

LAB 5 – Elastic Machine Learning

Estimated Time for This Lab: 45 Min

Introduction

Elastic machine learning features automate the analysis of time-series data by creating accurate baselines of normal behavior in the data and identifying anomalous patterns in that data. You can submit your data for analysis in batches or continuously in real-time datafeeds.

Using proprietary machine learning algorithms, the following circumstances are detected, scored, and linked with statistically significant influencers in the data:

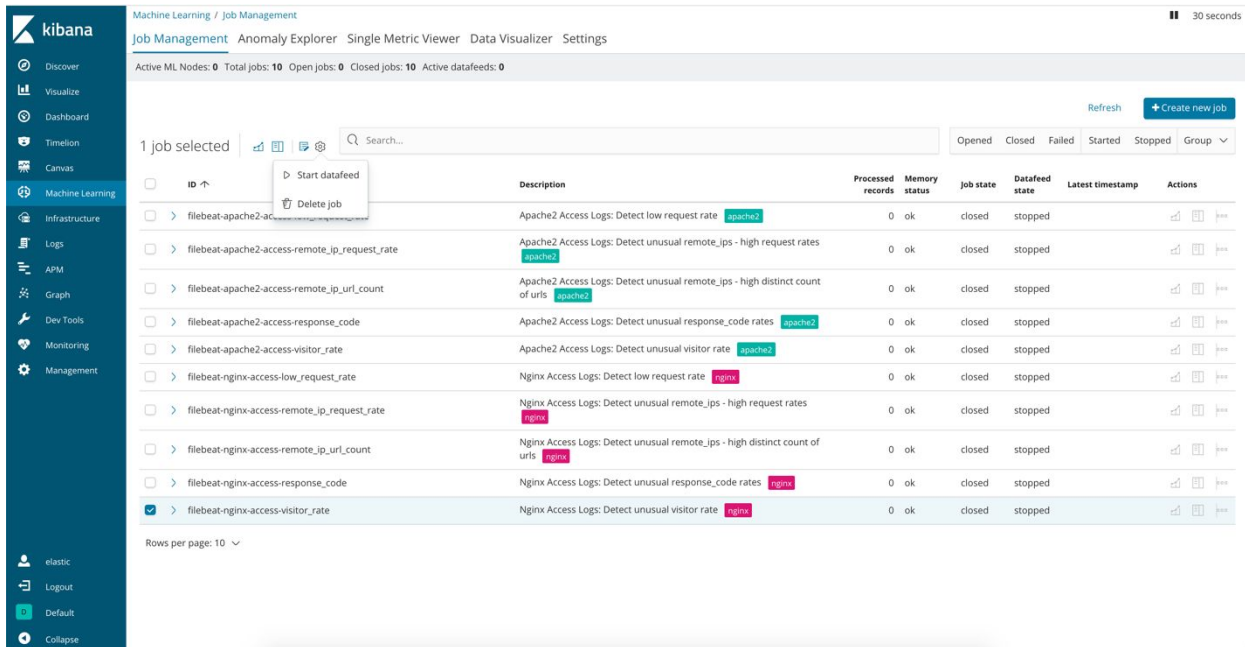
- Anomalies related to temporal deviations in values, counts, or frequencies
- Statistical rarity
- Unusual behaviors for a member of a population
- Automated periodicity detection and quick adaptation to changing data ensure that you don't need to specify algorithms, models, or other data science-related configurations in order to get the benefits of machine learning.

You can view the machine learning results in Kibana where, for example, charts illustrate the actual data values, the bounds for the expected values, and the anomalies that occur outside these bounds.

Machine Learning is a very powerful addition to your toolkit when working with log data. In this lab we will use Elastic Machine Learning to review the EGINX logs we loaded in Lab 1.

Let's get started!

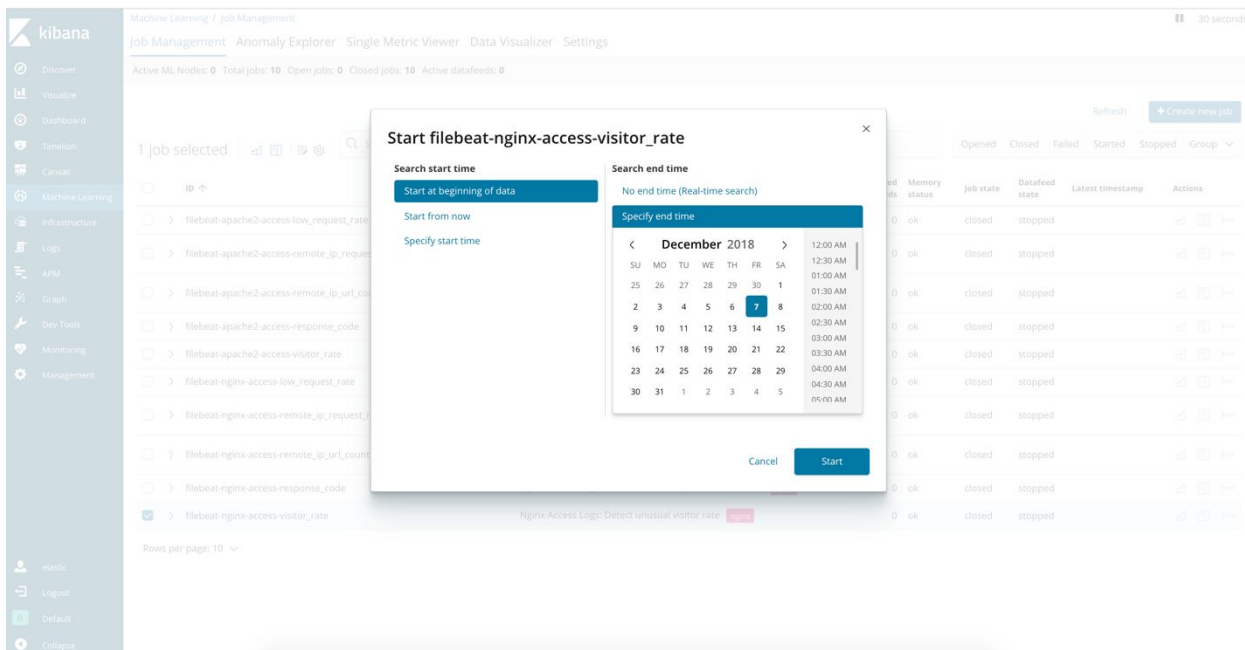
1. In Kibana click on “Machine Learning” item in the menu. In the list of jobs select “filebeat-nginx-access-visitor_rate” and click on “Start datafeed”.



The screenshot shows the Kibana Machine Learning Job Management interface. The left sidebar contains the Kibana logo and a navigation menu with options: Discover, Visualize, Dashboard, Timeline, Canvas, Machine Learning (selected), Infrastructure, Logs, APM, Graph, Dev Tools, Monitoring, and Management. The main content area is titled 'Machine Learning / Job Management' and includes tabs for Job Management, Anomaly Explorer, Single Metric Viewer, Data Visualizer, and Settings. Below the tabs, it displays 'Active ML Nodes: 0', 'Total jobs: 10', 'Open jobs: 0', 'Closed jobs: 10', and 'Active datafeeds: 0'. A table lists 10 jobs, with 'filebeat-nginx-access-visitor_rate' selected. A context menu is open over this job, showing 'Start datafeed' and 'Delete job' options. The table columns are: ID, Description, Processed records, Memory status, Job state, Datafeed state, Latest timestamp, and Actions.

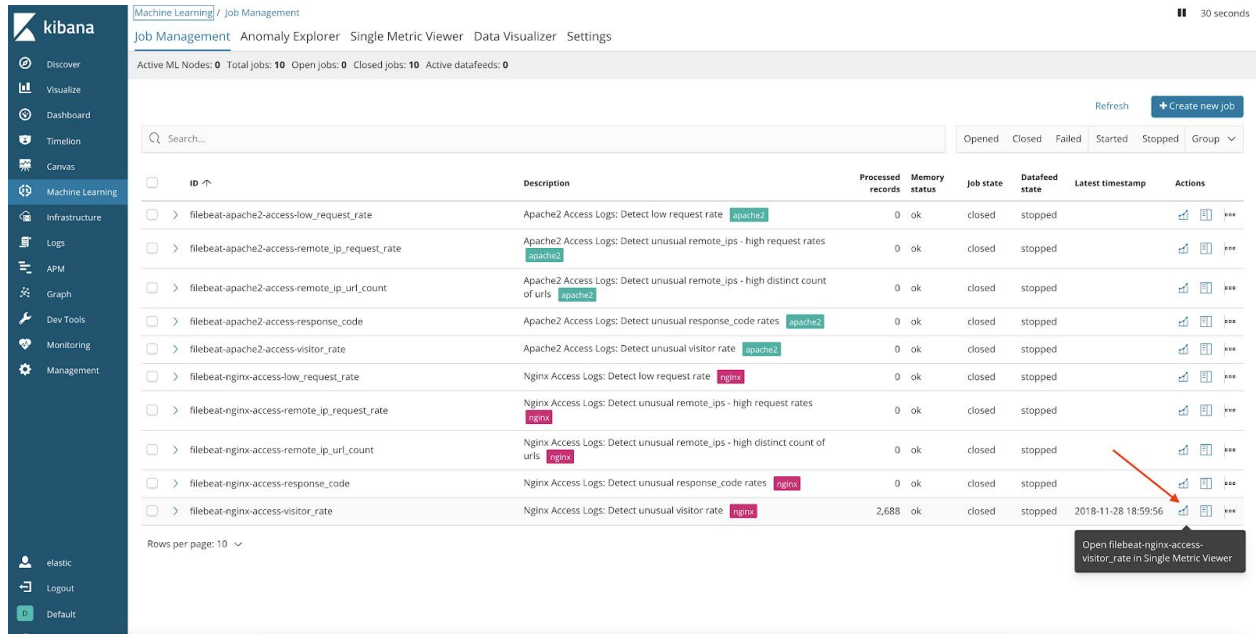
ID	Description	Processed records	Memory status	Job state	Datafeed state	Latest timestamp	Actions
filebeat-apache2-access-low_request_rate	Apache2 Access Logs: Detect low request rate	0	ok	closed	stopped		
filebeat-apache2-access-remote_ip_request_rate	Apache2 Access Logs: Detect unusual remote_ips - high request rates	0	ok	closed	stopped		
filebeat-apache2-access-remote_ip_url_count	Apache2 Access Logs: Detect unusual remote_ips - high distinct count of urls	0	ok	closed	stopped		
filebeat-apache2-access-response_code	Apache2 Access Logs: Detect unusual response_code rates	0	ok	closed	stopped		
filebeat-apache2-access-visitor_rate	Apache2 Access Logs: Detect unusual visitor rate	0	ok	closed	stopped		
filebeat-nginx-access-low_request_rate	Nginx Access Logs: Detect low request rate	0	ok	closed	stopped		
filebeat-nginx-access-remote_ip_request_rate	Nginx Access Logs: Detect unusual remote_ips - high request rates	0	ok	closed	stopped		
filebeat-nginx-access-remote_ip_url_count	Nginx Access Logs: Detect unusual remote_ips - high distinct count of urls	0	ok	closed	stopped		
filebeat-nginx-access-response_code	Nginx Access Logs: Detect unusual response_code rates	0	ok	closed	stopped		
filebeat-nginx-access-visitor_rate	Nginx Access Logs: Detect unusual visitor rate	0	ok	closed	stopped		

2. Select time period “Start at beginning of data” and click on “Start”.



The screenshot shows the same Kibana Machine Learning Job Management page, but with a modal dialog open for the selected job 'filebeat-nginx-access-visitor_rate'. The dialog is titled 'Start filebeat-nginx-access-visitor_rate' and has two sections: 'Search start time' and 'Search end time'. Under 'Search start time', the option 'Start at beginning of data' is selected. Under 'Search end time', the option 'No end time (Real-time search)' is selected. A calendar widget is visible, showing the date '7' selected. The 'Start' button is highlighted in blue. The background table and sidebar are partially visible behind the dialog.

- Wait until the Datafeed state changes to “stopped” and click on the results icon to view the output of the job in Single Metrics viewer



Machine Learning / Job Management

Job Management Anomaly Explorer Single Metric Viewer Data Visualizer Settings

Active ML Nodes: 0 Total jobs: 10 Open jobs: 0 Closed jobs: 10 Active datafeeds: 0

Search...

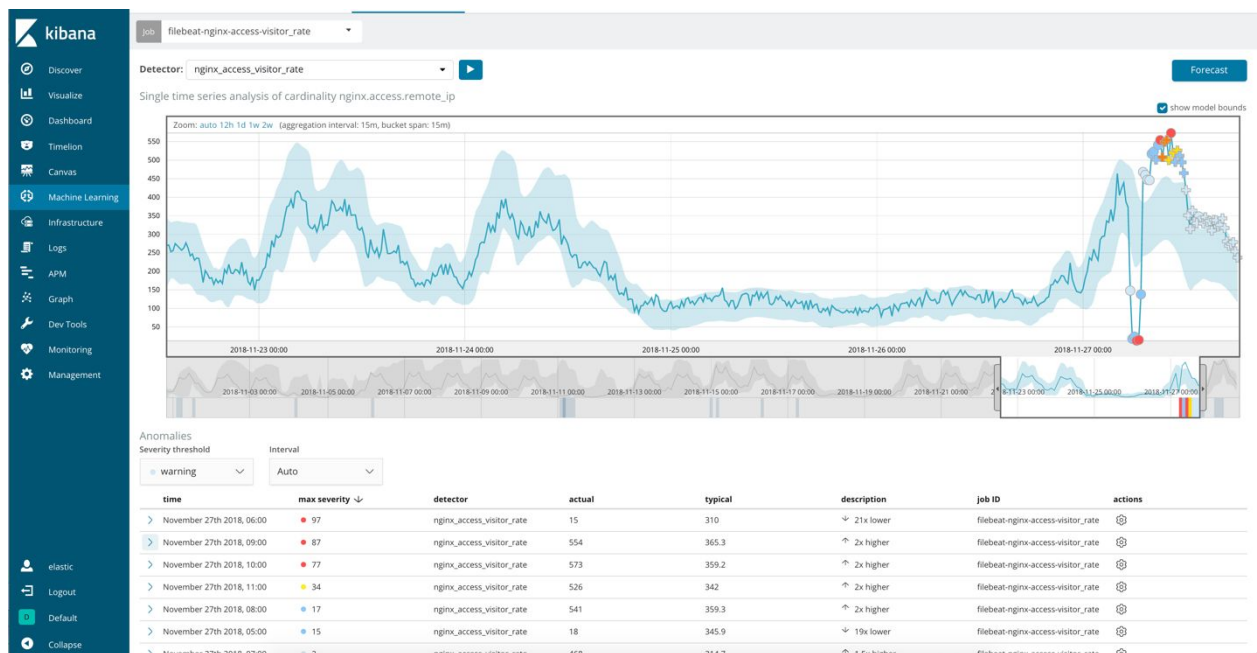
Refresh Create new job

ID	Description	Processed records	Memory status	Job state	Datafeed state	Latest timestamp	Actions
filebeat-apache2-access-low_request_rate	Apache2 Access Logs: Detect low request rate apache2	0	ok	closed	stopped		
filebeat-apache2-access-remote_ip_request_rate	Apache2 Access Logs: Detect unusual remote_ips - high request rates apache2	0	ok	closed	stopped		
filebeat-apache2-access-remote_ip_url_count	Apache2 Access Logs: Detect unusual remote_ips - high distinct count of urls apache2	0	ok	closed	stopped		
filebeat-apache2-access-response_code	Apache2 Access Logs: Detect unusual response_code rates apache2	0	ok	closed	stopped		
filebeat-apache2-access-visitor_rate	Apache2 Access Logs: Detect unusual visitor rate apache2	0	ok	closed	stopped		
filebeat-nginx-access-low_request_rate	Ngix Access Logs: Detect low request rate nginx	0	ok	closed	stopped		
filebeat-nginx-access-remote_ip_request_rate	Ngix Access Logs: Detect unusual remote_ips - high request rates nginx	0	ok	closed	stopped		
filebeat-nginx-access-remote_ip_url_count	Ngix Access Logs: Detect unusual remote_ips - high distinct count of urls nginx	0	ok	closed	stopped		
filebeat-nginx-access-response_code	Ngix Access Logs: Detect unusual response_code rates nginx	0	ok	closed	stopped		
filebeat-nginx-access-visitor_rate	Ngix Access Logs: Detect unusual visitor rate nginx	2,688	ok	closed	stopped	2018-11-28 18:59:56	

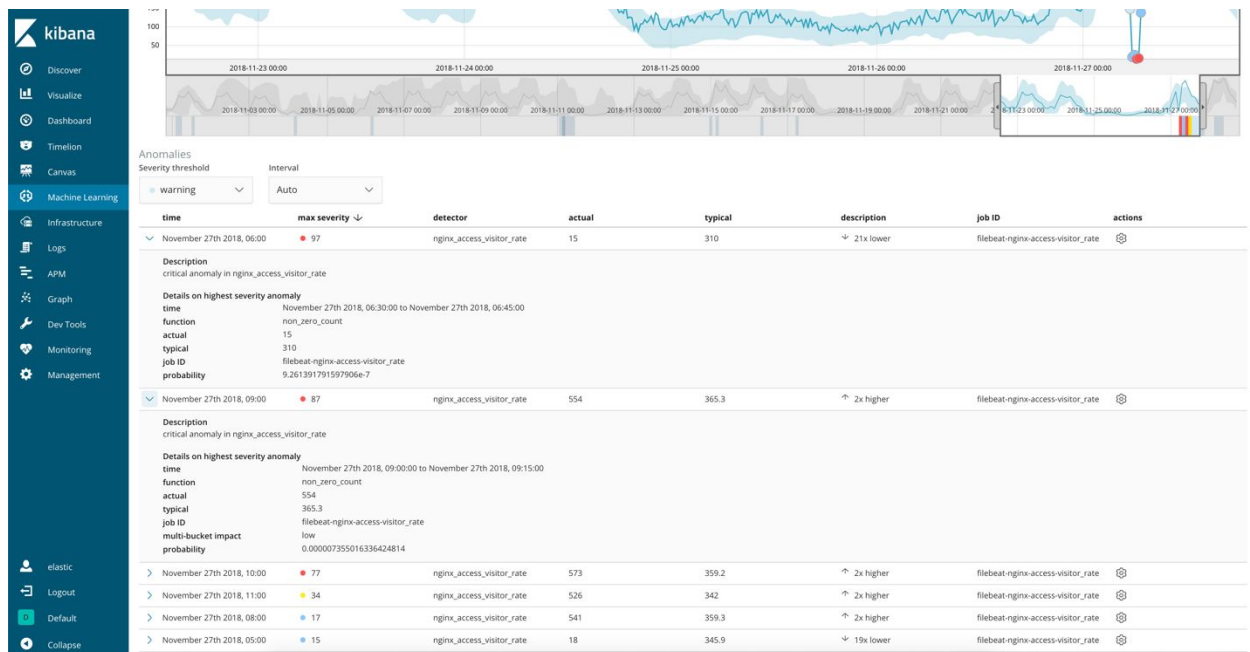
Rows per page: 10

Open filebeat-nginx-access-visitor_rate in Single Metric Viewer

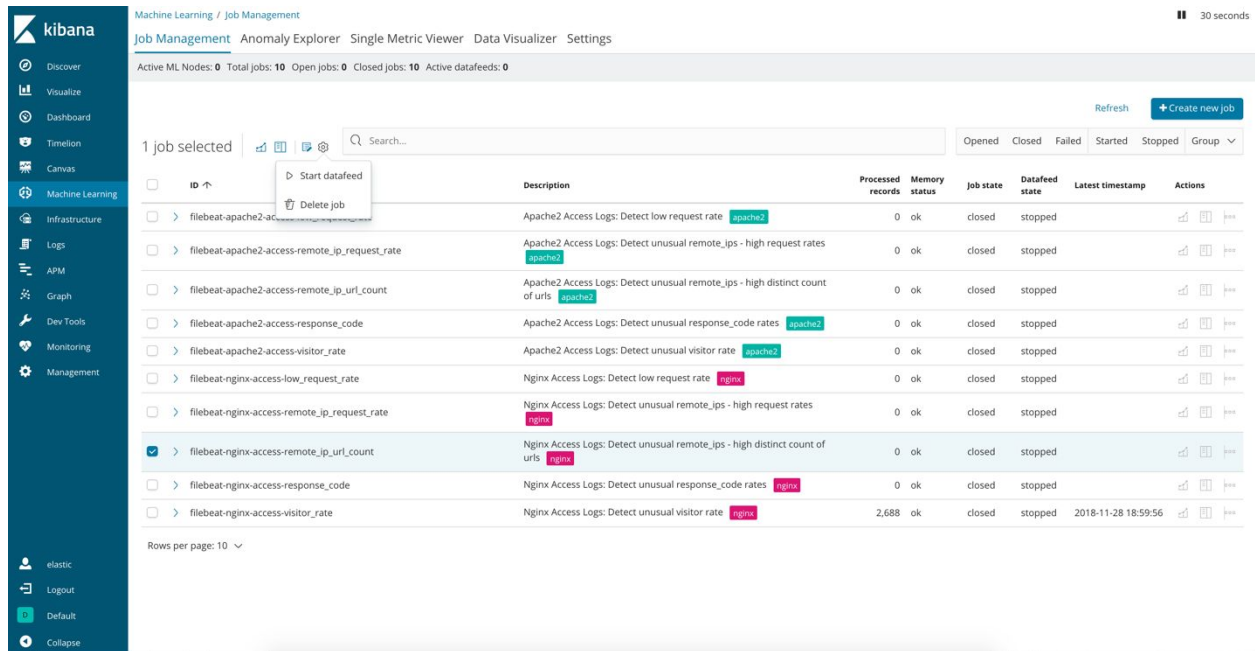
- Slide the window to see how the model (shaded blue area) was built and how it got narrower (i.e. more precise) over time. The anomaly that you will see towards the end is the sudden drop in visitor rate and then the spike.



- Expand top two anomalies by clicking on them. Note how both anomalies are severe, despite the drop is 21x lower, and the spike is 2x higher. The reason for that is because what contributes to anomaly is not how much higher/lower the actual value is against the expected value, but the probability of its occurrence. The probability value is provided once you click on Anomalies to expand the results.



- Click on “Machine Learning” menu in Kibana again. Select “filebeat-nginx-access-remote_ip_url_count” job and start datafeed. Start from the beginning of data and click on Start.



Machine Learning / Job Management

Job Management Anomaly Explorer Single Metric Viewer Data Visualizer Settings

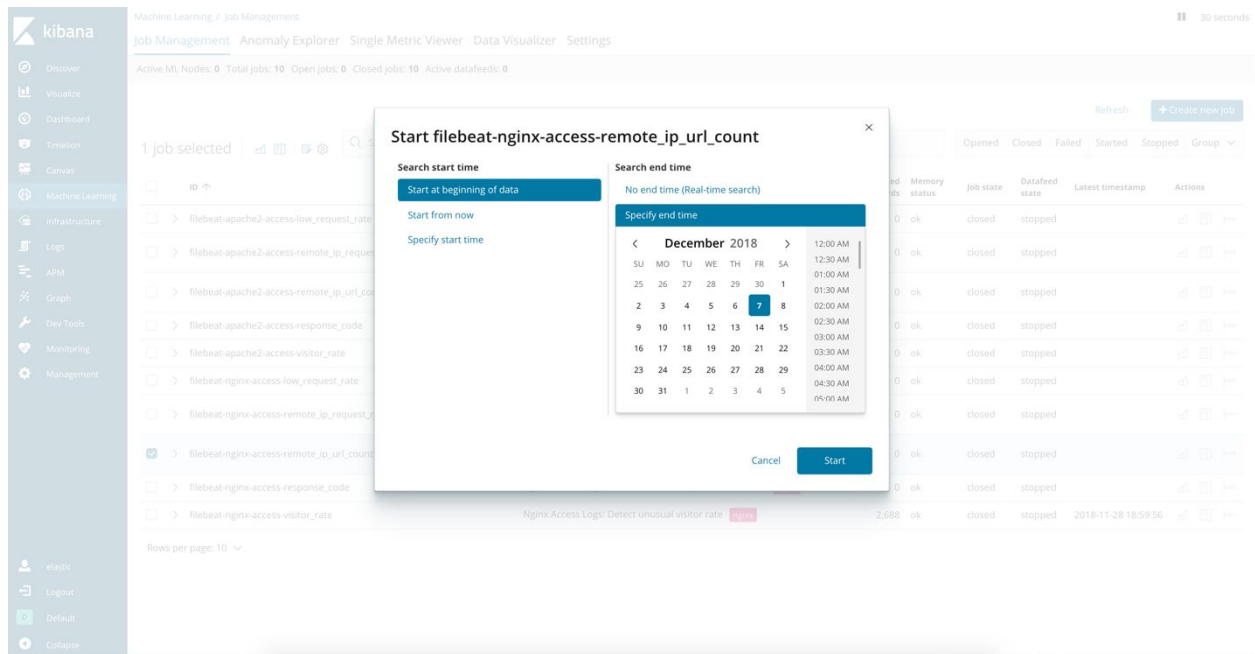
Active ML Nodes: 0 Total jobs: 10 Open jobs: 0 Closed jobs: 10 Active datafeeds: 0

Refresh Create new job

1 job selected

ID	Description	Processed records	Memory status	Job state	Datafeed state	Latest timestamp	Actions
filebeat-apache2-access-low_request_rate	Apache2 Access Logs: Detect low request rate	0	ok	closed	stopped		
filebeat-apache2-access-remote_ip_request_rate	Apache2 Access Logs: Detect unusual remote_ips - high request rates	0	ok	closed	stopped		
filebeat-apache2-access-remote_ip_url_count	Apache2 Access Logs: Detect unusual remote_ips - high distinct count of urls	0	ok	closed	stopped		
filebeat-apache2-access-response_code	Apache2 Access Logs: Detect unusual response_code rates	0	ok	closed	stopped		
filebeat-apache2-access-visitor_rate	Apache2 Access Logs: Detect unusual visitor rate	0	ok	closed	stopped		
filebeat-nginx-access-low_request_rate	Ngix Access Logs: Detect low request rate	0	ok	closed	stopped		
filebeat-nginx-access-remote_ip_request_rate	Ngix Access Logs: Detect unusual remote_ips - high request rates	0	ok	closed	stopped		
filebeat-nginx-access-remote_ip_url_count	Ngix Access Logs: Detect unusual remote_ips - high distinct count of urls	0	ok	closed	stopped		
filebeat-nginx-access-response_code	Ngix Access Logs: Detect unusual response_code rates	0	ok	closed	stopped		
filebeat-nginx-access-visitor_rate	Ngix Access Logs: Detect unusual visitor rate	2,688	ok	closed	stopped	2018-11-28 18:59:56	

Rows per page: 10



Machine Learning / Job Management

Job Management Anomaly Explorer Single Metric Viewer Data Visualizer Settings

Active ML Nodes: 0 Total jobs: 10 Open jobs: 0 Closed jobs: 10 Active datafeeds: 0

Refresh Create new job

1 job selected

Start filebeat-nginx-access-remote_ip_url_count

Search start time

Start at beginning of data

Start from now

Specify start time

Search end time

No end time (Real-time search)

Specify end time

< December 2018 >

SU MO TU WE TH FR SA

25 26 27 28 29 30 1

2 3 4 5 6 7 8

9 10 11 12 13 14 15

16 17 18 19 20 21 22

23 24 25 26 27 28 29

30 31 1 2 3 4 5

12:00 AM

12:30 AM

01:00 AM

01:30 AM

02:00 AM

02:30 AM

03:00 AM

03:30 AM

04:00 AM

04:30 AM

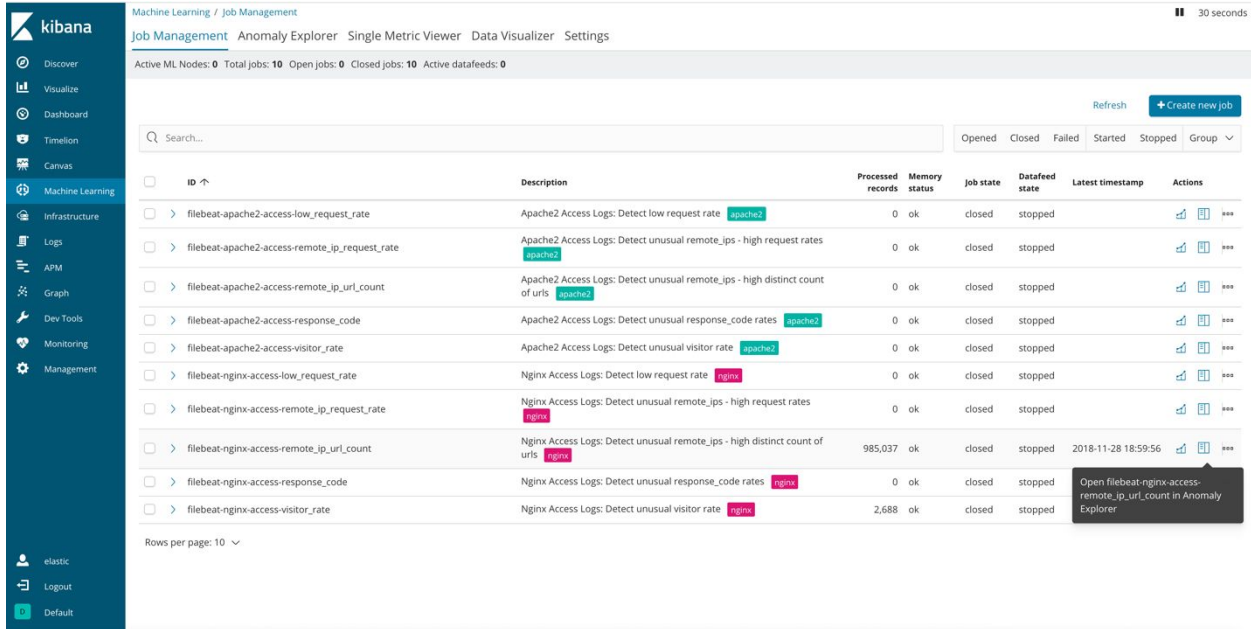
05:00 AM

Cancel Start

ID	Description	Processed records	Memory status	Job state	Datafeed state	Latest timestamp	Actions
filebeat-apache2-access-low_request_rate	Apache2 Access Logs: Detect low request rate	0	ok	closed	stopped		
filebeat-apache2-access-remote_ip_request_rate	Apache2 Access Logs: Detect unusual remote_ips - high request rates	0	ok	closed	stopped		
filebeat-apache2-access-remote_ip_url_count	Apache2 Access Logs: Detect unusual remote_ips - high distinct count of urls	0	ok	closed	stopped		
filebeat-apache2-access-response_code	Apache2 Access Logs: Detect unusual response_code rates	0	ok	closed	stopped		
filebeat-apache2-access-visitor_rate	Apache2 Access Logs: Detect unusual visitor rate	0	ok	closed	stopped		
filebeat-nginx-access-low_request_rate	Ngix Access Logs: Detect low request rate	0	ok	closed	stopped		
filebeat-nginx-access-remote_ip_request_rate	Ngix Access Logs: Detect unusual remote_ips - high request rates	0	ok	closed	stopped		
filebeat-nginx-access-remote_ip_url_count	Ngix Access Logs: Detect unusual remote_ips - high distinct count of urls	0	ok	closed	stopped		
filebeat-nginx-access-response_code	Ngix Access Logs: Detect unusual response_code rates	0	ok	closed	stopped		
filebeat-nginx-access-visitor_rate	Ngix Access Logs: Detect unusual visitor rate	2,688	ok	closed	stopped	2018-11-28 18:59:56	

Rows per page: 10

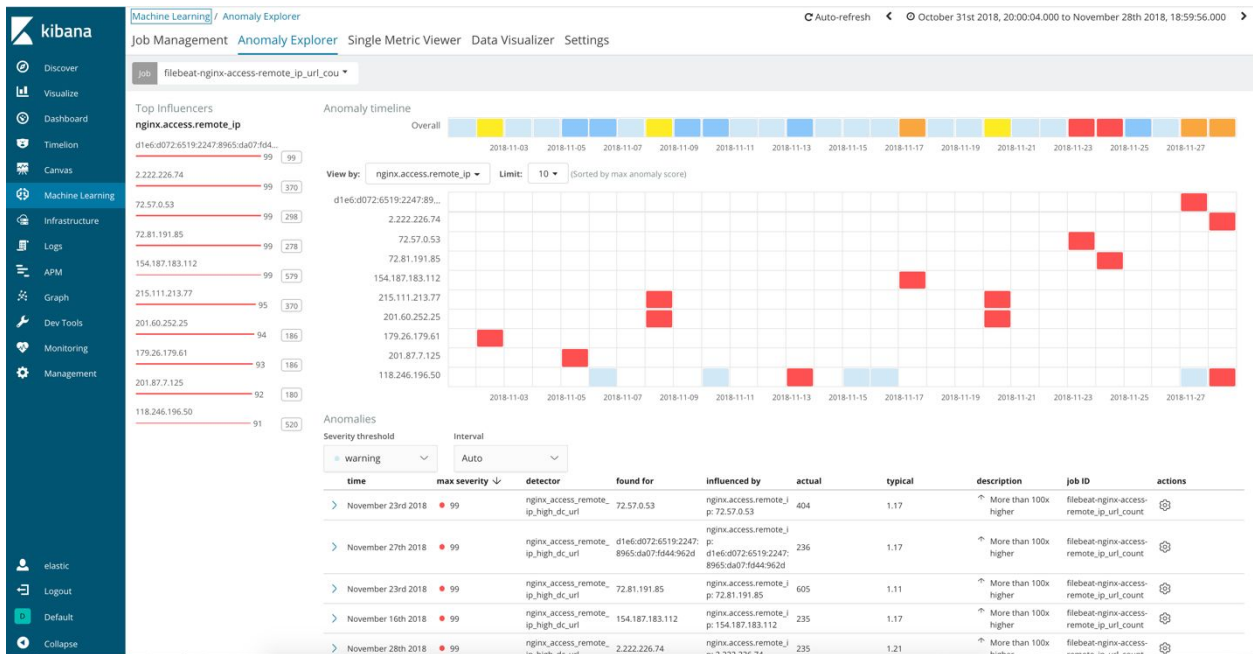
- Wait until the datafeed state changes to “stopped” and click on the results icon to view the output of the job in Anomaly Explorer.



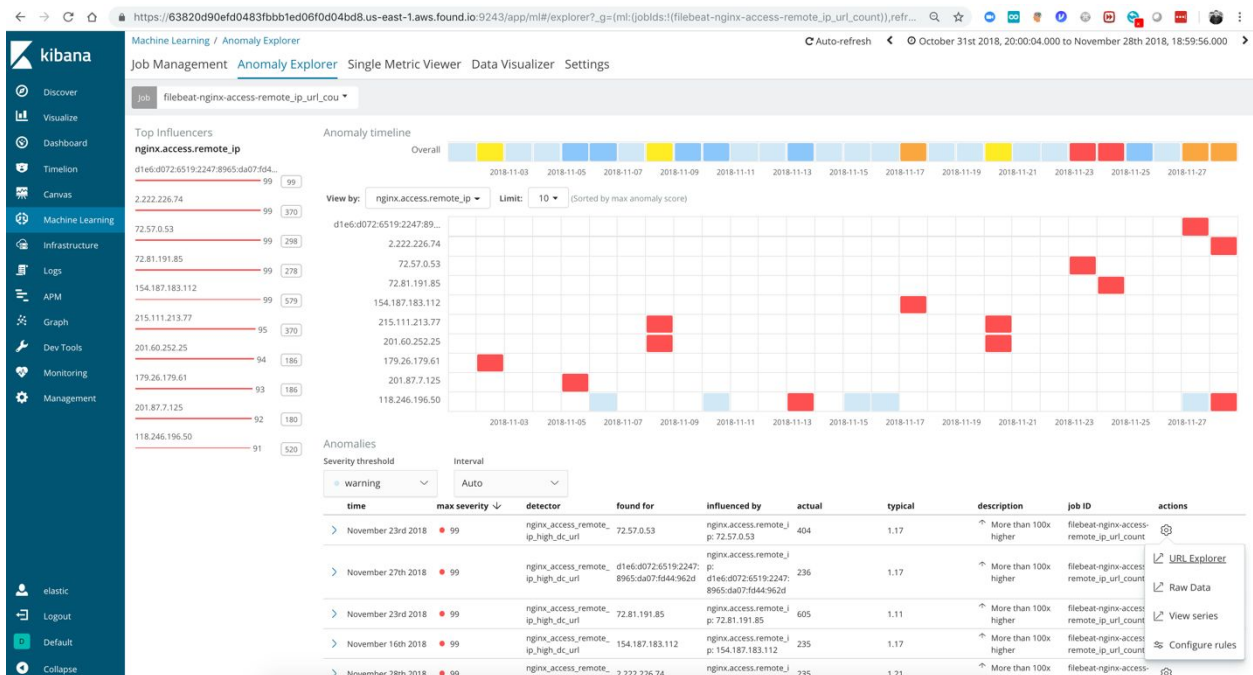
The screenshot shows the Kibana Machine Learning interface. The left sidebar contains navigation links for Discover, Visualize, Dashboard, Timeline, Canvas, Machine Learning, Infrastructure, Logs, APM, Graph, Dev Tools, Monitoring, and Management. The main content area is titled 'Machine Learning / Job Management' and includes a search bar, a 'Refresh' button, and a '+ Create new job' button. Below these is a table of jobs with columns for ID, Description, Processed records, Memory status, Job state, Datafeed state, Latest timestamp, and Actions. The table lists several jobs, including 'filebeat-apache2-access-low_request_rate', 'filebeat-apache2-access-remote_ip_request_rate', 'filebeat-apache2-access-remote_ip_url_count', 'filebeat-apache2-access-response_code', 'filebeat-apache2-access-visitor_rate', 'filebeat-nginx-access-low_request_rate', 'filebeat-nginx-access-remote_ip_request_rate', 'filebeat-nginx-access-remote_ip_url_count', 'filebeat-nginx-access-response_code', and 'filebeat-nginx-access-visitor_rate'. A tooltip is visible over the 'filebeat-nginx-access-remote_ip_url_count' job, indicating it can be opened in the Anomaly Explorer.

ID	Description	Processed records	Memory status	Job state	Datafeed state	Latest timestamp	Actions
filebeat-apache2-access-low_request_rate	Apache2 Access Logs: Detect low request rate apache2	0	ok	closed	stopped		
filebeat-apache2-access-remote_ip_request_rate	Apache2 Access Logs: Detect unusual remote_ips - high request rates apache2	0	ok	closed	stopped		
filebeat-apache2-access-remote_ip_url_count	Apache2 Access Logs: Detect unusual remote_ips - high distinct count of urls apache2	0	ok	closed	stopped		
filebeat-apache2-access-response_code	Apache2 Access Logs: Detect unusual response_code rates apache2	0	ok	closed	stopped		
filebeat-apache2-access-visitor_rate	Apache2 Access Logs: Detect unusual visitor rate apache2	0	ok	closed	stopped		
filebeat-nginx-access-low_request_rate	Ngix Access Logs: Detect low request rate nginx	0	ok	closed	stopped		
filebeat-nginx-access-remote_ip_request_rate	Ngix Access Logs: Detect unusual remote_ips - high request rates nginx	0	ok	closed	stopped		
filebeat-nginx-access-remote_ip_url_count	Ngix Access Logs: Detect unusual remote_ips - high distinct count of urls nginx	985,037	ok	closed	stopped	2018-11-28 18:59:56	
filebeat-nginx-access-response_code	Ngix Access Logs: Detect unusual response_code rates nginx	0	ok	closed	stopped		
filebeat-nginx-access-visitor_rate	Ngix Access Logs: Detect unusual visitor rate nginx	2,688	ok	closed	stopped		

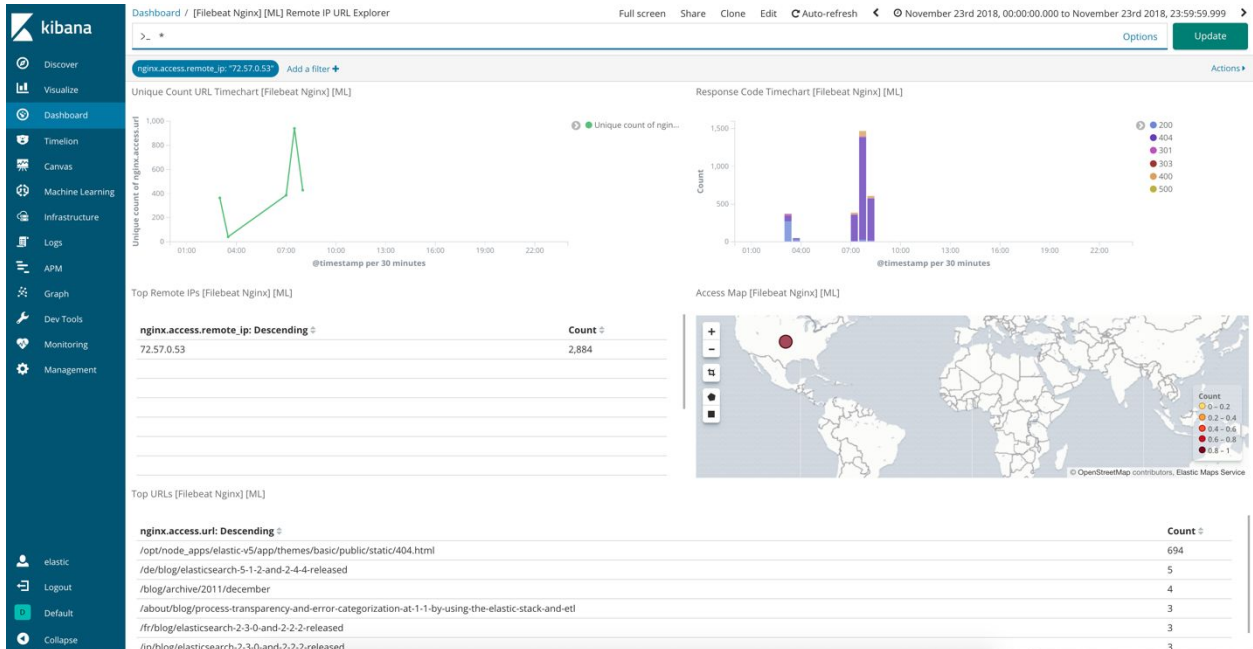
- You will see the anomaly explorer where the results are broken down by particular ips. The influencers that you see are statistically significant elements in our data set contributing to the anomaly. You have an option to select a particular field in your data as contributors when you're creating a Machine Learning job. On top of the Anomaly Timeline you see Overall result for the job, and inside the matrix you see the result of the job for a particular IP. Note that results are not the same. Hover over the overall results to see the score across the whole population and over particular IPs contributing to that same score. When the score for IP "154.187.183.112" is 99 overall anomaly score is only 55.



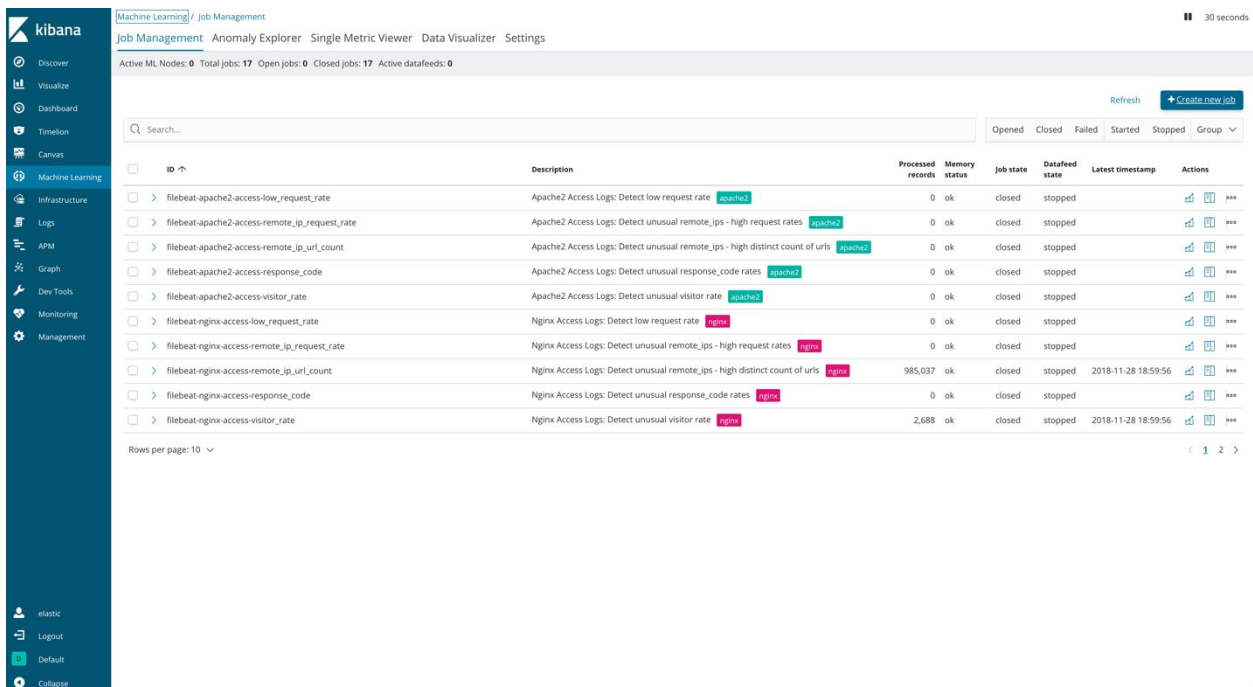
9. We have the ability to link the anomaly results of ML job with particular views when we setup the job. When opening the views the metadata from the job can be passed to it. Click on “actions” next to top anomaly and click on “URL Explorer”.



Review the Remote IP URL Explorer dashboard, note how its view was customized based on the IP for the anomaly (i.e. on the metadata from the ML job on the previous screen)



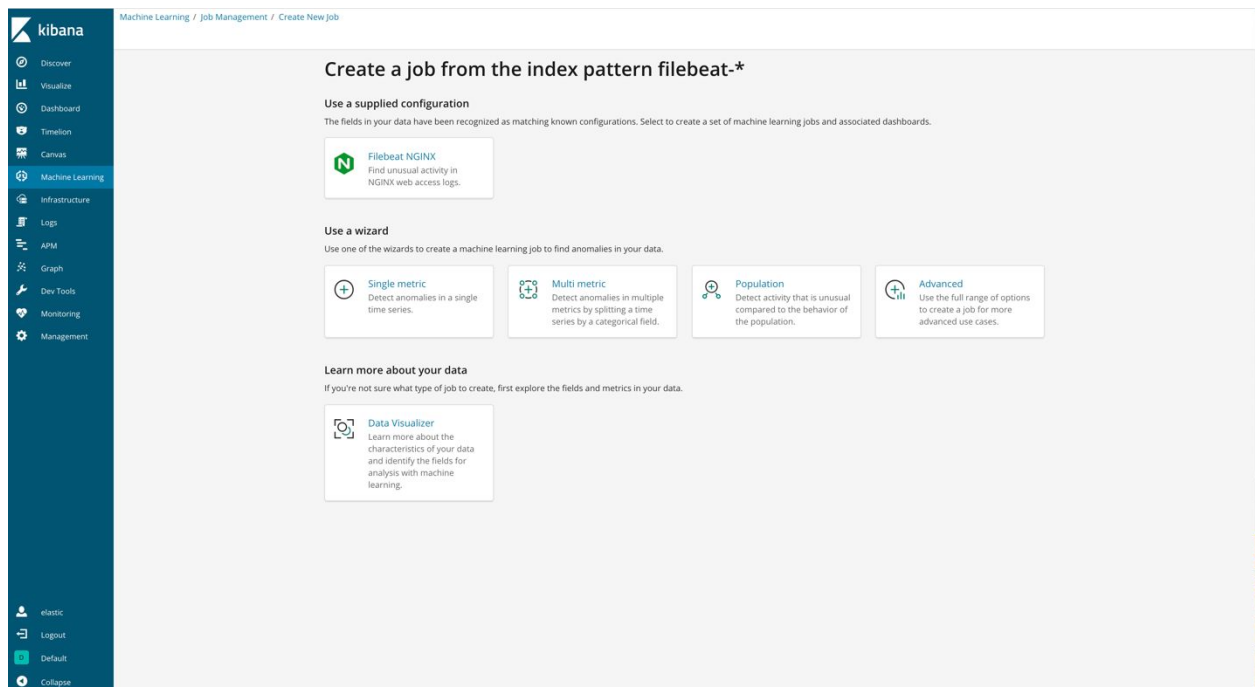
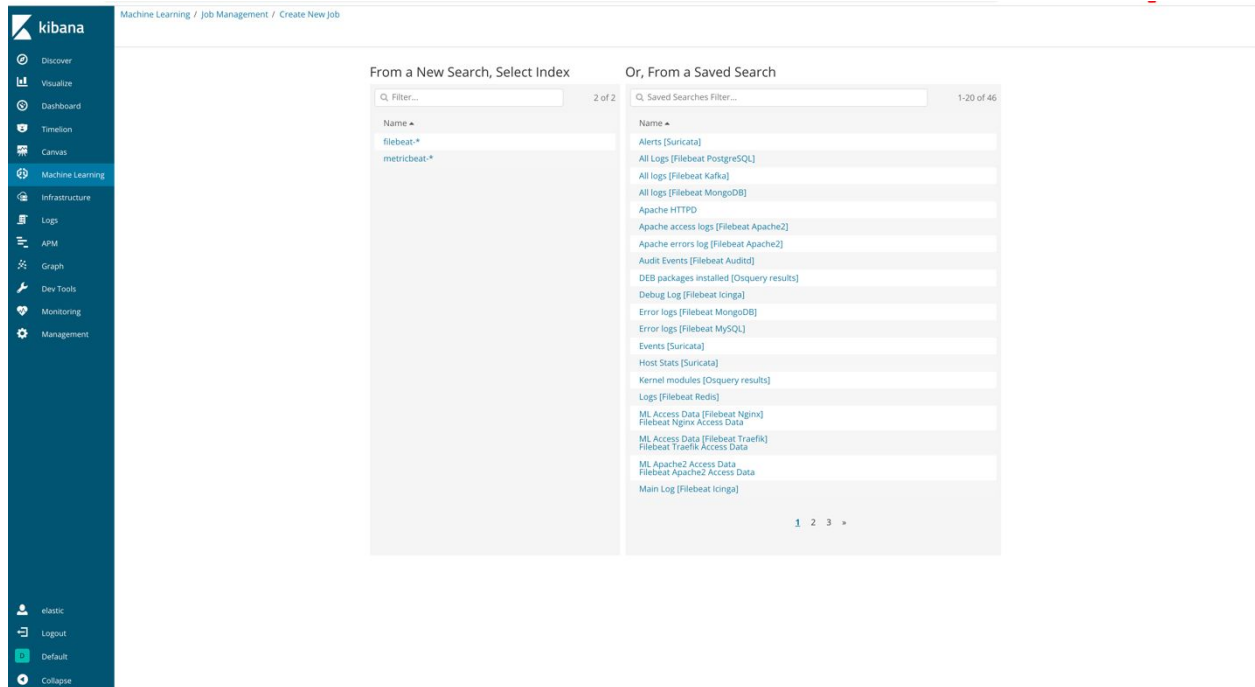
10. (This and next steps are optional) Create your own Single Metrics Machine Learning job. Click on Machine Learning in Kibana menu. Click on Create Job.



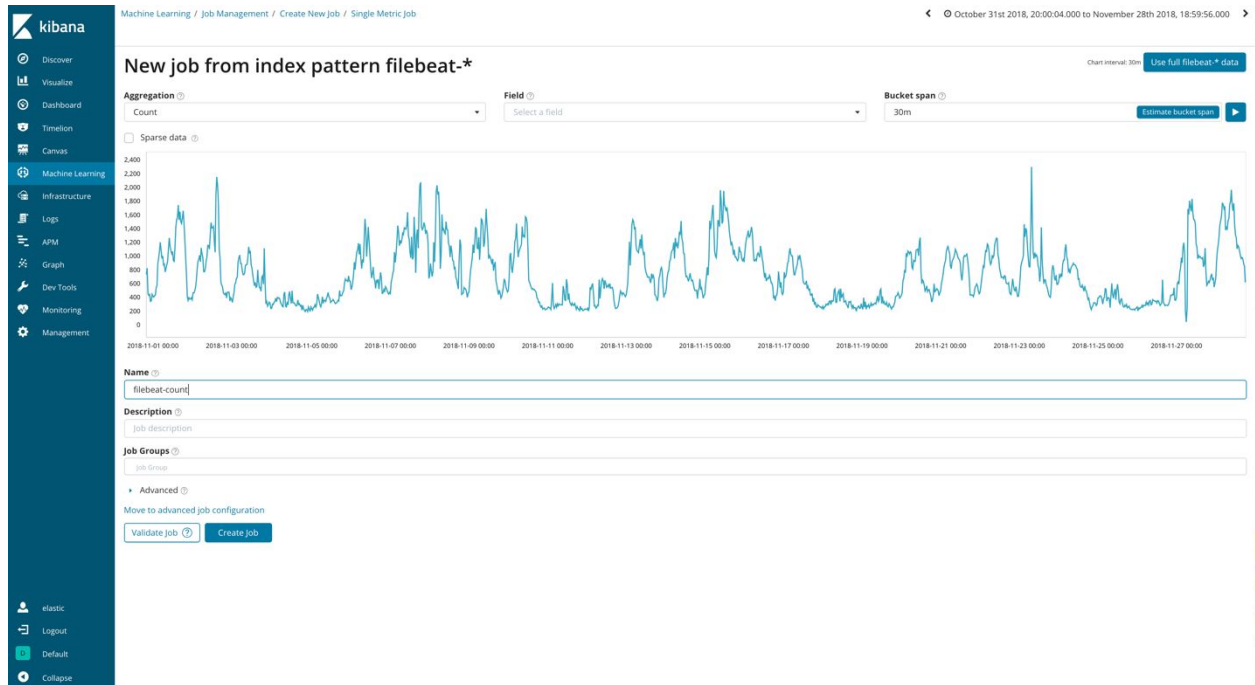
The Machine Learning Job Management page shows a list of jobs with the following columns: ID, Description, Processed records, Memory status, Job state, Datafeed state, Latest timestamp, and Actions.

ID	Description	Processed records	Memory status	Job state	Datafeed state	Latest timestamp	Actions
> filebeat-apache2-access-low_request_rate	Apache2 Access Logs: Detect low request rate	0	ok	closed	stopped		View Edit Delete
> filebeat-apache2-access-remote_ip_request_rate	Apache2 Access Logs: Detect unusual remote_ips - high request rates	0	ok	closed	stopped		View Edit Delete
> filebeat-apache2-access-remote_ip_url_count	Apache2 Access Logs: Detect unusual remote_ips - high distinct count of urls	0	ok	closed	stopped		View Edit Delete
> filebeat-apache2-access-response_code	Apache2 Access Logs: Detect unusual response_code rates	0	ok	closed	stopped		View Edit Delete
> filebeat-apache2-access-visitor_rate	Apache2 Access Logs: Detect unusual visitor rate	0	ok	closed	stopped		View Edit Delete
> filebeat-nginx-access-low_request_rate	Nginx Access Logs: Detect low request rate	0	ok	closed	stopped		View Edit Delete
> filebeat-nginx-access-remote_ip_request_rate	Nginx Access Logs: Detect unusual remote_ips - high request rates	0	ok	closed	stopped		View Edit Delete
> filebeat-nginx-access-remote_ip_url_count	Nginx Access Logs: Detect unusual remote_ips - high distinct count of urls	985,037	ok	closed	stopped	2018-11-28 18:59:56	View Edit Delete
> filebeat-nginx-access-response_code	Nginx Access Logs: Detect unusual response_code rates	0	ok	closed	stopped		View Edit Delete
> filebeat-nginx-access-visitor_rate	Nginx Access Logs: Detect unusual visitor rate	2,688	ok	closed	stopped	2018-11-28 18:59:56	View Edit Delete

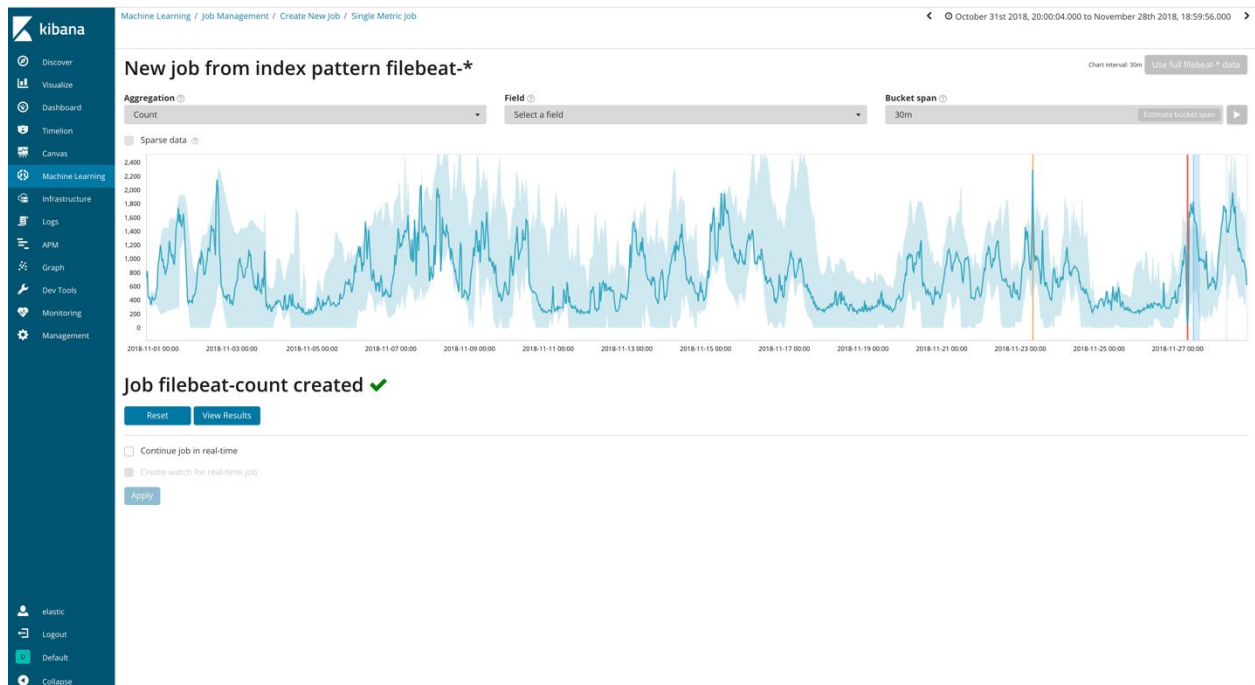
Select “filebeat*” index pattern and click on creating Single metric job.



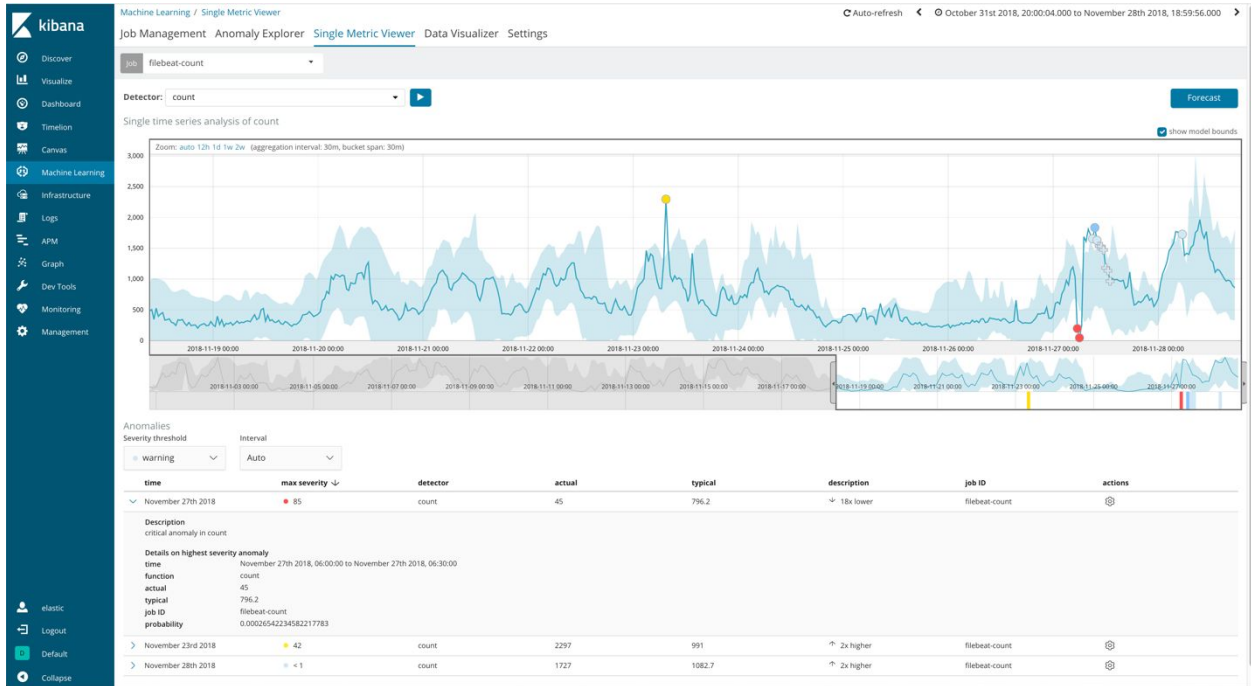
As function select “count”, leave field blank, set bucket span to 30m and click on “Use full filebeat-* data”. Give your job a name and click on “Create Job”.



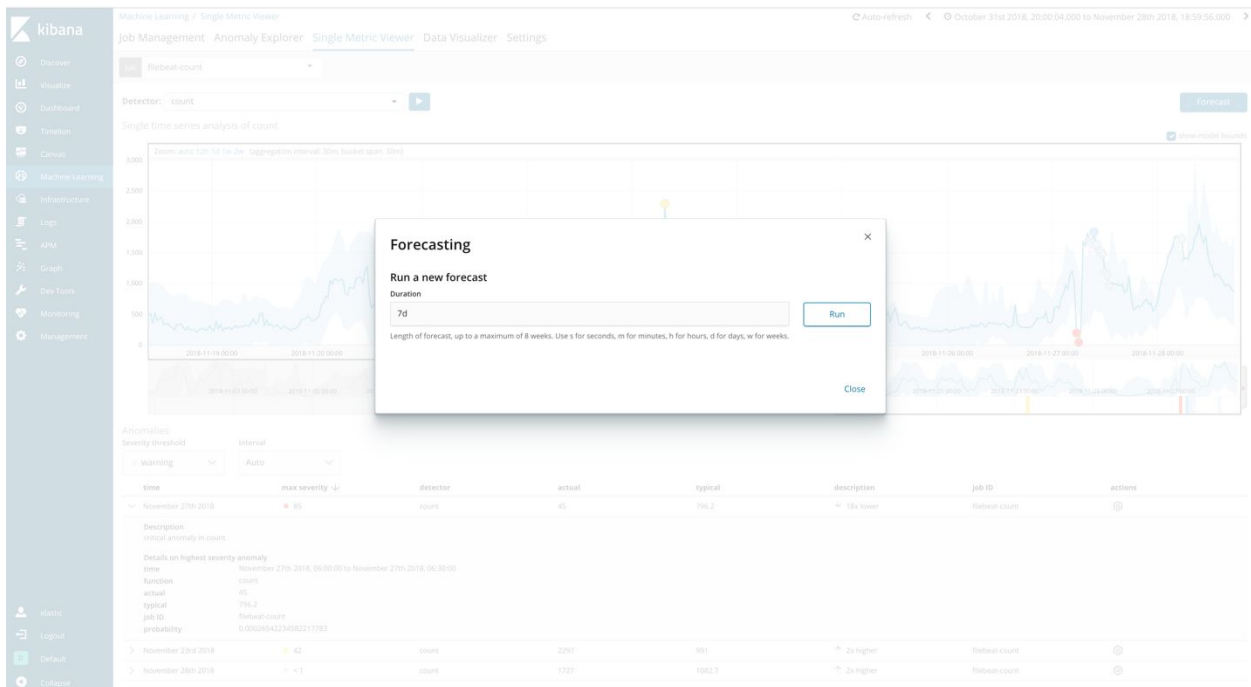
After the job completes click on “View Results”.



On the results screen slide the window in the Single Metrics viewer and click on the anomalies.
What does this anomaly represent? Remember that we've picked the function count of all events over 30min buckets.



Click on Forecast button and give it a length of 7d and click on Run.



View the forecasting result. The forecast part of the metrics and the model is in brown.

