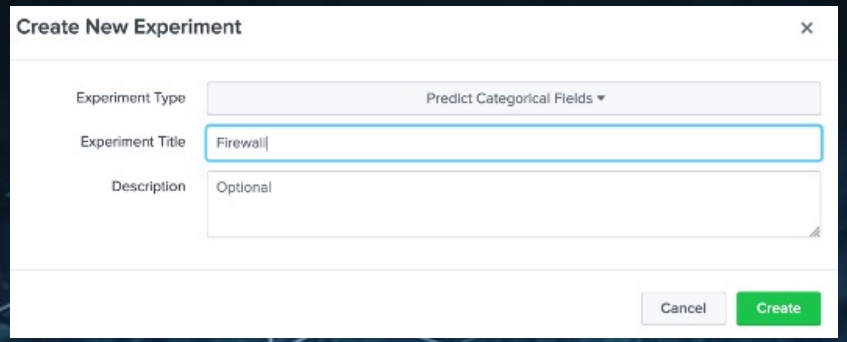
"From bits to information"

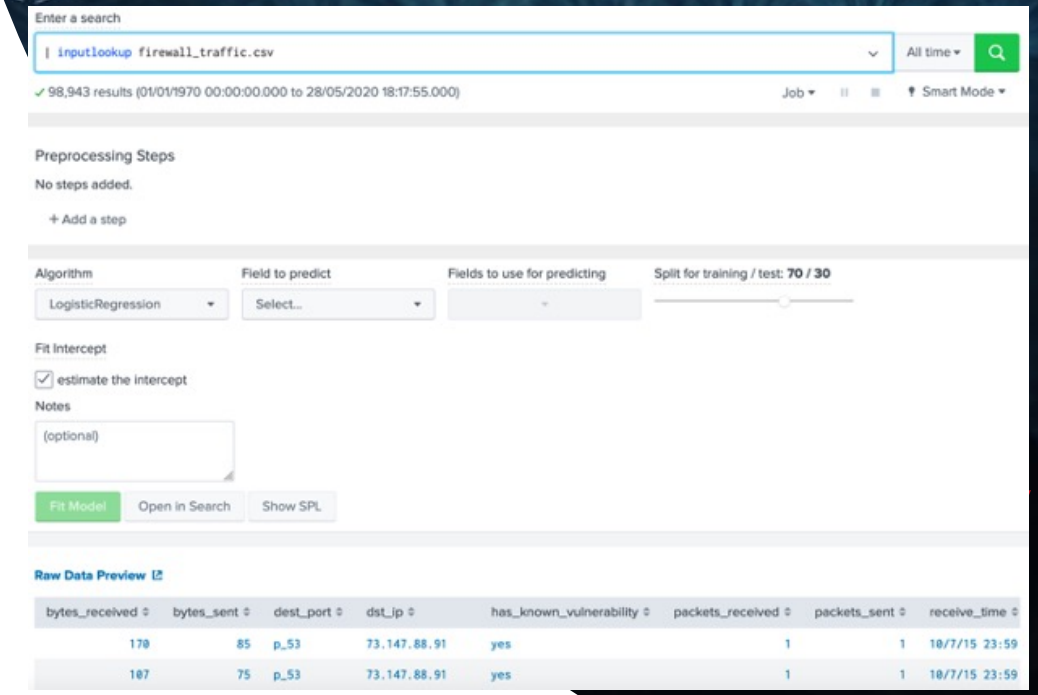
Predicting (Categories)

Splunk ML



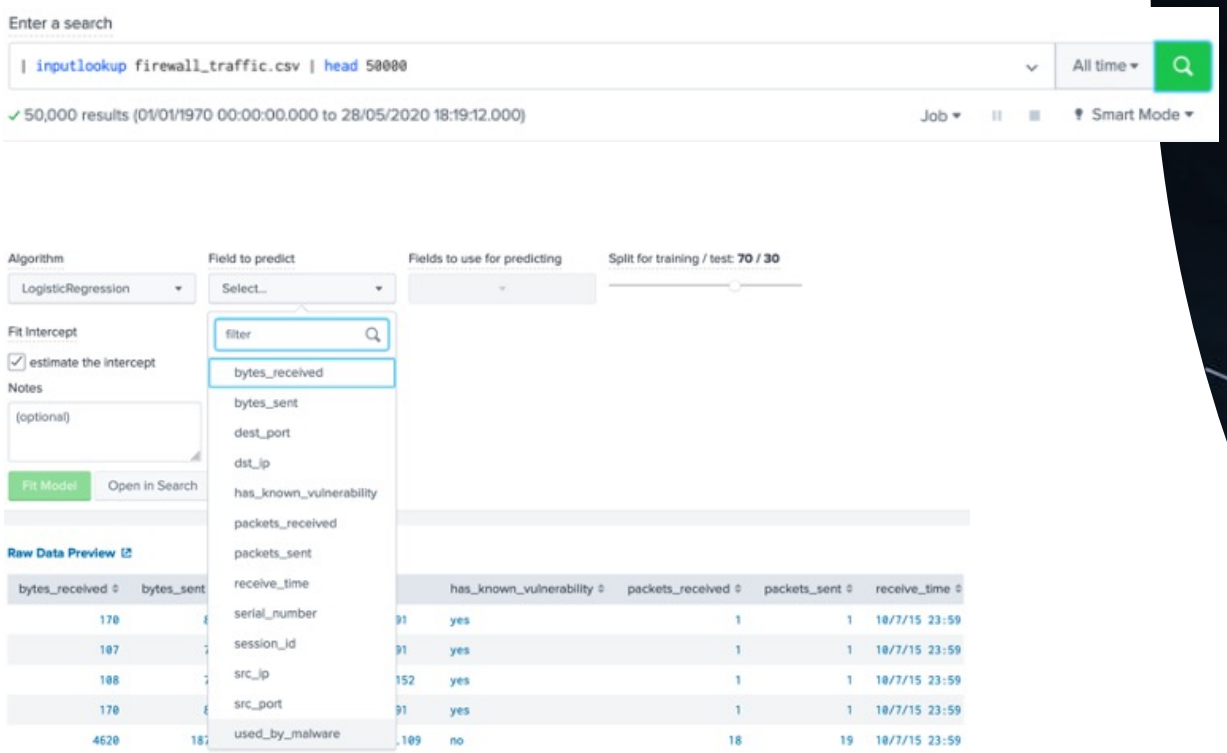


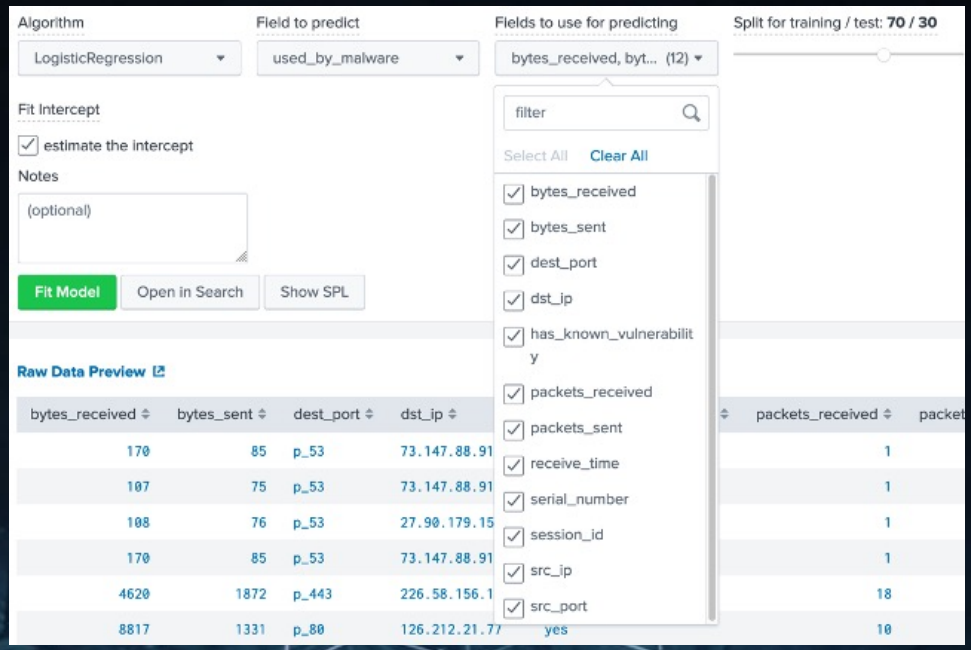




Site

Logistic Regression





used_by_malware	predicted(used_by_malware)	bytes_received	bytes_sent	dest_port	dst_ip	has_known_vulnerability
no	no	4620	1872	p_443	226.58.156.109	no
yes	no	4160	1243	p_443	47.242.134.132	yes
yes	no	507	976	p_443	204.243.248.73	yes
yes	yes	3950	2447	p_443	84.216.108.116	yes
yes	yes	3950	2447	p_443	84.216.108.116	yes
yes	yes	98	86	p_53	73.147.88.91	yes
yes	yes	98	86	p_53	73.147.88.91	yes
yes	yes	3950	2447	p_443	84.216.108.116	yes
yes	yes	456	669	p_80	72.8.163.120	yes
yes	no	572	1018	p_443	32.246.18.81	yes

Precision	Recall	Accuracy	F1
0.83	0.82	0.82	0.83

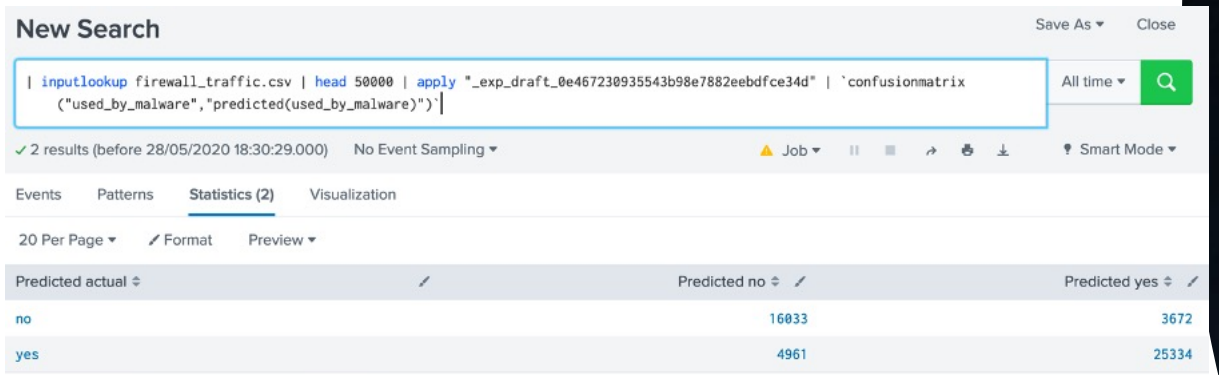
Classification Results (Confusion Matrix)

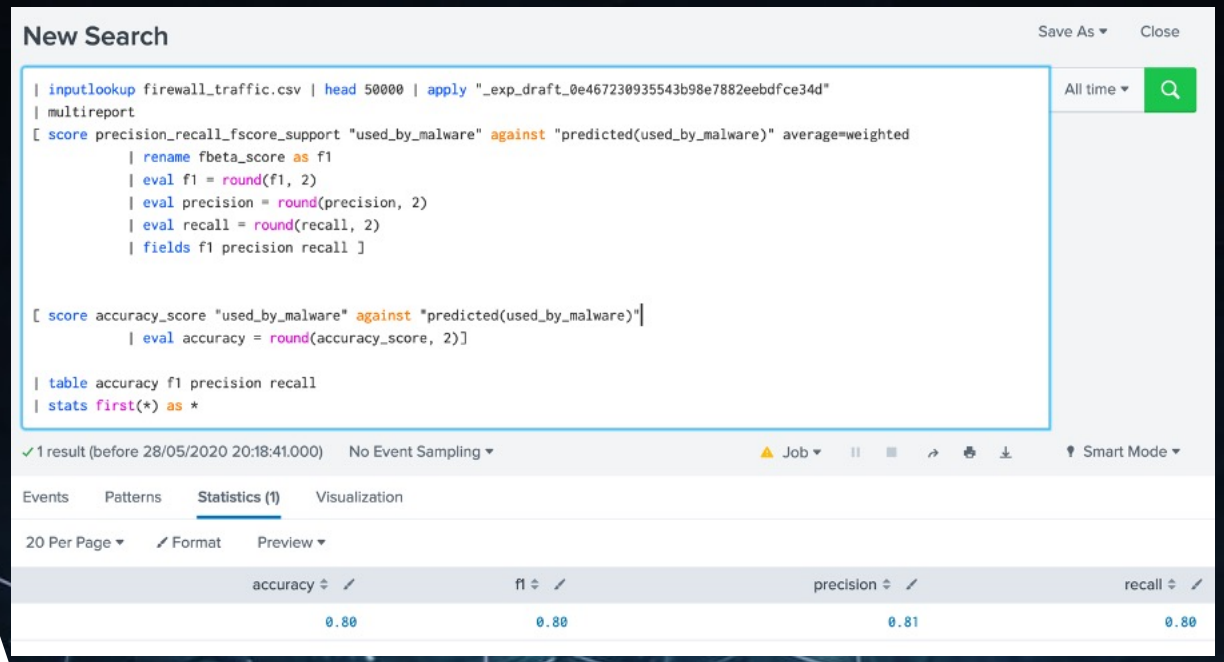
Predicted actual	Predicted no	Predicted yes
no	4818 (80.7%)	1152 (19.3%)
yes	1486 (16.4%)	7580 (83.6%)

Site

&cyber data

Logistic Regression





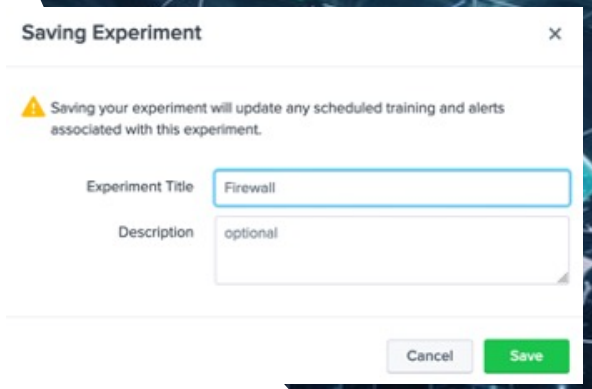

```
| inputlookup firewall_traffic.csv | head 50000 | fit LogisticRegression fit_intercept=true "used_by_malware" from
    "bytes_received" "bytes_sent" "dest_port" "dst_ip" "has_known_vulnerability" "packets_received" "packets_sent" "receive_time"
    "serial_number" "session_id" "src_ip" "src_port" into "_exp_draft_0e467230935543b98e7882eebdfce34d"
```

```
| inputlookup firewall_traffic.csv | head 50000
| fit LogisticRegression fit_intercept=true "used_by_malware" from "bytes_received" "bytes_sent" "dest_port" "dst_ip"
    "has_known_vulnerability" "packets_received" "packets_sent" "receive_time" "serial_number" "session_id" "src_ip" "src_port"
    into "_exp_draft_0e467230935543b98e7882eebdfce34d"
```

```
| inputlookup firewall_traffic.csv | head 50000
| fit LogisticRegression fit_intercept=true "used_by_malware" from "bytes_received" "bytes_sent" "dest_port" "dst_ip"
    "has_known_vulnerability" "packets_received" "packets_sent" "receive_time" "serial_number" "session_id" "src_ip" "src_port"
    into "_exp_draft_0e467230935543b98e7882eebdfce34d"
| multireport
[ score precision_recall_fscore_support "used_by_malware" against "predicted(used_by_malware)" average=weighted
            | rename fbeta_score as f1
            | eval f1 = round(f1, 2)
            | eval precision = round(precision, 2)
            | eval recall = round(recall, 2)
            | fields f1 precision recall ]

[ score accuracy_score "used_by_malware" against "predicted(used_by_malware)"
            | eval accuracy = round(accuracy_score, 2)]

| table accuracy f1 precision recall
| stats first(*) as *
```



Site

SVM
(Support Vector Machine)

Algorithm

LogisticRegression

- ✓ LogisticRegression
- SVM
- RandomForestClassifier
- GaussianNB
- BernoulliNB
- DecisionTreeClassifier

Precision	Recall	Accuracy	F1
0.96	0.96	0.96	0.96

Classification Results (Confusion Matrix)

Predicted actual	Predicted no	Predicted yes
no	5357 (91.7%)	484 (8.3%)
yes	75 (0.8%)	9071 (99.2%)

Algorithm

RandomForestClassifier

- LogisticRegression
- SVM
- ✓ RandomForestClassifier
- GaussianNB
- BernoulliNB
- DecisionTreeClassifier

Precision	Recall	Accuracy	F1
0.99	0.99	0.99	0.99

Classification Results (Confusion Matrix)

Predicted actual	Predicted no	Predicted yes
no	5852 (98.4%)	95 (1.6%)
yes	112 (1.2%)	8923 (98.8%)

Site

&cyber data

cyber & data

"From bits to information"

Race Day

Race Day

Smart Forecasting	Smart Outlier Detection	Smart Clustering	Smart Prediction	Predict Numeric Fields	Predict Categorical Fields	Detect Numeric Outliers	Detect Categorical Outliers	Forecast Time Series	Cluster Numeric Events
1	0	1	0	4	3	1	1	3	3

Create New Experiment

Experiment Type: Predict Categorical Fields

Experiment Title: cars

Description: Optional

Enter a search

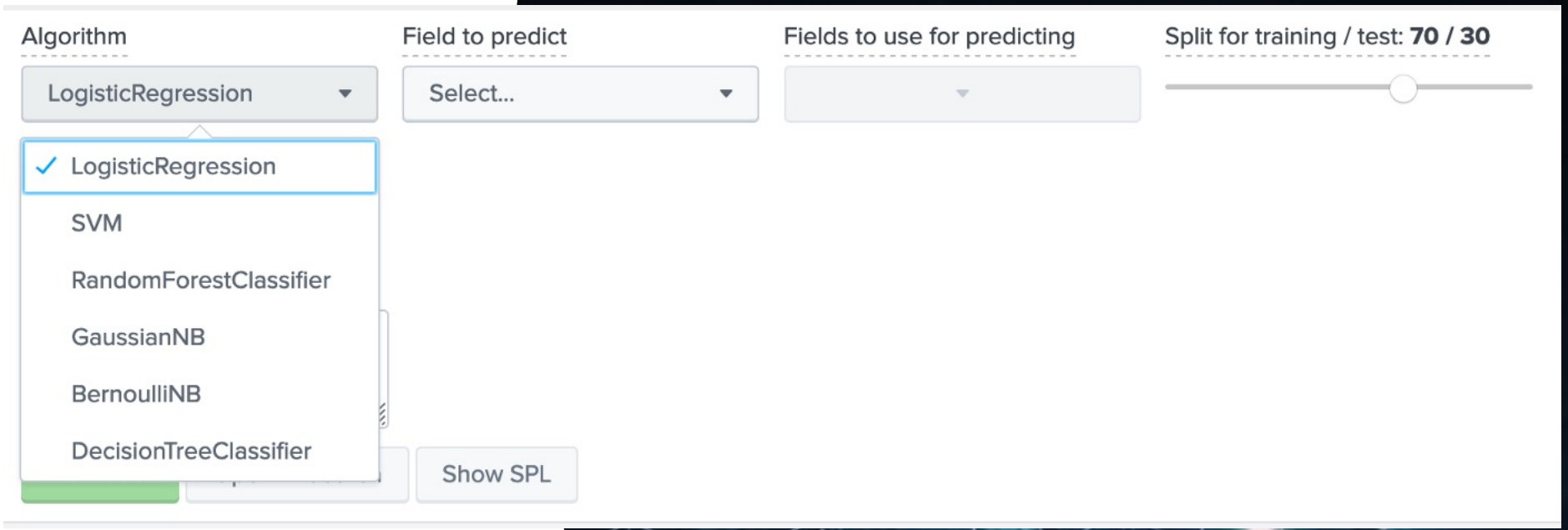
```
| inputlookup track_day.csv
```

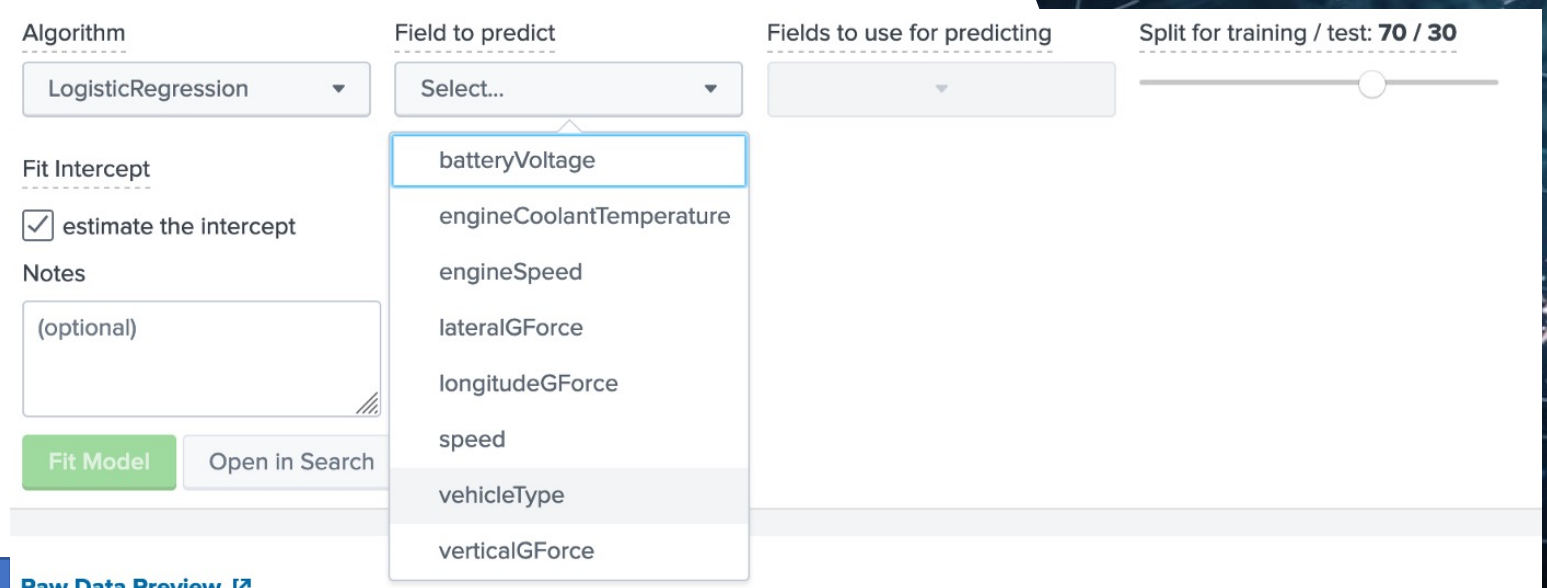
✓ 50,000 results (01/01/1970 00:00:00.000 to 13/06/2020 08:14:30.000)

batteryVoltage	engineCoolantTemperature	engineSpeed	lateralGForce	longitudeGForce	speed	vehicleType	verticalGForce
13.785	93	6060	1.11	0.5	69	2015 Porsche GT3	-2.0
13.937	94	4957	0.56	0.7	56	2013 Audi RS5	0.95
13.827	93	6163	0.71	0.26	70	2015 Porsche GT3	-2.0
14.035	87	2846	0.81	-0.71	47	2011 Ford Mustang GT500	-2.0
13.827	93	6542	0.49	-0.18	76	2015 Porsche GT3	-2.0
14.624	105	4425	0.32	0.05	100	2014 Chevrolet Corvette	-0.17
13.827	93	7763	0.24	-0.18	91	2015 Porsche GT3	-2.0
13.827	93	6365	0.38	-0.2	95	2015 Porsche GT3	-2.0
13.827	93	6713	0.04	0.13	100	2015 Porsche GT3	-2.0
14.262	-10	1508	0.65	-0.01	6	2013 Audi RS5	0.77

Site

Race Day





Site

Race Day

Algorithm: LogisticRegression

Field to predict: vehicleType

Fields to use for predicting: batteryVoltage, engi... (7)

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

Fit Intercept

- [x] estimate the intercept

Notes

(optional)

Fit Model | Open in Search | Show SPL

filter

Select All | Clear All

- [x] batteryVoltage
- [x] engineCoolantTemperature
- [x] engineSpeed
- [x] lateralGForce
- [x] longitudeGForce
- [x] speed

Raw Data Preview

vehicleType	predicted(vehicleType)	batteryVoltage	engineCoolantTemperature	engineSpeed	lateralGForce	longitudeGForce	speed	verticalGForce
2011 Ford Mustang GT500	2011 Ford Mustang GT500	14.035	87.0	2846	0.81	-0.71	47	-2.0
2015 Porsche GT3	2015 Porsche GT3	13.827	93.0	6365	0.38	-0.2	95	-2.0
2015 Porsche GT3	2015 Porsche GT3	13.808	92.0	4303	-0.27	0.54	60	-2.0
2015 Porsche GT3	2015 Porsche GT3	13.808	92.0	4472	-0.64	0.25	47	-2.0
2015 Porsche GT3	2015 Porsche GT3	13.808	92.0	4472	-0.89	-0.31	47	-2.0
2015 Porsche GT3	2015 Porsche GT3	13.808	92.0	3499	-0.78	-0.42	39	-2.0
2015 Porsche GT3	2015 Porsche GT3	13.716	93.0	3810	-0.81	-0.46	44	-2.0
2015 Porsche GT3	2015 Porsche GT3	14.22	92.0	7344	0.0	0.21	107	-2.0
2015 Porsche GT3	2015 Porsche GT3	13.772	92.0	5167	0.32	0.25	75	-2.0
2015 Porsche GT3	2015 Porsche GT3	13.772	92.0	5091	-0.92	-0.53	75	-2.0

« Prev 1 2 3 4 5 6 7 8 9 10 Next »

Open in Search | Show SPL

Precision	Recall	Accuracy	F1
0.97	0.97	0.97	0.97

Classification Results (Confusion Matrix)

Predicted actual	Predicted 2008 BMW M3	Predicted 2011 Ferrari 458	Predicted 2011 Ford M
2008 BMW M3	2679 (99.9%)	0 (0%)	
2011 Ferrari 458	0 (0%)	2025 (97.1%)	
2011 Ford Mustang GT500	0 (0%)	2 (0.1%)	
2013 Audi RS5	50 (2.2%)	33 (1.4%)	
2014 Chevrolet Corvette	17 (0.5%)	7 (0.2%)	
2015 Porsche GT3	0 (0%)	0 (0%)	

Site

Race Day

Prediction Results

vehicleType	predicted(vehicleType)	batteryVoltage	engineCoolantTemperature	engineSpeed	lateralGForce	longitudeGForce	speed	verticalGForce
2011 Ford Mustang GT500	2011 Ford Mustang GT500	14.035	87.0	2846	0.81	-0.71	47	-2.0
2015 Porsche GT3	2015 Porsche GT3	13.827	93.0	6542	0.49	-0.18	76	-2.0
2014 Chevrolet Corvette	2014 Chevrolet Corvette	14.624	105.0	4425	0.32	0.05	100	-0.17
2015 Porsche GT3	2015 Porsche GT3	13.827	93.0	6365	0.38	-0.2	95	-2.0
2015 Porsche GT3	2015 Porsche GT3	13.827	93.0	6713	0.04	0.13	100	-2.0
2013 Audi RS5	2013 Audi RS5	14.262	-10.0	1508	0.65	-0.01	6	0.77
2015 Porsche GT3	2015 Porsche GT3	13.808	92.0	4472	-0.64	0.25	47	-2.0
2015 Porsche GT3	2015 Porsche GT3	13.808	92.0	4472	-0.89	-0.31	47	-2.0
2015 Porsche GT3	2015 Porsche GT3	13.808	92.0	3536	-0.86	-0.39	40	-2.0
2015 Porsche GT3	2015 Porsche GT3	13.716	93.0	3663	-0.79	-0.54	42	-2.0

« Prev 1 2 3 4 5 6 7 8 9 10 Next »

Open in Search | Show SPL

Precision	Recall	Accuracy	F1
1.00	1.00	1.00	1.00

Classification Results (Confusion Matrix)

Predicted actual	Predicted 2008 BMW M3	Predicted 2011 Ferrari 458	Predicted 2011 Ford M
2008 BMW M3	2654 (100%)	0 (0%)	
2011 Ferrari 458	0 (0%)	2065 (100%)	
2011 Ford Mustang GT500	0 (0%)	0 (0%)	
2013 Audi RS5	0 (0%)	0 (0%)	
2014 Chevrolet Corvette	0 (0%)	0 (0%)	
2015 Porsche GT3	0 (0%)	0 (0%)	

Site

&cyber data

cyber & data

"From bits to information"

Predicting (Predicting Numeric Fields)

Numeric Prediction

```
| inputlookup df.csv
```

All time

✓ 51 results (before 29/05/2020 20:20:24.000) No Event Sampling

Job Smart Mode

Events Patterns Statistics (51) Visualization

20 Per Page Format Preview

< Prev 1 2 3 Next >

Av Income	CO2 emissions	Cancer DR	Drug Poisoning DR	Heart Disease DR	Homicide DR	Infant MR	Motor Vech DR	Population	Population density	RegionState	Smoking deaths per 100K	Stroke DR	Suicide DR	Unemployment
41415	26.85	177.6	15.2	224	8.1	8.6	16.9	4858979	95.9	Alabama	532	48.3	14.5	7.2
69825	52.71	164.2	16.8	146.6	4.7	5.8	9.9	738432	1.3	Alaska	215	32.3	22.1	6.9
46709	14.23	142.7	18.2	136.4	5	5.2	11.4	6828065	60.1	Arizona	342	28.3	18	7.7
38758	22.8	183.1	12.6	217.5	7.7	7.8	15.7	2978204	57.2	Arkansas	549	45.4	17.3	7.3
67458	9.18	144.1	11.1	142.2	4.6	4.8	7.9	39144818	251.3	California	340	33.9	10.5	8.9
55387	17.98	136	16.3	130.3	3.3	5.1	9.1	5456574	52.6	Colorado	274	33.4	19.9	6.8
65753	9.84	146.7	17.6	145.6	2.8	4.8	6.9	3590886	741.6	Connecticut	440	26.3	9.8	7.8
57954	14.44	167.3	20.9	168.7	6.6	6.4	12.9	945934	485.3	Delaware	432	38.8	13.2	6.7
65124	5.12	178.6	14.2	207.8	13.7	6.7	3.5	672228	89.5	District of Columbia	377	33.6	7.8	8.5

Experiments

Create New Experiment

Smart Forecasting	Smart Outlier Detection	Smart Clustering	Smart Prediction	Predict Numeric Fields	Predict Categorical Fields	Detect Numeric Outliers	Detect Categorical Outliers	Forecast Time Series	Cluster Numeric Events
0	0	0	0	2	3	0	0	0	0

Create New Experiment

Experiment Type: Predict Numeric Fields

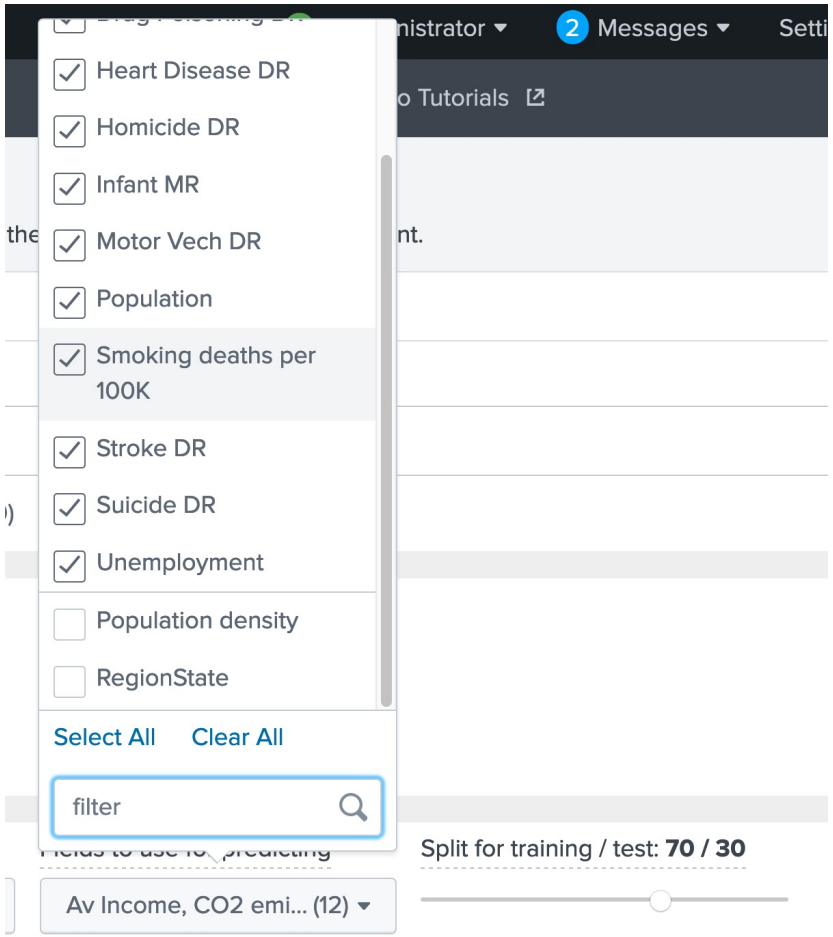
Experiment Title: cancer

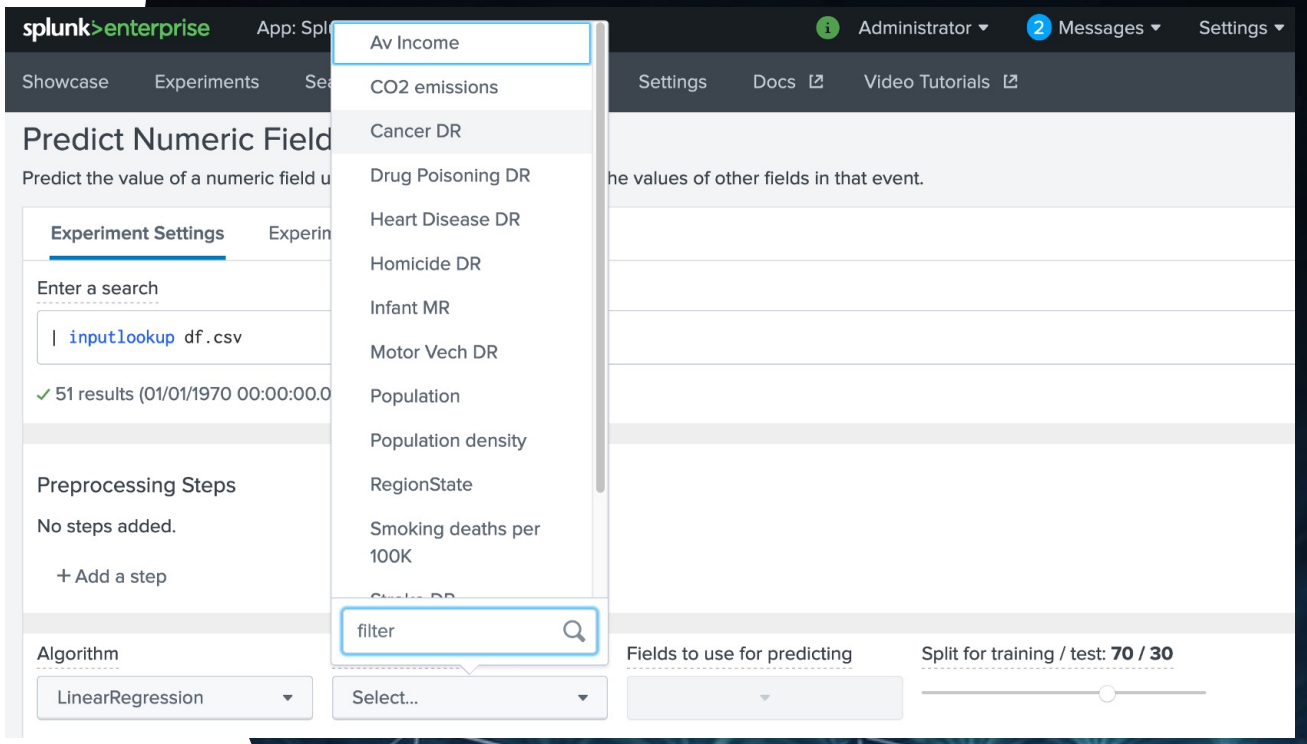
Description: Optional

Site

&cyber data

Numeric Prediction

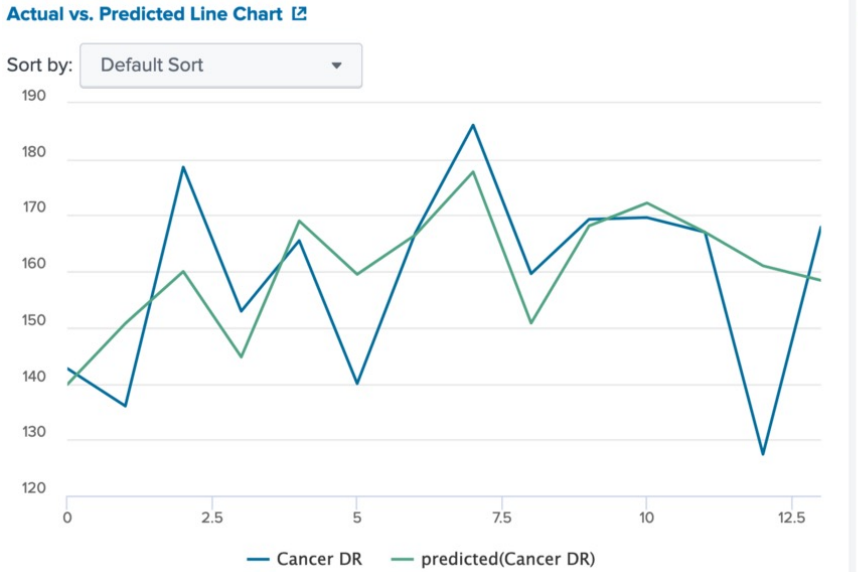


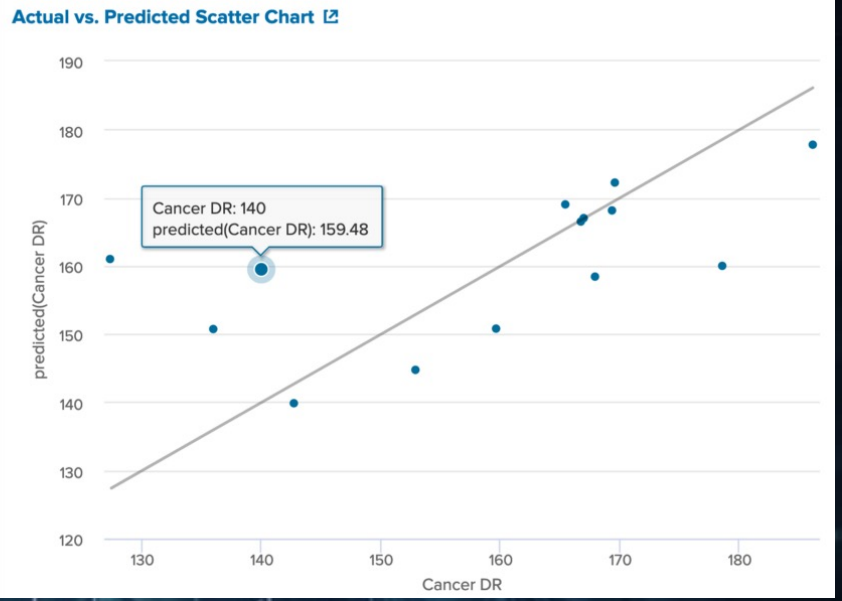


Cancer DR	predicted(Cancer DR)	residual	Av Income	CO2 emissions	Drug Poisoning DR	Heart Disease DR	Homicide DR	Infant MR
142.7	139.85	2.8	46709	14.23	18.2	136.4	5.0	5.2
136.0	150.73	-14.7	55387	17.98	16.3	130.3	3.3	5.1
178.6	159.99	18.6	65124	5.12	14.2	207.8	13.7	6.7
152.9	144.74	8.2	44299	11.9	13.2	151.3	6.2	6.1
165.5	169.00	-3.5	46007	15.69	11.9	179.7	6.6	7.0
140.0	159.48	-19.5	62814	14.01	10.9	136.7	2.2	6.4
166.8	166.49	0.3	48964	25.21	11.7	157.4	3.6	6.5
186.1	177.76	8.3	41734	48.75	16.9	216.3	11.7	8.7
159.6	150.79	8.8	50296	28.37	7.2	143.0	3.4	5.2
169.3	168.12	1.2	43916	12.74	13.8	158.7	5.6	7.0

Site

Numeric Prediction





R^2 Statistic

0.3631

Root Mean Squared Error (RMSE)

13.10

Site

&cyber data

Numeric Pred

New Search

Save As Close

| summary "_exp_draft_3a2667ce044840f89c19bc0ce58d2fef" | table feature *

All time

✓ 13 results (before 29/05/2020 20:51:04.000) No Event Sampling

Job Smart Mode

Events Patterns Statistics (13) Visualization

50 Per Page Format Preview

Algorithm: LinearRegression

Field to predict: Cancer DR

Fields to use for predicting: Av Income, CO2 emi... (12)

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

- ✓ LinearRegression
- RandomForestRegressor
- Lasso
- KernelRidge
- ElasticNet
- Ridge
- DecisionTreeRegressor

Show SPL

coefficient
36.32774588686257
2.073133680918171
1.7728643949462002
1.6872775538375058
0.26266757736959545
0.18733642382044466
0.18365251502561808
0.11860259466942144
0.04225673173770444
0.00013768431401350815
-9.353521050222717e-07
-1.133913720246636
-1.9987270193886013

Popul

Motor

Homic

Actual vs. Predicted Line Chart

Sort by: Default Sort

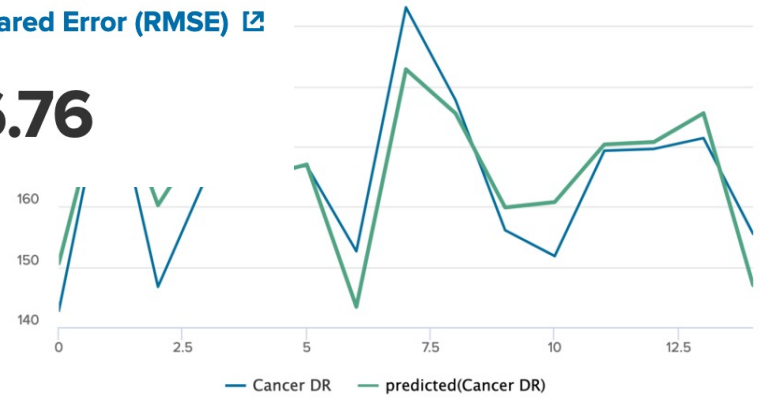
200

R^2 Statistic

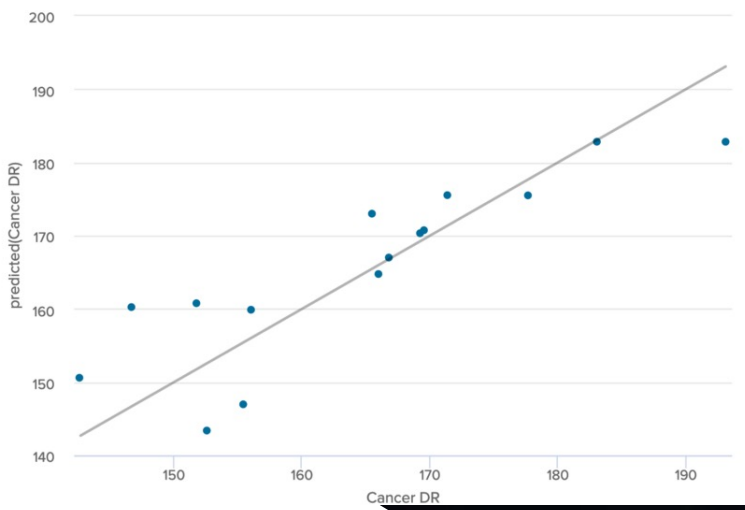
0.7449

Root Mean Squared Error (RMSE)

6.76



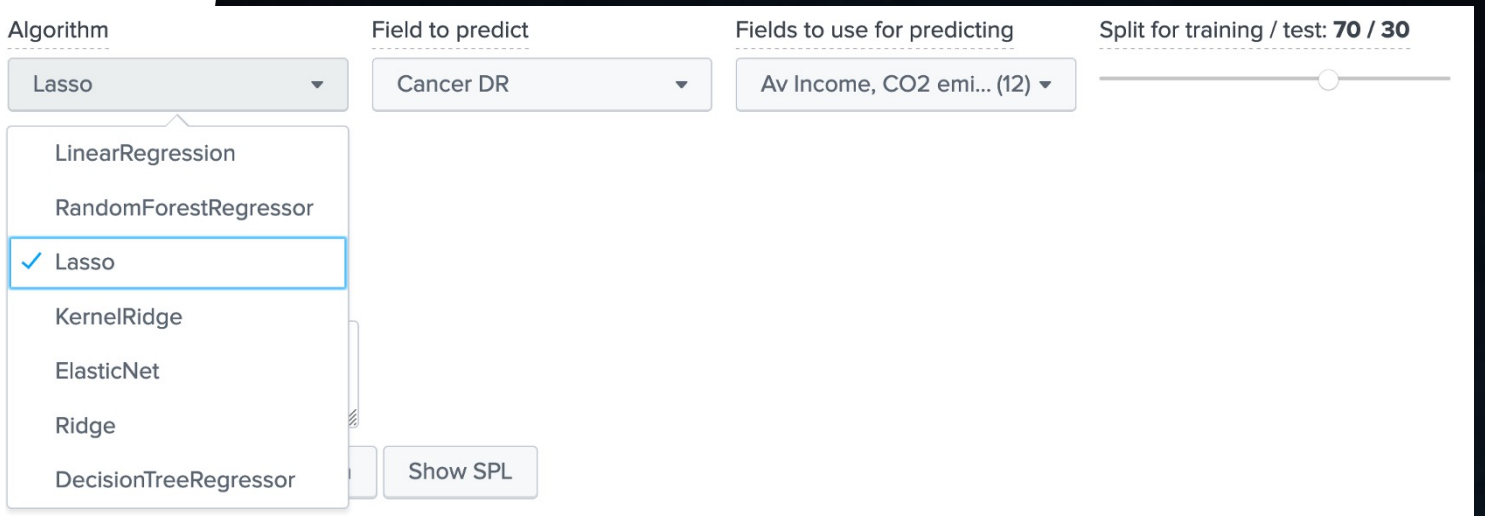
Actual vs. Predicted Scatter Chart

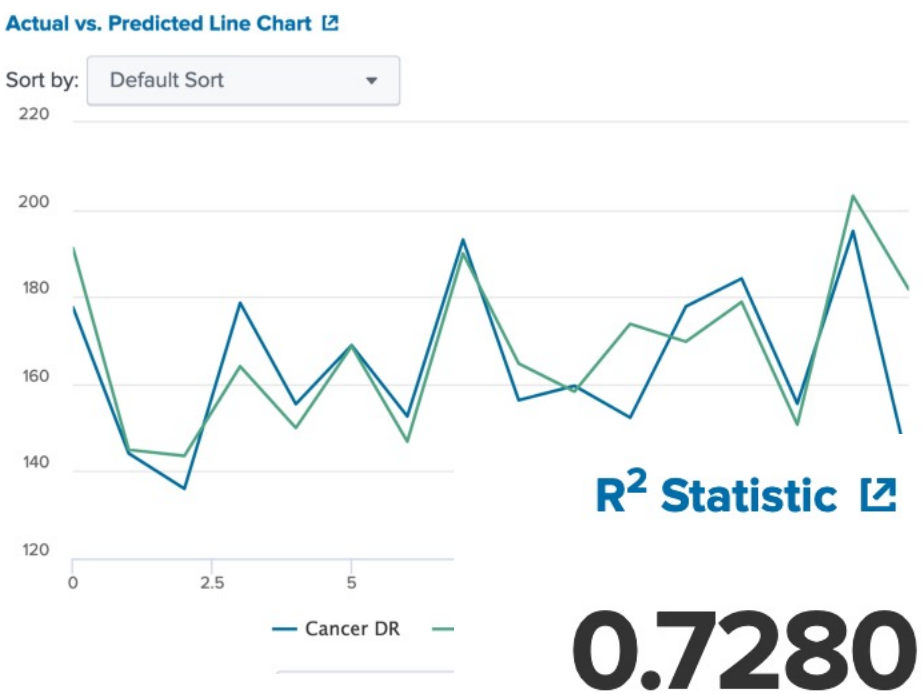


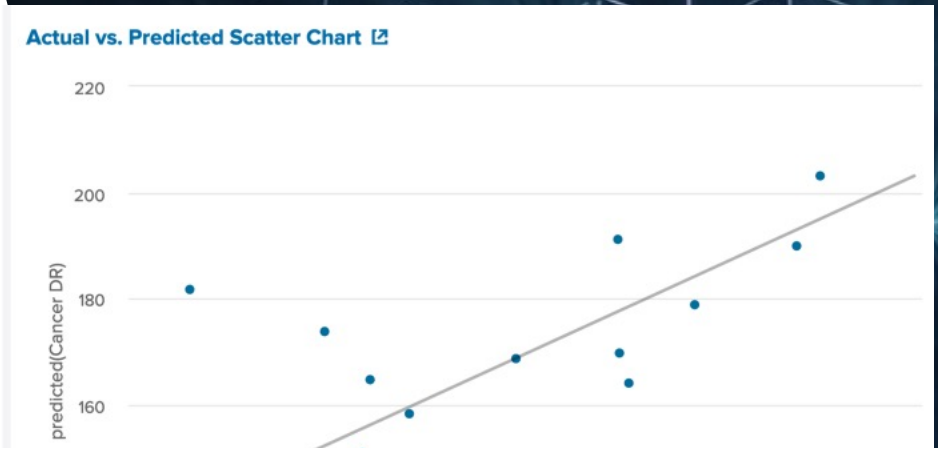
Site

yber
data

Numeric Prediction





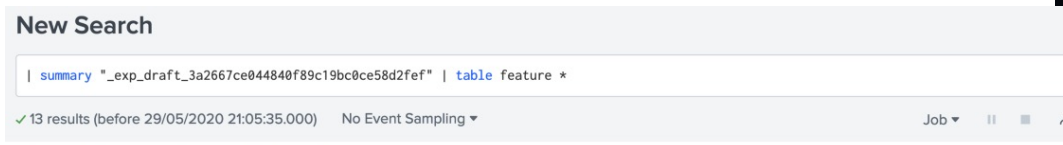


R^2 Statistic

0.7280

Root Mean Squared Error (RMSE)

7.70

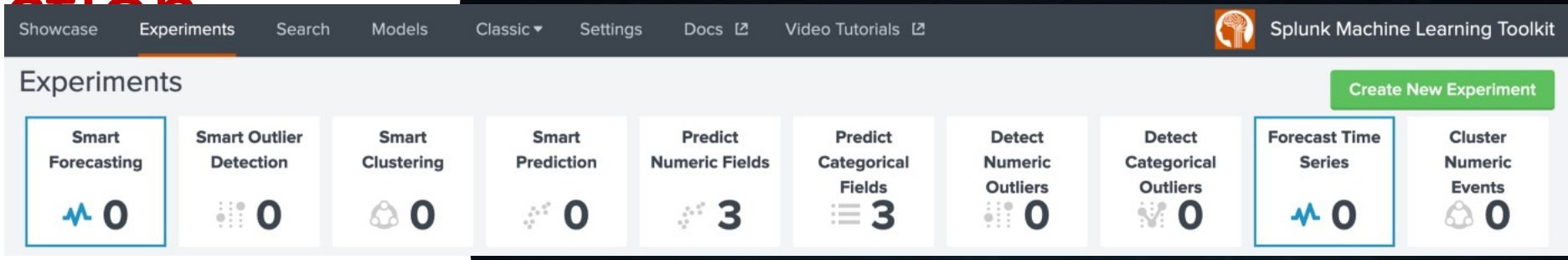


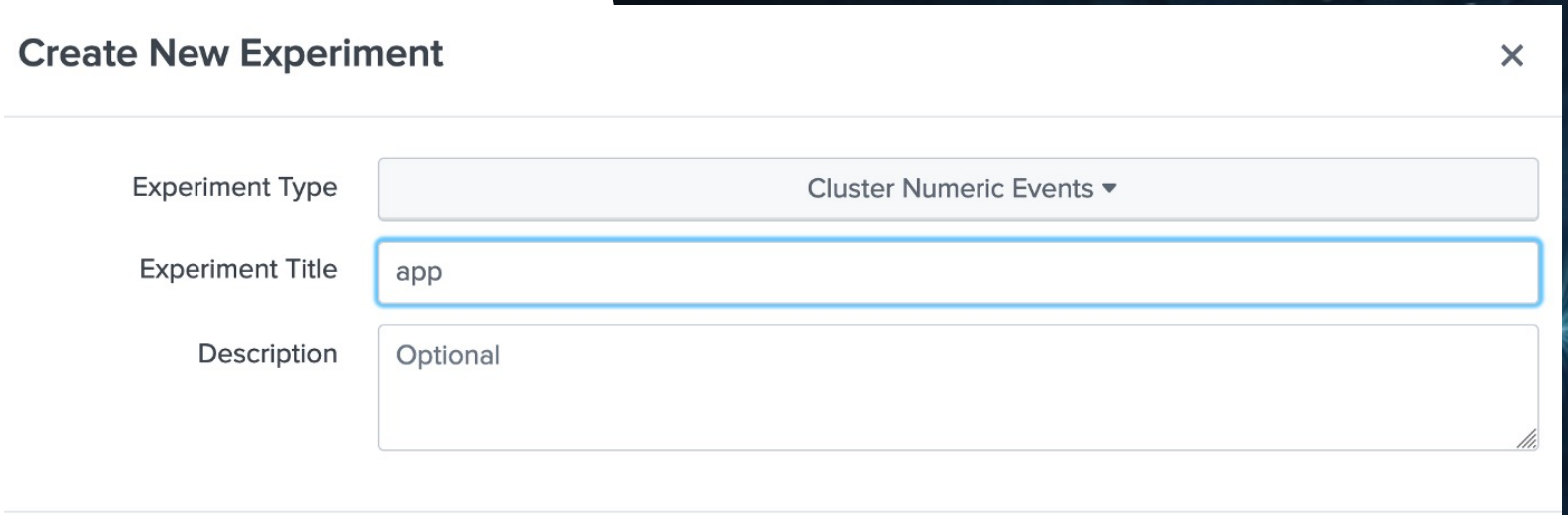
Site

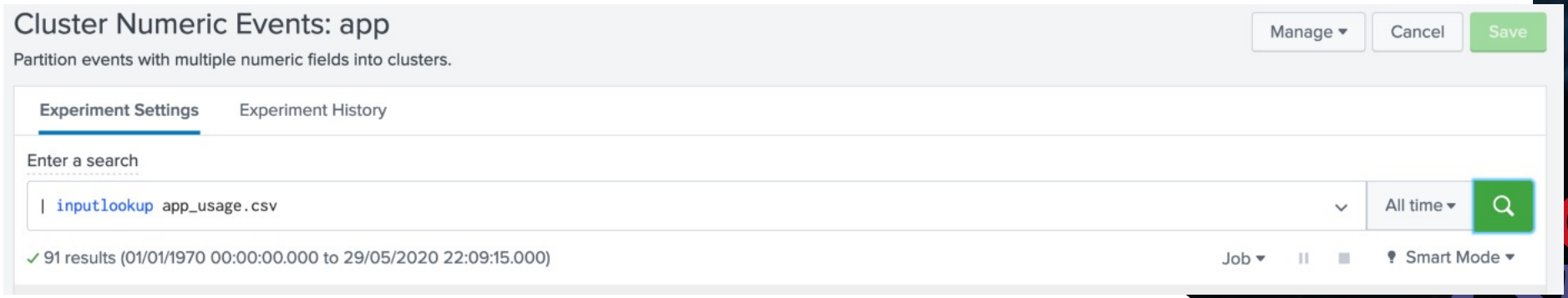
"From bits to information"

Grouping

Numeric Prediction







Site

cyber & data

Numeric Predi

CRM	CloudDrive	ERP	Expenses		HR2	ITOps	OTHER	Recruiting	RemoteAccess	Webmail	_time
49	99	17	38	0	0	18	144	33	283	141	2015-06-06
107	148	28	54	0	0	38	188	30	430	213	2015-06-07
639	796	221	216	0	0	133	1175	297	732	579	2015-06-08
653	767	203	191	0	0	139	1475	308	738	549	2015-06-09
670	738	196	140	0	0	128	1111	305	781	678	2015-06-10
562	672	218	173	0	0	110	994	313	663	843	2015-06-11
547	537	148	174	0	0	81	977	252	631	588	2015-06-12
51	108	8	40	0	0	13	362	27	235	148	2015-06-13
									298	191	2015-06-14
									825	670	2015-06-15
									732	708	2015-06-16
									616	539	2015-06-17
									624	625	2015-06-18
											2015-06-19

Algorithm

K-means

Fields to use for clustering

K (# of centroids)

2

- ✓ K-means
- DBSCAN
- Birch
- Spectral Clustering

Show SPL

Showcase Experiments Se... Settings Docs

Cluster Numeric Ever...

Partition events with multiple numeri...

Experiment Settings Experin...

Enter a search

| inputlookup app_usage.csv

✓ 91 results (01/01/1970 00:00:00.0...

Preprocessing Steps

No steps added.

+ Add a step

- ✓ CRM
- ✓ CloudDrive
- ✓ ERP
- ✓ Expenses
- ✓ HR1
- ✓ HR2
- ✓ ITOps
- ✓ OTHER
- ✓ Recruiting
- ✓ RemoteAccess
- ✓ Webmail

Select All Clear All

filter

Algorithm

K-means

CRM, CloudDrive, ER... (11)

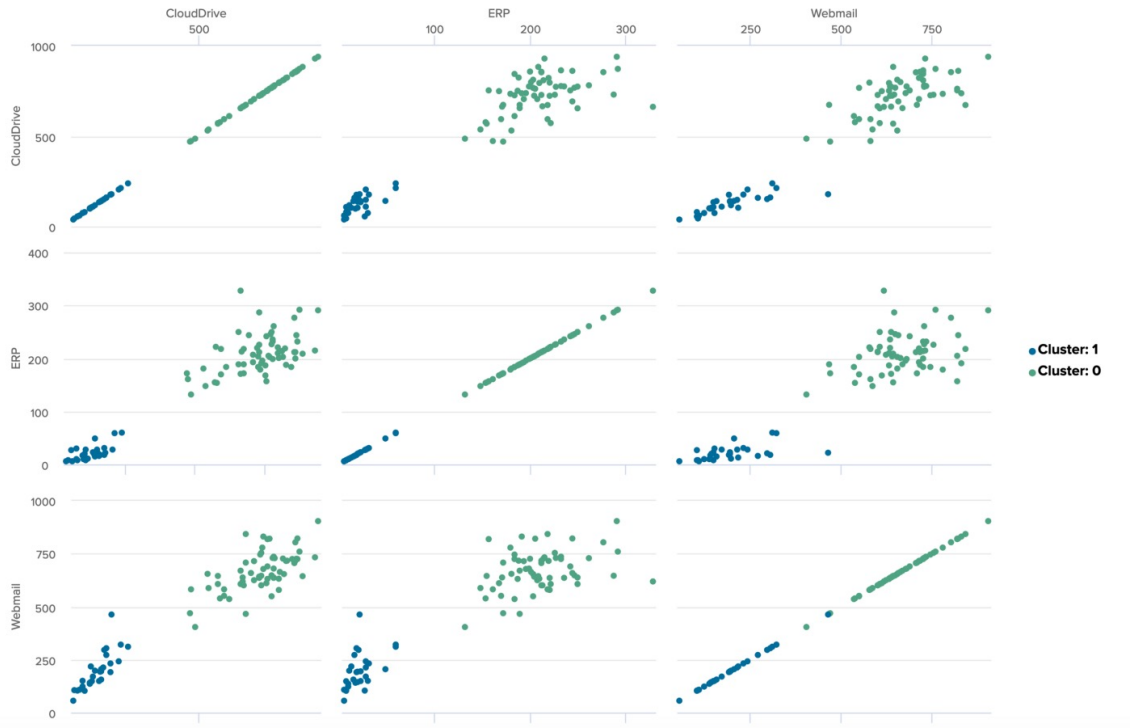
K (# of centroids)

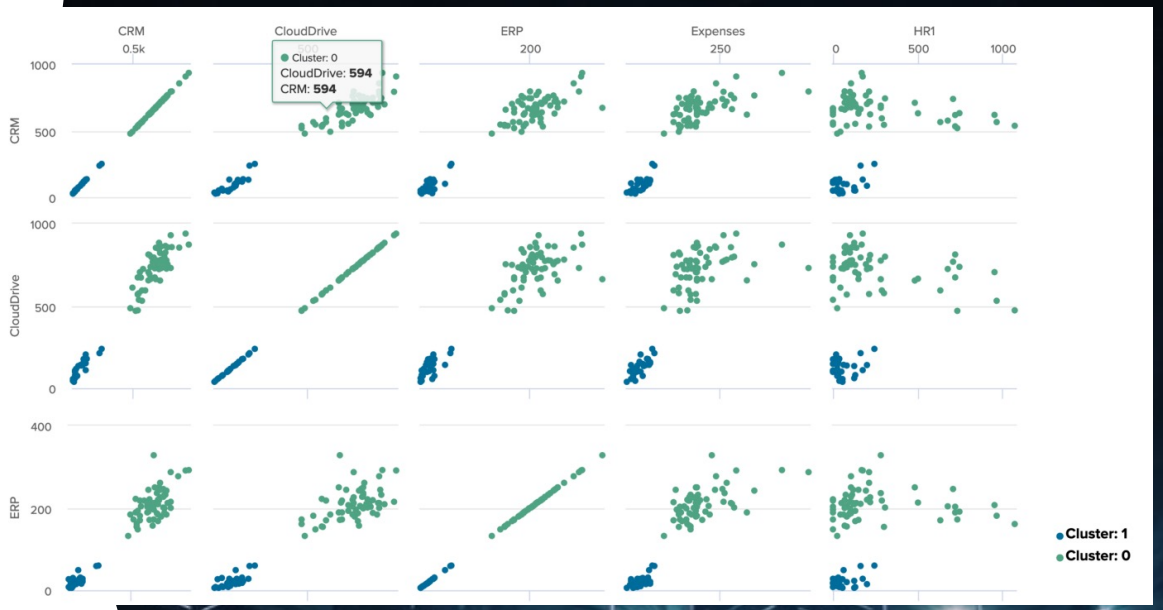
2

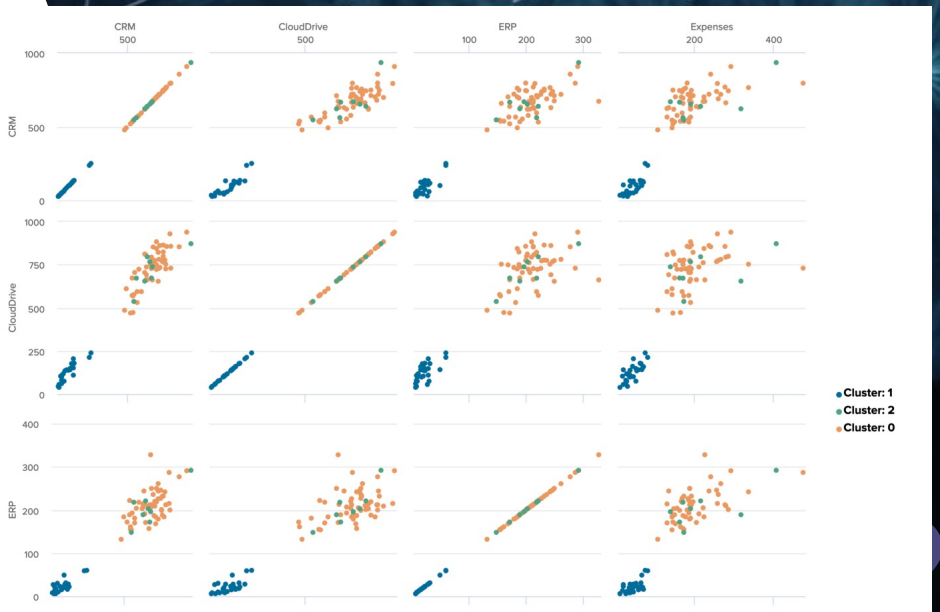
Site

cyber & data

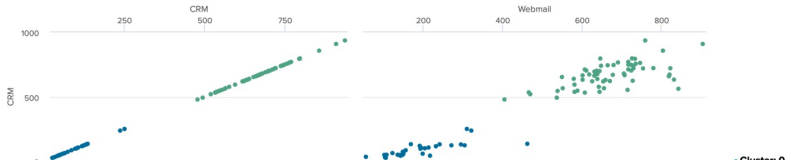
Numeric Prediction







Site



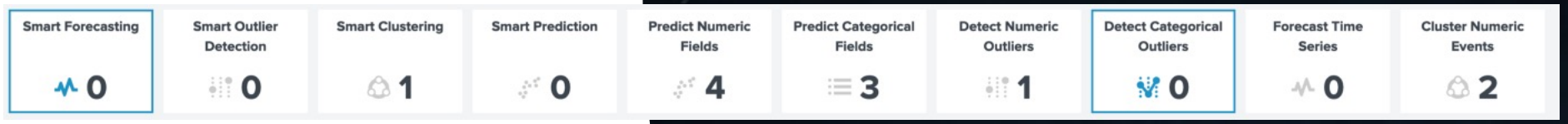
cyber
&data

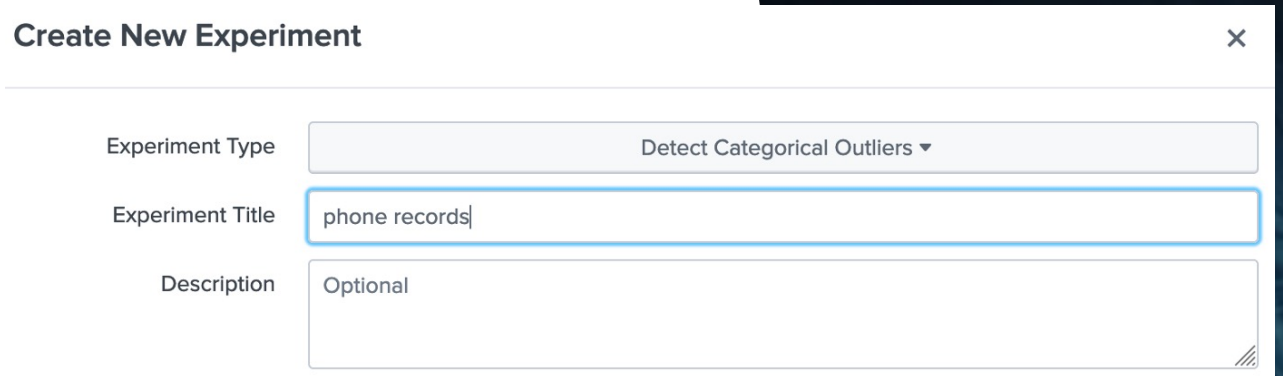
cyber & data

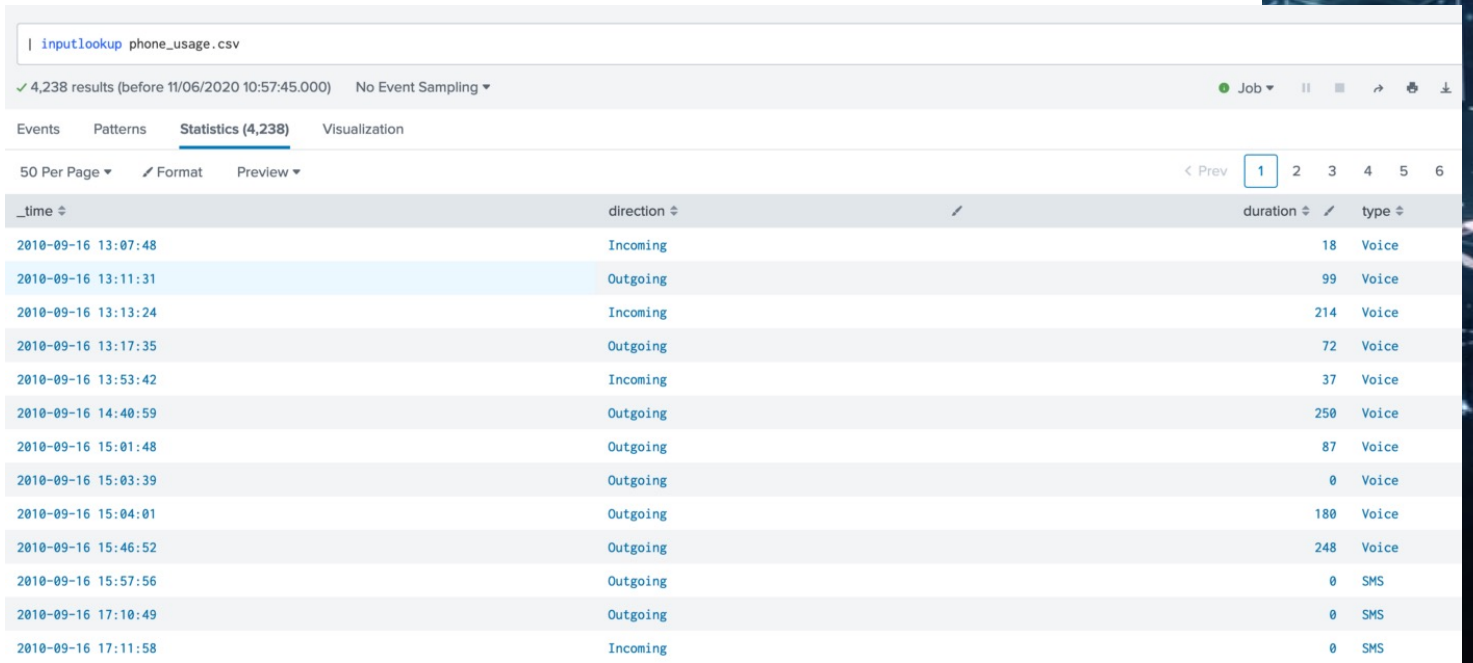
"From bits to information"

Outliers

Detect Categorical Outliner



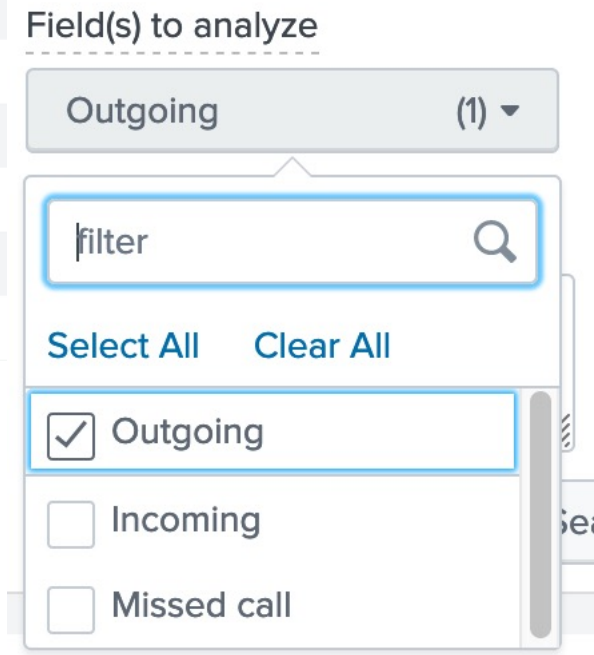




Site

Detect Categorical Outliner

_time	Incoming	Missed call	Outgoing
2010-09-16 12:00	3	1	8
2010-09-17 00:00	0	1	4
2010-09-17 12:00	2	1	8
2010-09-18 00:00			2
2010-09-18 12:00			1
2010-09-19 00:00			0
2010-09-19 12:00			3
2010-09-20 00:00			1
2010-09-20 12:00			



Outlier(s)

2

Outlier(s)

Open in Search | Show SPL

Total Event(s)

285

Total Event(s)

Open in Search | Show SPL

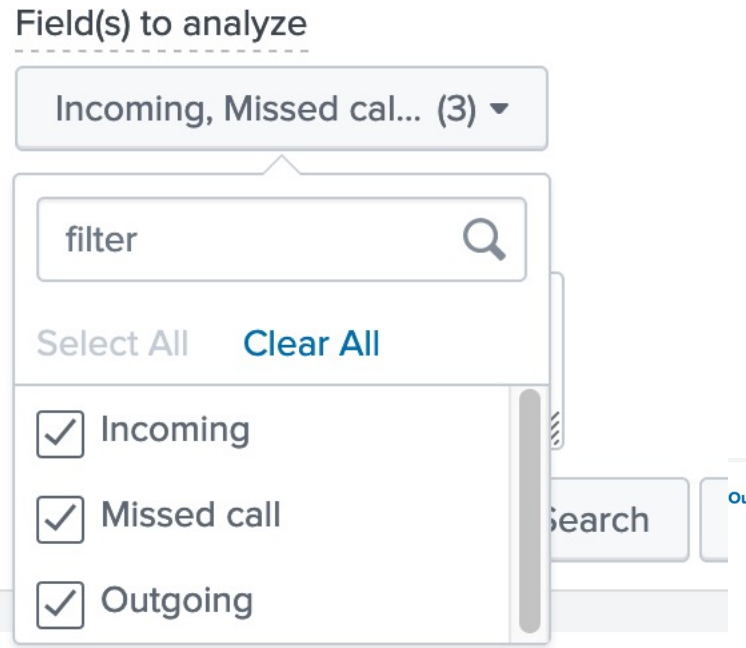
Data and Outliers

Outgoing	probable_cause	isOutlier
17	Outgoing	1
31	Outgoing	1
8		0
4		0
8		0
2		0
1		0
0		0
3		0
1		0

Site

yber
data

Detect Categorical Outliner



Outlier(s)

1

Outlier(s)

Open in Search | Show SPL

Total Event(s)

285

Total Event(s)

Open in Search | Show SPL

Data and Outliers

Incoming	Missed call	Outgoing	probable_cause	isOutlier
9	8	31	Missed call	1
3	1	8		0
0	1	4		0
2	1	8		0
0	1	2		0
0	0	1		0
0	0	0		0
3	0	3		0
1	0	1		0
0	0	3		0

Site

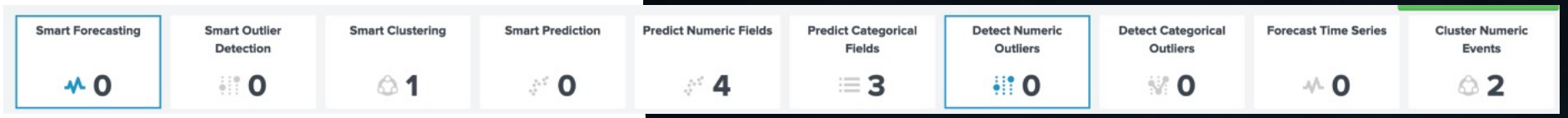
cyber data

cyber & data

"From bits to information"

Outliers (Numeric Outlier)

Detect Numerical Outliner



Create New Experiment

Experiment Type: Detect Numeric Outliers

Experiment Title: Server response

Description: Optional

_time	hoststate	pl	rta	rtmax	rtmin	src_host	timestamp
2015-02-18 14:17:26	UP	0%	1.056	1.191	0.909	host40	1424297846
2015-02-18 14:18:27	UP	0%	1.128	1.275	0.965	host40	1424297907
2015-02-18 14:19:28	UP	0%	1.088	1.254	1.005	host40	1424297968
2015-02-18 14:20:34	UP	0%	1.088	1.499	0.932	host40	1424298034
2015-02-18 14:21:35	UP	0%	1.773	2.182	1.115	host40	1424298095
2015-02-18 14:22:39	UP	0%	1.642	3.816	0.862	host40	1424298159
2015-02-18 14:23:41	UP	0%	2.597	5.238	1.421	host40	1424298221
2015-02-18 14:24:42	UP	0%	1.529	2.414	1.191	host40	1424298282
2015-02-18 14:25:43	UP	0%	2.001	5.933	0.936	host40	1424298343
2015-02-18 14:26:44	UP	0%	2.217	5.081	0.954	host40	1424298404
2015-02-18 14:27:45	UP	0%	1.516	2.390	1.008	host40	1424298465
2015-02-18 14:28:46	UP	0%	0.992	1.220	0.852	host40	1424298526
2015-02-18 14:29:47	UP	0%	1.204	1.333	1.048	host40	1424298587

Site

cyber data

Detect Numerical Outliner

```
| inputlookup hostperf.csv | eval _time=strptime(_time, "%Y-%m-%dT%H:%M:%S.%3Q%z") | timechart span=10m max(rtmax) as responsetime | head 1000
```

✓ 1,000 results (01/01/1970 00:00:00.000 to 10/06/2020 21:56:48.000)

Field to analyze	Threshold method	Threshold multiplier	Sliding window (# of values)		Fields to split by
Select...	Standard Deviation	2	(optional)	Include current point	

responsetime

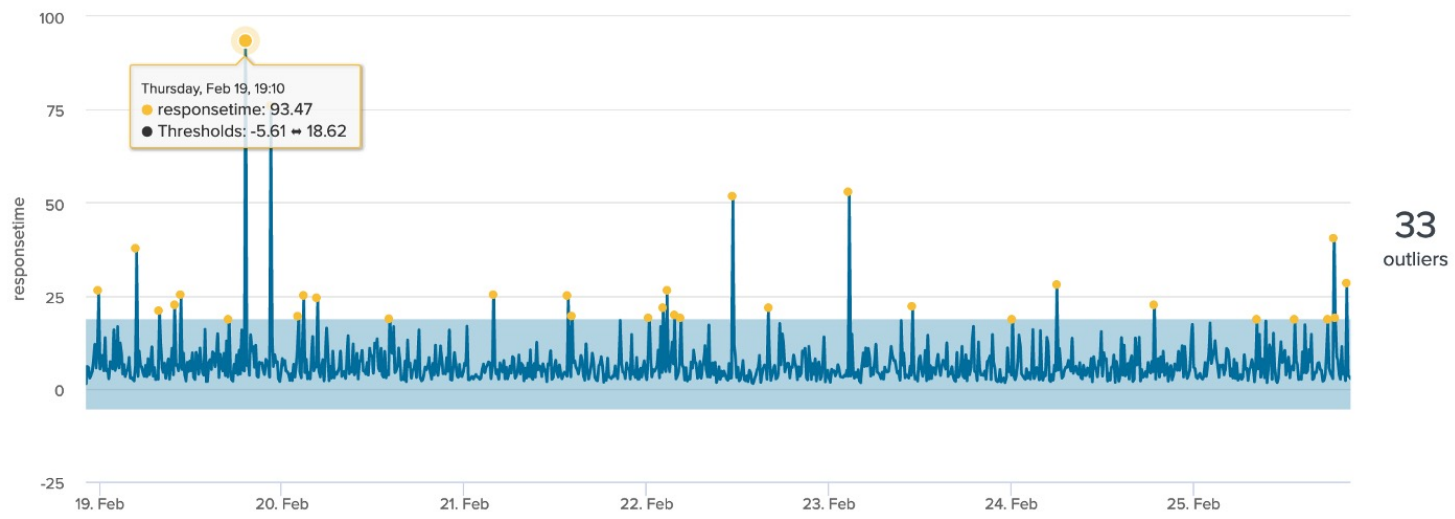
Could not create search.

Field to analyze	Threshold method	Threshold multiplier	Sliding window (# of values)		Fields to split by
responsetime	Standard Deviation	2	(optional)	Include current point	

Notes

(optional)

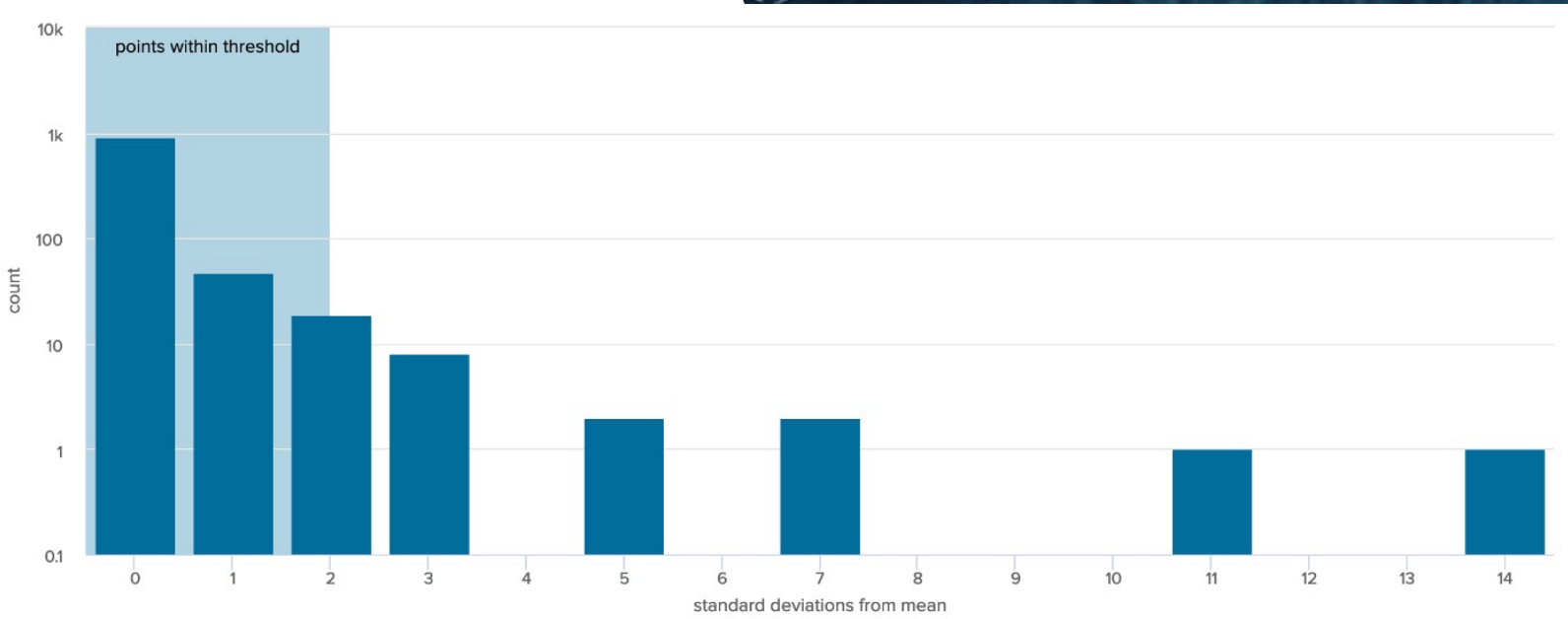
- ✓ Standard Deviation
- Median Absolute Deviation
- Interquartile Range



Site

cyber & data

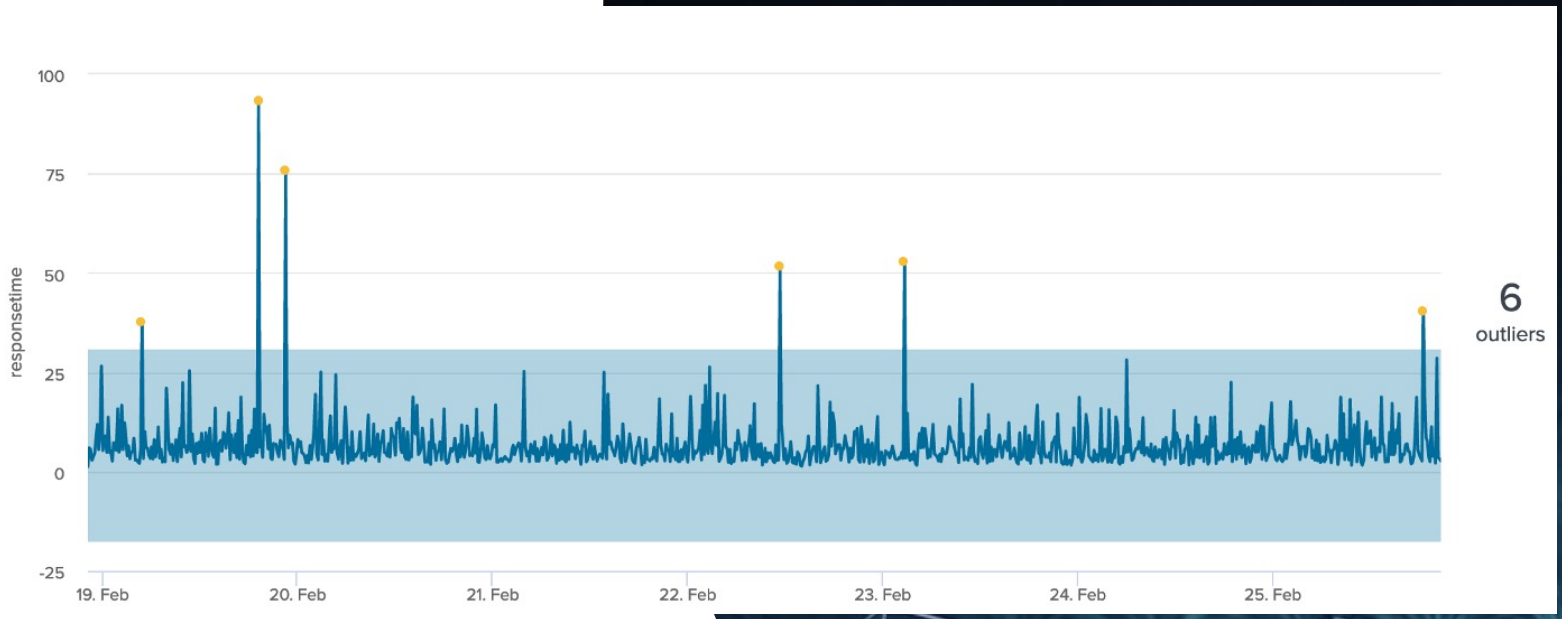
Detect Numerical Outliner



Site

cyber & data

Detect Numerical Outliner



New Search

Save As ▾ Close

```
| inputlookup hostperf.csv
| eval _time=strptime(_time, "%Y-%m-%dT%H:%M:%S.%3Q%z")
| timechart span=10m max(rtmax) as responsetime
| head 1000 | eventstats avg("responsetime") as avg stdev("responsetime") as stdev
| eval lowerBound=(avg-stdev*exact(2)), upperBound=(avg+stdev*exact(2))
| eval isOutlier=if('responsetime' < lowerBound OR 'responsetime' > upperBound, 1, 0)
```

All time ▾

✓ 1,000 results (before 11/06/2020 09:07:03.000) No Event Sampling ▾ Job ▾ Smart Mode ▾

Events Patterns **Statistics (1,000)** Visualization

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

_time	responsetime	avg	isOutlier	lowerBound	stdev	upperBound
2015-02-18 22:10:00	1.275	6.504857000000001	0	-5.607236756629078	6.05604687831454	18.61695075662908
2015-02-18 22:20:00	5.933	6.504857000000001	0	-5.607236756629078	6.05604687831454	18.61695075662908
2015-02-18 22:30:00	5.599	6.504857000000001	0	-5.607236756629078	6.05604687831454	18.61695075662908
2015-02-18 22:40:00	2.839	6.504857000000001	0	-5.607236756629078	6.05604687831454	18.61695075662908
2015-02-18 22:50:00	3.702	6.504857000000001	0	-5.607236756629078	6.05604687831454	18.61695075662908

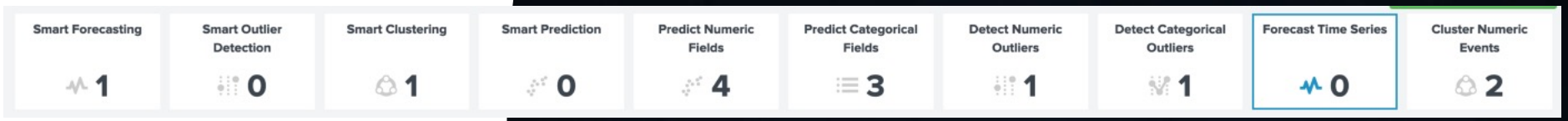
Site

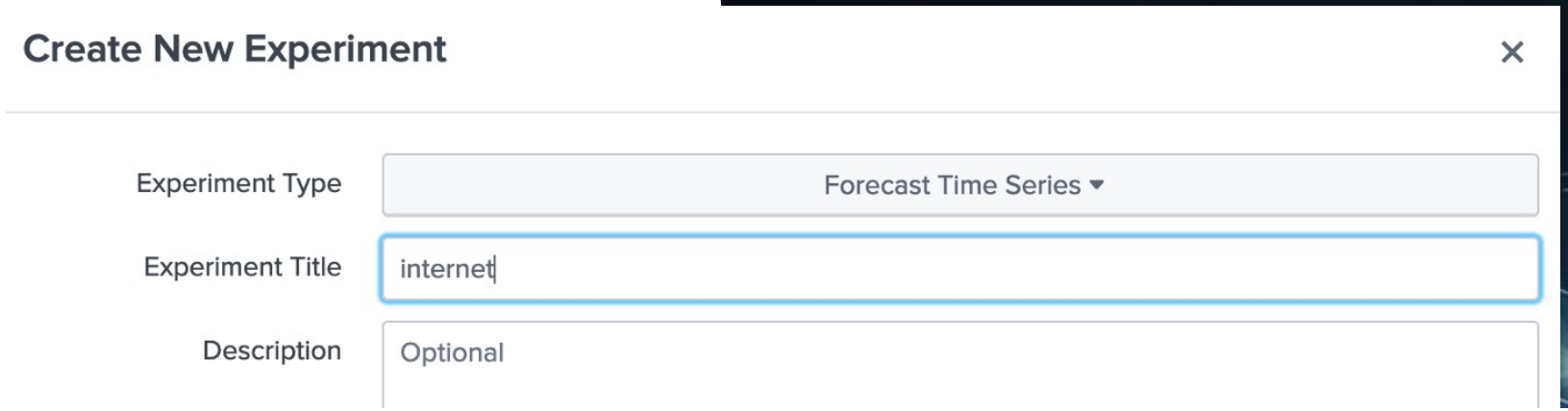
cyber & data

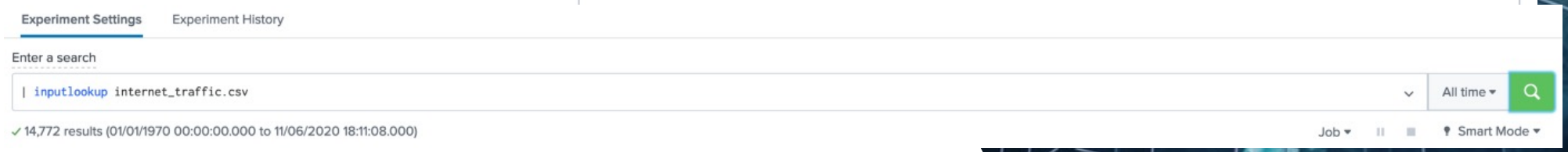
"From bits to information"

Forecasting

Forecasting







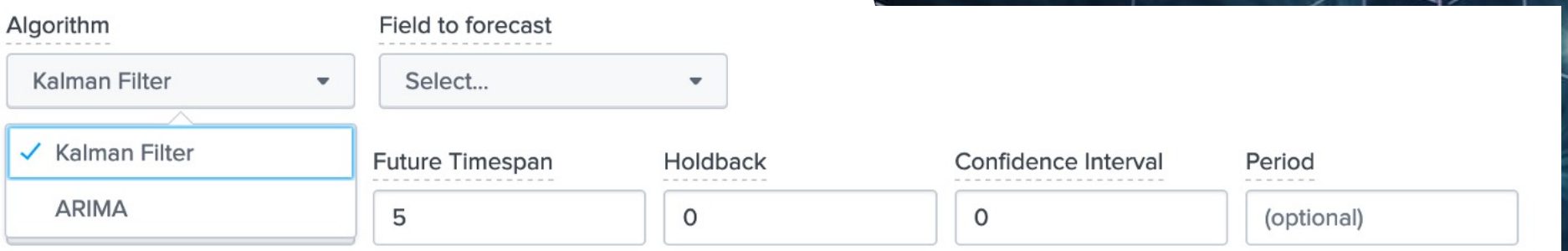
_time	bits_transferred
2005-06-07 14:00:00	3562279127
2005-06-07 14:05:00	3710215571
2005-06-07 14:10:00	3877469703
2005-06-07 14:15:00	3876354871
2005-06-07 14:20:00	4582542581
2005-06-07 14:25:00	5016336869
2005-06-07 14:30:00	5202513642
2005-06-07 14:35:00	5410604985
2005-06-07 14:40:00	5408071320
2005-06-07 14:45:00	5598048880

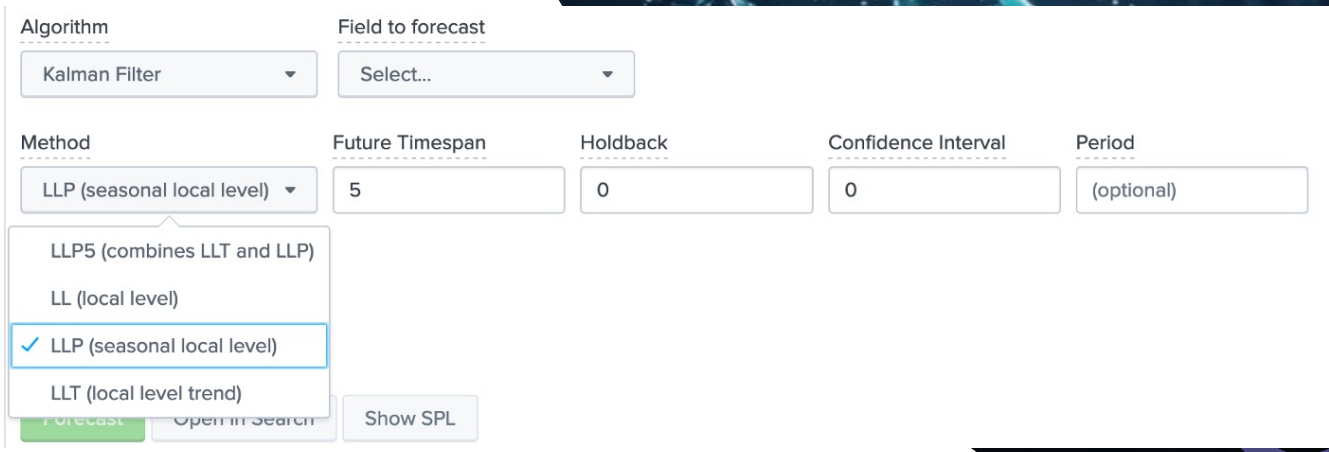
Site

yber
data

Forecasting

_time	bits_transferred
2005-06-07 14:00	5548947933
2005-06-07 16:00	7138793066
2005-06-07 18:00	7516418311
2005-06-07 20:00	7517711314
2005-06-07 22:00	6700932379
2005-06-08 00:00	4366913671





Site

Forecasting

R^2 Statistic	Root Mean Squared Error (RMSE)	Forecast Outliers
0.9992	59,571,591.14	533

R^2 Statistic	Root Mean Squared Error (RMSE)	Forecast Outliers
0.9998	33,136,888.15	0

Site

cyber & data

"From bits to information"

Forecasting (Bluetooth)

Forecasting

_time	address	probe
2006-01-11 20:06:41	2165d8cc825de52667506e2be28f24d0	AxisBoard-5
2006-01-11 20:07:11	2165d8cc825de52667506e2be28f24d0	AxisBoard-5
2006-01-11 20:07:41	2165d8cc825de52667506e2be28f24d0	AxisBoard-5
2006-01-11 20:08:12	2165d8cc825de52667506e2be28f24d0	AxisBoard-5
2006-01-11 20:08:41	2165d8cc825de52667506e2be28f24d0	AxisBoard-5
2006-01-11 20:09:11	2165d8cc825de52667506e2be28f24d0	AxisBoard-5
2006-01-11 20:09:41	2165d8cc825de52667506e2be28f24d0	AxisBoard-5
2006-01-11 20:10:12	2165d8cc825de52667506e2be28f24d0	AxisBoard-5

_time	distinct_addresses
2006-01-11 18:00	5
2006-01-11 21:00	15
2006-01-12 00:00	7
2006-01-12 03:00	8
2006-01-12 06:00	0
2006-01-12 09:00	0
2006-01-12 12:00	0
2006-01-12 15:00	0
2006-01-12 18:00	0
2006-01-12 21:00	0

Site

yber
data

Forecasting

Enter a search

```
| inputlookup bluetooth.csv
| where probe="AxisBoard-5"
| timechart dc(address) as distinct_addresses span=3h
```

All time

✓ 708 results (01/01/1970 00:00:00.000 to 11/06/2020 19:15:52.000)

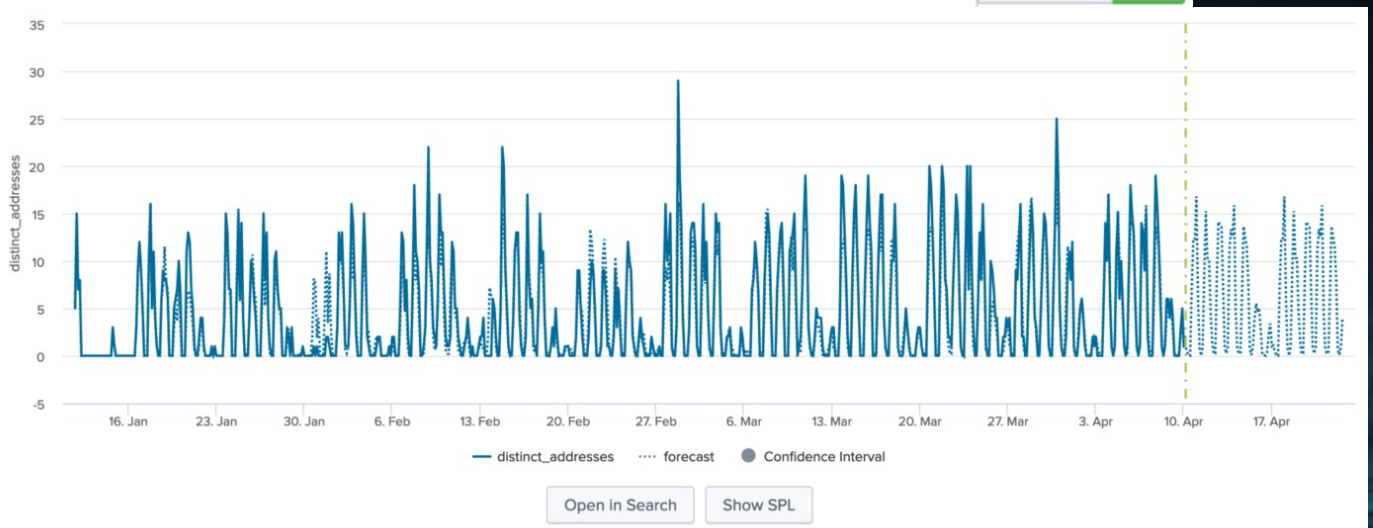
Algorithm: Kalman Filter

Field to forecast: distinct_addresses

Method: LLP (seasonal local level)

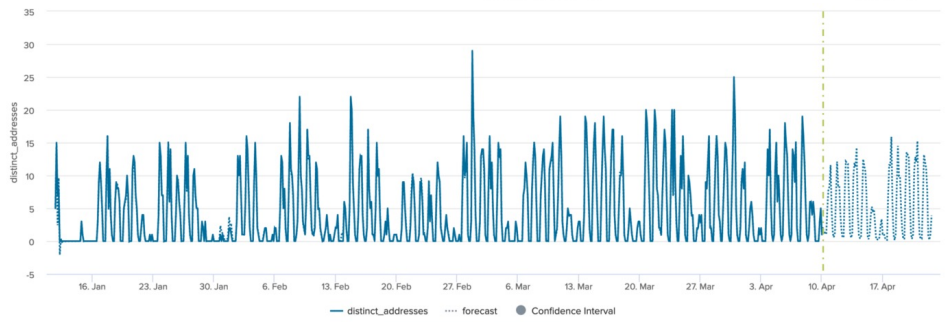
Future Timespan: 100

Holdb: 0



Open in Search | Show SPL

R^2 Statistic	Root Mean Squared Error (RMSE)	Forecast Outliers
0.8784	1.94	535



Open in Search | Show SPL

R^2 Statistic	Root Mean Squared Error (RMSE)	Forecast Outliers
0.9946	0.41	706

Site

&cyber data

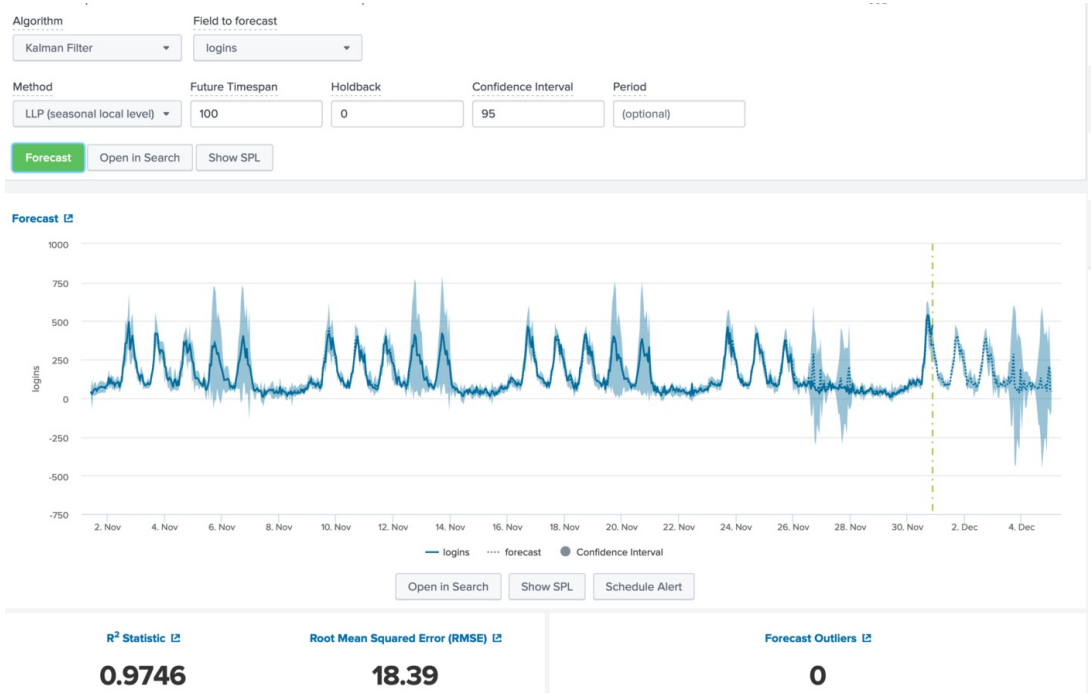
cyber & data

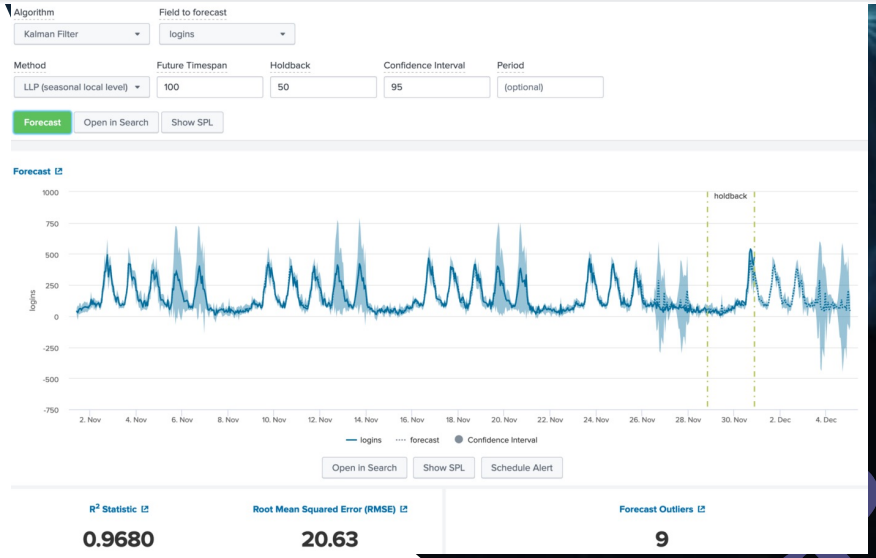
"From bits to information"

Forecasting (Predicting Login)

Forecasting

_time	logins
2015-11-01 10:00	39
2015-11-01 11:00	27
2015-11-01 12:00	49
2015-11-01 13:00	59
2015-11-01 14:00	62
2015-11-01 15:00	49
	63
	68
	73
	75





Site

cyber & data

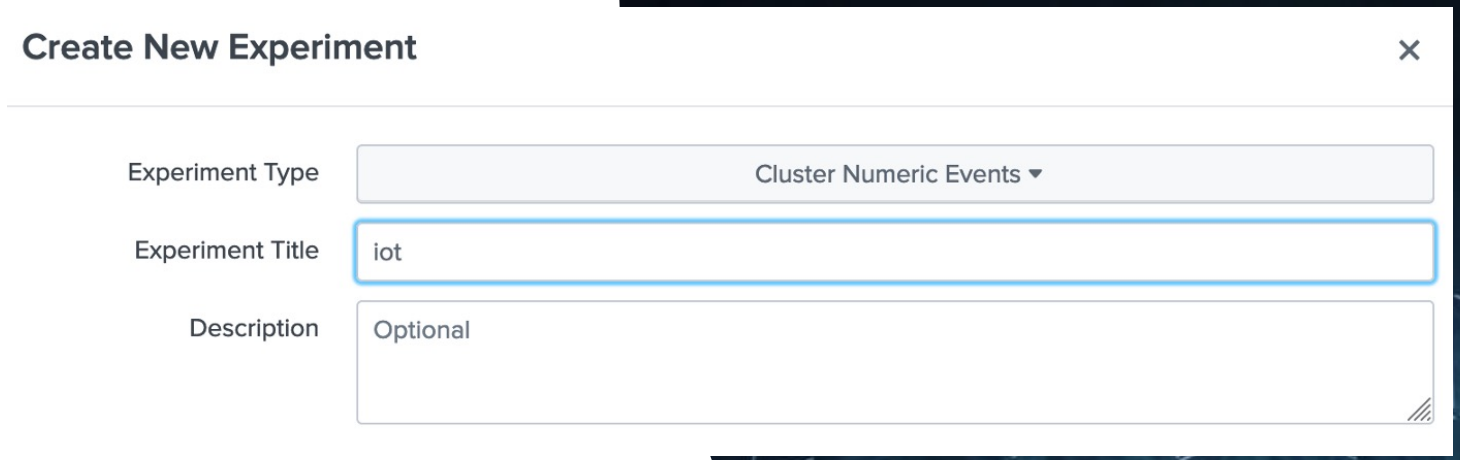
cyber & data

"From bits to information"

Pre-processing

Forecasting

Smart Forecasting	Smart Outlier Detection	Smart Clustering	Smart Prediction	Predict Numeric Fields	Predict Categorical Fields	Detect Numeric Outliers	Detect Categorical Outliers	Forecast Time Series	Cluster Numeric Events
1	0	1	0	4	3	1	1	3	2

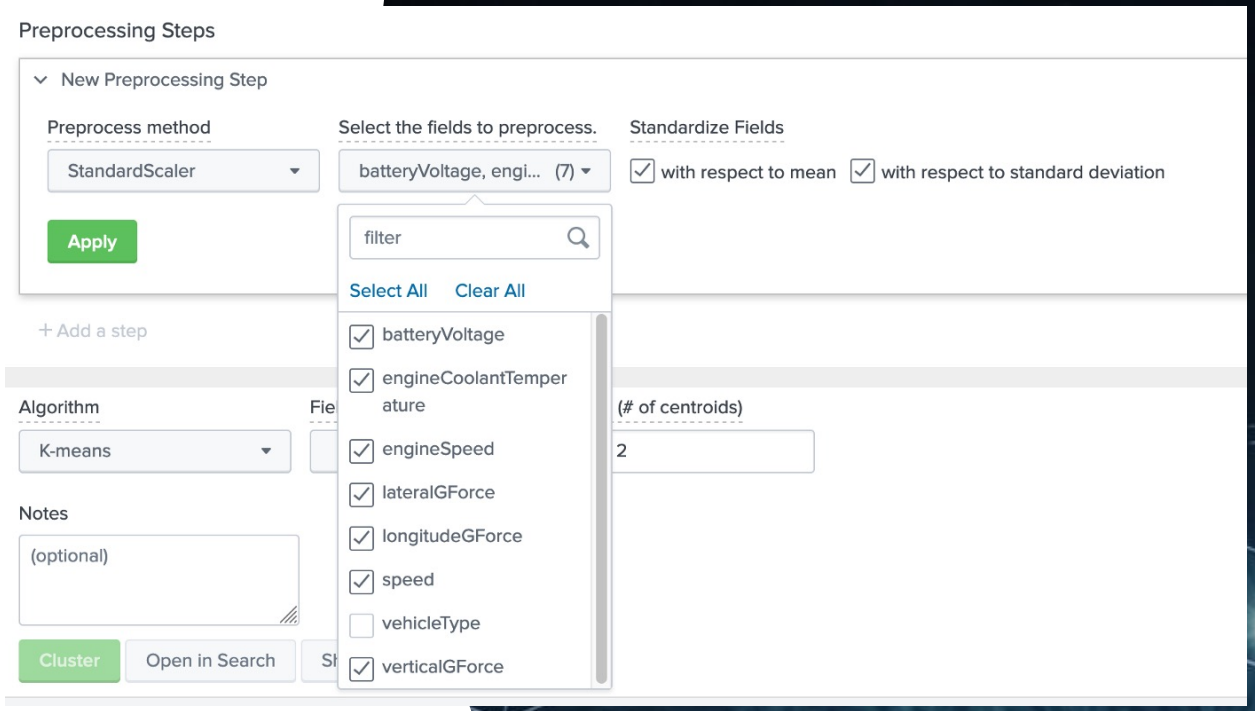


batteryVoltage	engineCoolantTemperature	engineSpeed	lateralGForce	longitudeGForce	speed	vehicleType	verticalGForce
13.785	93	6060	1.11	0.5	69	2015 Porsche GT3	-2.0
13.937	94	4957	0.56	0.7	56	2013 Audi RS5	0.95
13.827	93	6163	0.71	0.26	70	2015 Porsche GT3	-2.0
14.035	87	2846	0.81	-0.71	47	2011 Ford Mustang GT500	-2.0
13.827	93	6542	0.49	-0.18	76	2015 Porsche GT3	-2.0
14.624	105	4425	0.32	0.05	100	2014 Chevrolet Corvette	-0.17
13.827	93	7763	0.24	-0.18	91	2015 Porsche GT3	-2.0
13.827	93	6365	0.38	-0.2	95	2015 Porsche GT3	-2.0
13.827	93	6713	0.04	0.13	100	2015 Porsche GT3	-2.0
14.262	-10	1508	0.65	-0.01	6	2013 Audi RS5	0.77
13.931	6	4374	0.9	-0.82	49	2013 Audi RS5	0.6
13.808	92	5217	-0.3	0.57	86	2015 Porsche GT3	-2.0

Site

yber
&data

Forecasting



SS_batteryVoltage	SS_engineCoolantTemperature	SS_engineSpeed	SS_lateralGForce	SS_longitudeGForce	SS_speed
-0.9200116144159544	0.2781771818735084	1.2353344902461127	1.3133013746051845	1.2108729500560975	0.4309714079813609
-0.4981623490893071	0.324644719228529	0.5448435195306173	0.12941572241985255	1.5534995236073381	-0.07809860216241242
-0.8034480016283294	0.2781771818735084	1.2998137014733984	0.4522936275613065	0.7997210617946089	0.470130639530882
-0.22618058591818396	-0.0006280422566155339	-0.7766673047489938	0.6675455643222761	-0.8620178199289069	-0.4305316861081017
-0.8034480016283294	0.2781771818735084	1.53707215831943	-0.02126063331282617	0.045942599981880015	0.7050860288280081
1.4084853172225644	0.8357876301337561	0.21180526348288964	-0.3871889258064741	0.4399631595658065	1.6449075860165128
-0.8034480016283294	0.2781771818735084	2.3014325166545344	-0.5593904752152498	0.045942599981880015	1.2924745020708237
-0.8034480016283294	0.2781771818735084	1.426268076890017	-0.2580377637498925	0.011679942626755944	1.4491114282689077
-0.8034480016283294	0.2781771818735084	1.6441201691919138	-0.9898943487371886	0.5770137889863026	1.6449075860165128
0.4038179879577925	-4.5079791656936194	-1.6142710389442183	0.323142465504725	0.33717518750043435	-2.036060179638464

Site

cyber & data

Forecasting

Preprocessing Steps

StandardScaler

Preprocess method: StandardScaler

Select the fields to preprocess.: batteryVoltage, engi... (7)

Standardize Fields: with respect to mean, with respect to standard deviation

Apply

+ Add a step Preview Results

Algorithm: Birch

Fields to use for clustering: SS_batteryVoltage, S... (7)

K (# of centroids): 6

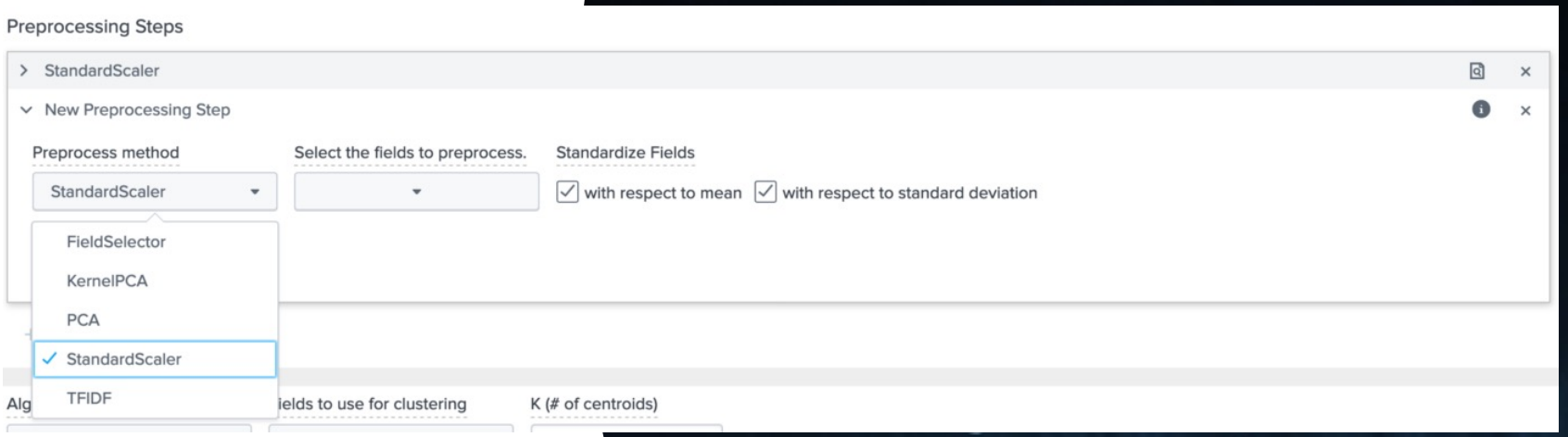
Notes: (optional)

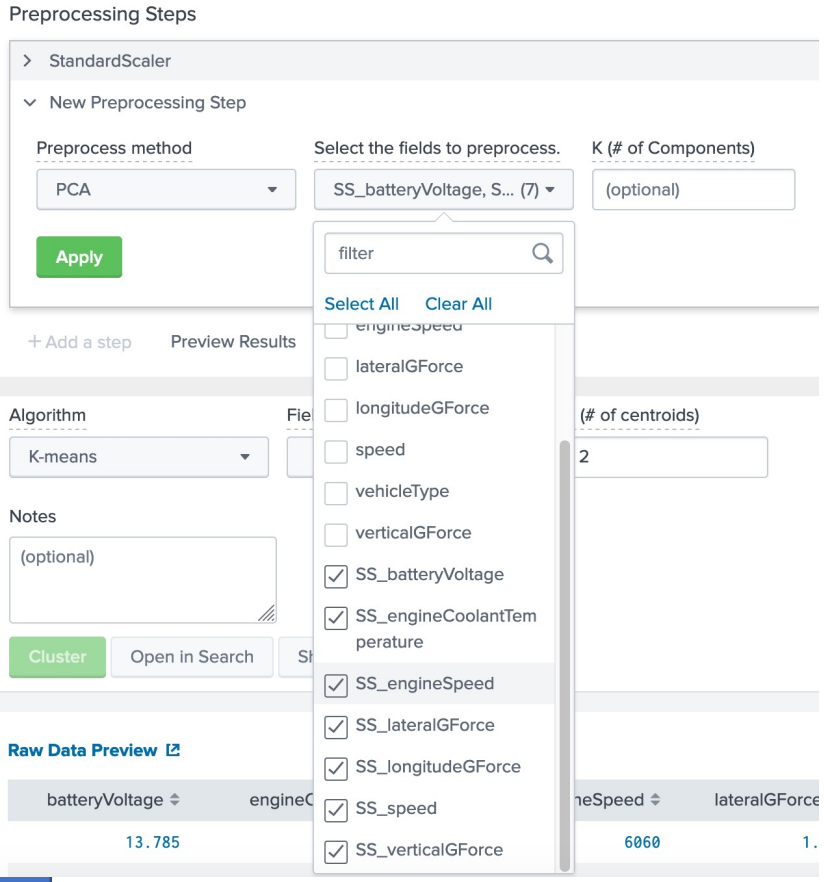
Cluster Open in Search Show SPL

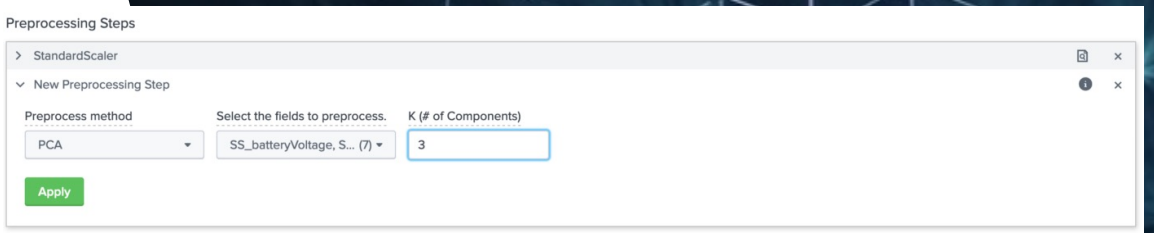
Site

cyber & data

Forecasting





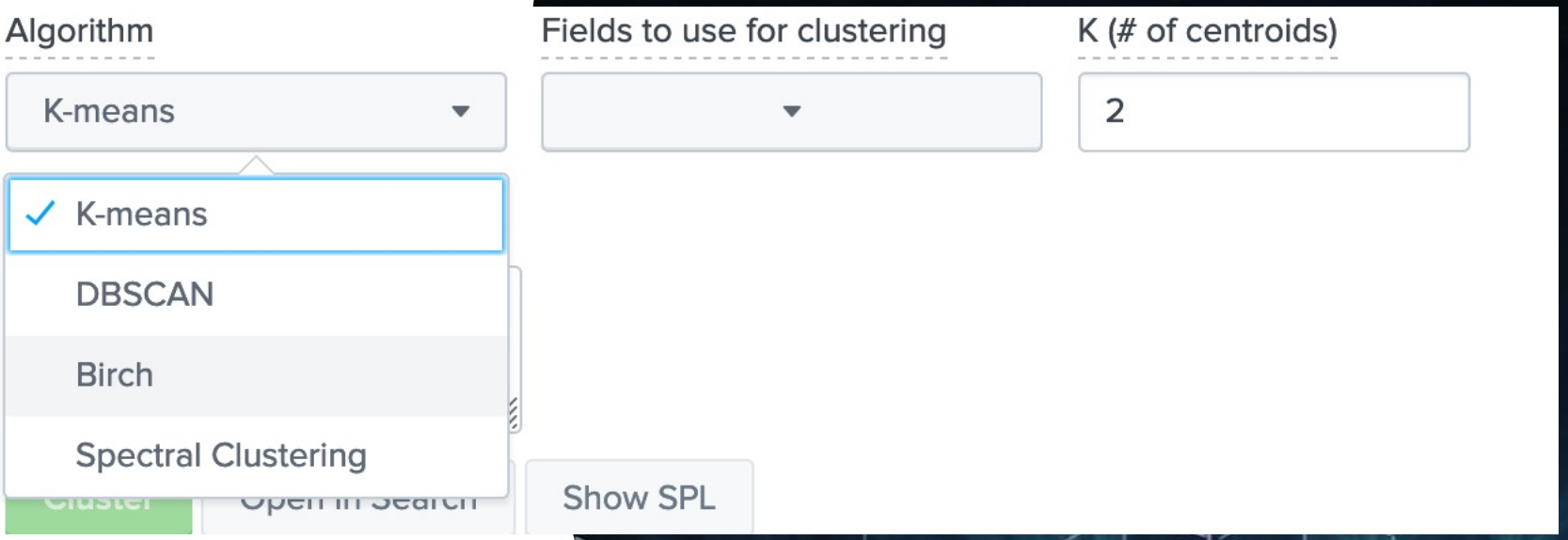


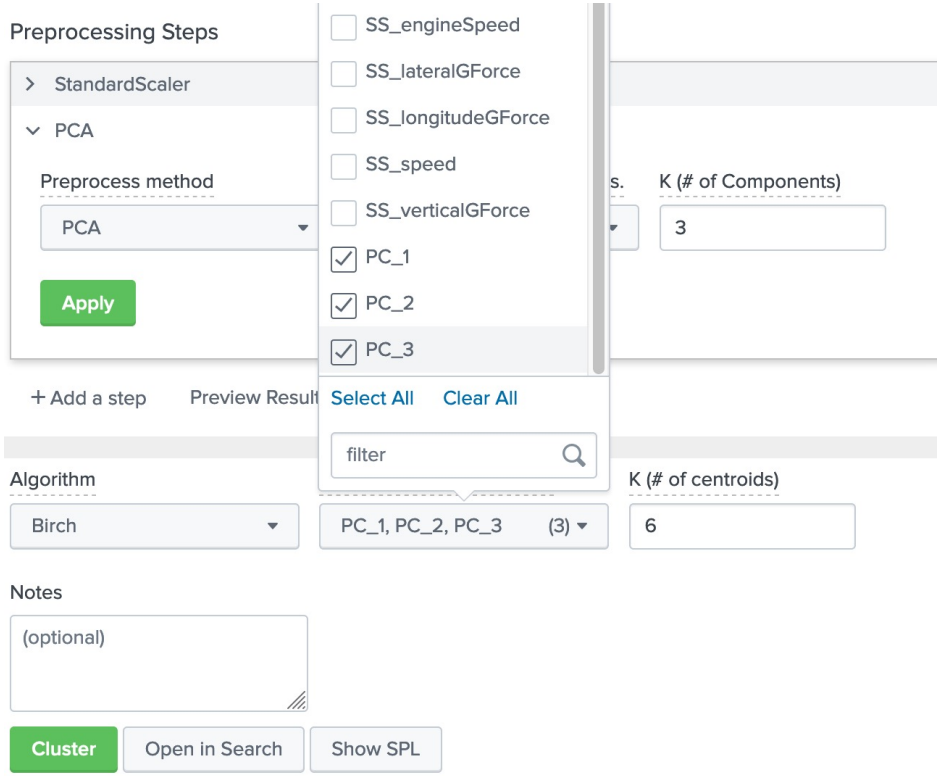
SS_verticalGForce	PC_1	PC_2	PC_3
-1.427968350953411	1.4281002997859338	-1.125352623502914	0.2353207267509022
1.4407323826344405	0.42375720491054764	-0.5173762506356252	1.1157215602667847
-1.427968350953411	1.4720198017320876	-0.5973633786344341	-0.3806236325808356
-1.427968350953411	-0.8514146130632574	-0.04044451718143394	-0.6244578333756619
-1.427968350953411	1.7163241051246994	0.0658788103242548	-0.8610567320289821
0.3515985447976967	1.2630352731698287	-1.1293102825533279	-0.38906036287221313
-1.427968350953411	2.6776737121095042	0.3449272197167691	-1.1556807183776086
-1.427968350953411	2.166604527943434	0.09322075146958986	-1.0084729739753058
-1.427968350953411	2.560278240655818	0.028979113959949565	-1.2703035251655508
1.2656930158392492	-2.542567650281204	1.7526714597753381	3.3827971149679685

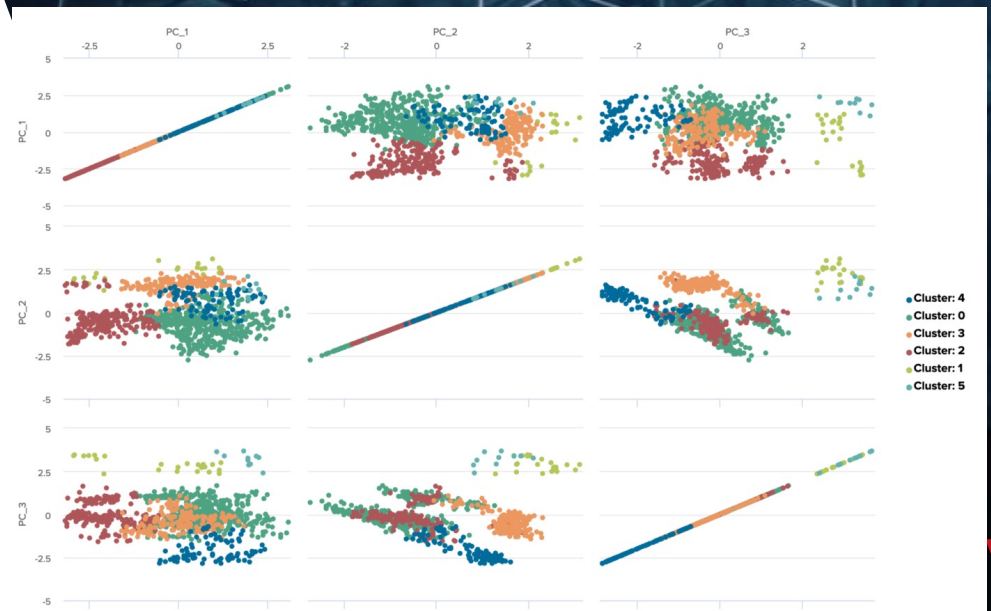
Site

&cyber data

Forecasting







Site

cyber & data

cyber & data

"From bits to information"

Introduction to Splunk and Machine Learning – Part 2