



AI Stream Evolution Guide

DE Operations

Overview

What is the stream evolution modifications tool?

Tool automatically determines what modifications can be made to optimize streams

Why was this built?

Reduces stream volume while preserving alerts

How do we use it?



What does it cover?

Filters supported by Dataminr pipeline remove a specific, narrow portion of the matched content and work best on streams with a medium-to-high tweet matching rate (i.e. 5-30 tweets per minute) and a sufficient number of matching historical alerts

What is the AI Stream Evolution Project?

Introduces algorithmic tool that suggests and automatically applies filters to streams.

How to use it?

What Modifications are Evaluated?

- From main menu at [redacted] click “View Modifications” to see over 100 filters that can be applied
 - E.g. exclude tweets with [redacted] exclude tweets geo-located in Northern Europe, etc.
 - The details of what filter types used.

Most commonly used will be:

- **set-2019Q1**

- [redacted]
- [redacted]
- [redacted]
- [redacted]
- [redacted]


- **set-2019Q2**


- [redacted]
- [redacted]
- [redacted]

Submitting Streams for Evaluation

1. Login at: 

NOTE: need to be connected to VPN on Dataminr network

1. Go to “Create New Modification Report.”
2. Specify all fields.
 - a. Name group of streams, e.g. MENA and then add description
 - b. Apply tags by specifying filter set () and (optionally) a geographic specifier.
3. Set status as “Active” and Submit. Wait for processing to complete in minutes / hours.

Set identifier / Creation Time / Created By UID	Name / Description	Streams	Tags	Status	Update
vofvzkolqvjjwtrd5km 2019-07-17 18:54:58 UTC 26709	MENA MENA primary monitoring		set-2019Q1 × specifier-MENA × ...	I	<div>Edit</div> <div>Delete</div>

Interpreting a Stream Modification Report

1. When ready, a notice for each stream evaluated will appear in the #stream-evolution-noti slack channel.

In-progress stream reports appear in #stream-evolution-info.

1. Login at:



NOTE: need to be connected to VPN on Dataminr network

1. Go to “View Completed Stream Modification Reports” (can also be notified through slack channel “stream-evolution-noti”)

Stream Group Name	Stream Group Creation	Dataminr User	Stream Name	Evaluation Report	Tweets Per Hour	Alerts Per Hour	Best Modification Tweet Reduction %	Best Modification Alert Reduction %	Report Completion Date	Is Most Current Stream Version
MENA	2019-06-12 21:49:08 UTC	26709		link	47.62	0.89	39.9 %	3.8 %	2019-06-13 01:11:44 UTC	no



Interpreting a Stream Modification Report

The tool focuses on filters that reduce the tweet matching rate (**0.5% or more reduction in tweet matching**) while keeping the alert matching rate on historical alerts sufficiently high (**at least 95% alert matching**).

Best Modification Combination (based on maximum tweet reduction)

This is the best combination of modifications that should be applied to the stream definition to provide the best stream volume reduction discovered in modification search.



Original stream version performance:

Tweet rate: 47.62 tweets per hour

Alert rate: 0.89 alerts per hour

Modification set (count: 2)



The stream  is optimized when two modifications are applied, excluding 



Interpreting a Stream Modification Report

The best modification combination for the stream results in:

Performance with Modification Applied:
Tweet matching reduction: 11.42 per hour
Alert matching reduction: 0.03 per hour

11.42 fewer tweets per hour in stream, or a 23% reduction in stream volume (11.42 tweets per hour reduced / 47.62 tweets per hour of original stream = 23%)

Statistics on Tweets Excluded after Modification



Alerts excluded per month for Urgent National : 10.000
Alerts excluded per month for Signal : 13.000
Alerts excluded per month for Urgent Update National : 6.500
Alerts excluded per month for Urgent Local : 4.000
Alerts excluded per month for Urgent Global : 2.000
Alerts excluded per month for Urgent Update Global : 0.500

Alerts excluded per month 36.000
Tweets excluded per hour 6.333

If these modifications are applied, 36.00 alerts would not have come through this stream over a month. They may, however, come through other streams.

How to Apply Stream Modifications in

Raw Streams:

1. Go to  settings.
2. Search for Raw Stream. E.g. 
3. Select “Keywords”
4. In Universal Terms “Exclude” box, include modifications information from both “Raw Stream Common Filter Exclude” sections

Modification set (count: 2)

1.

Description:

Powerstream Component:

Rawstream Common Filter Exclude:

2.

Description:

Powerstream Component:

Rawstream Common Filter Exclude:



EXCLUDE

(mentionCount:range:1:100 and channel:chatter,) (hashtagCount:range:1:100)

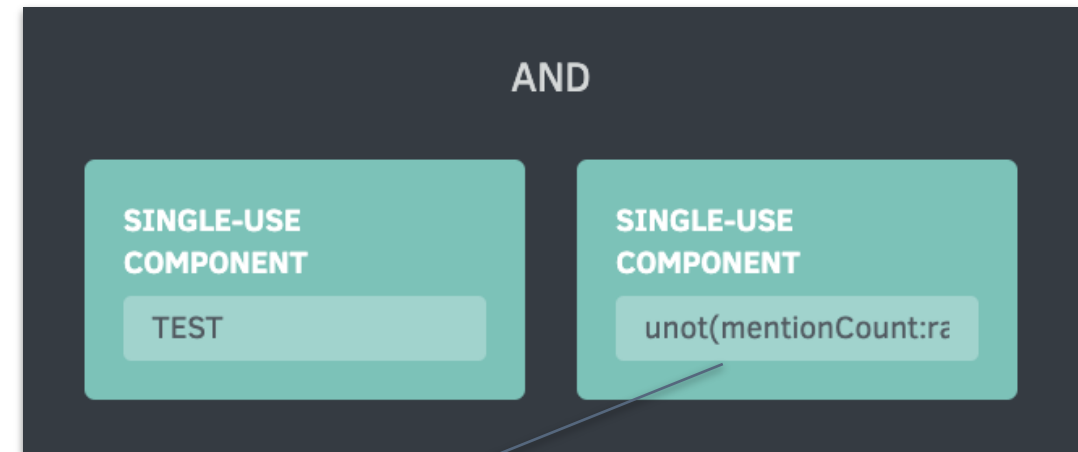


How to Apply Stream Modifications in

Power Streams:

1. Go to  settings.
2. Search for Power Stream. E.g. 
3. Add component with “AND” operator containing modifications from both “Powerstream Component” sections

```
Modification set (count: 2)
1.
    Description:
    Powerstream Component:
    Rawstream Common Filter Exclude:
2.
    Description:
    Powerstream Component:
    Rawstream Common Filter Exclude:
```



Considerations and Limitations

What does stream evolution do?

- The modification search is compatible with **modifications** that remove a **specific, narrow** portion of the matched content.
 - (e.g. tweet text with 3+ mentions, tweets geo-predicted to be in Southern Africa)
- All modifications must be supported by the Dataminr pipeline in order to be used here.


Stream Evolution works *well* on streams with:

- a medium-to-high tweet matching rate (i.e. 5-30 tweets per minute)
- a sufficient number of matching historical alerts

Stream Evolution works *less well* on?

- streams that target new content with few historical alerts

What does stream evolution not do?

- The search does not modify/simplify the overall boolean query.
- The search does not increase the matching (aka. *coverage*) of a stream, it only reduces matching (aka. *reduces noise*).
- 

Thank You

Questions? Email

