COMMSEC

HITB SECCONF
AMSTERDAM - 2021

One-Stop Anomaly Shop #OSAS

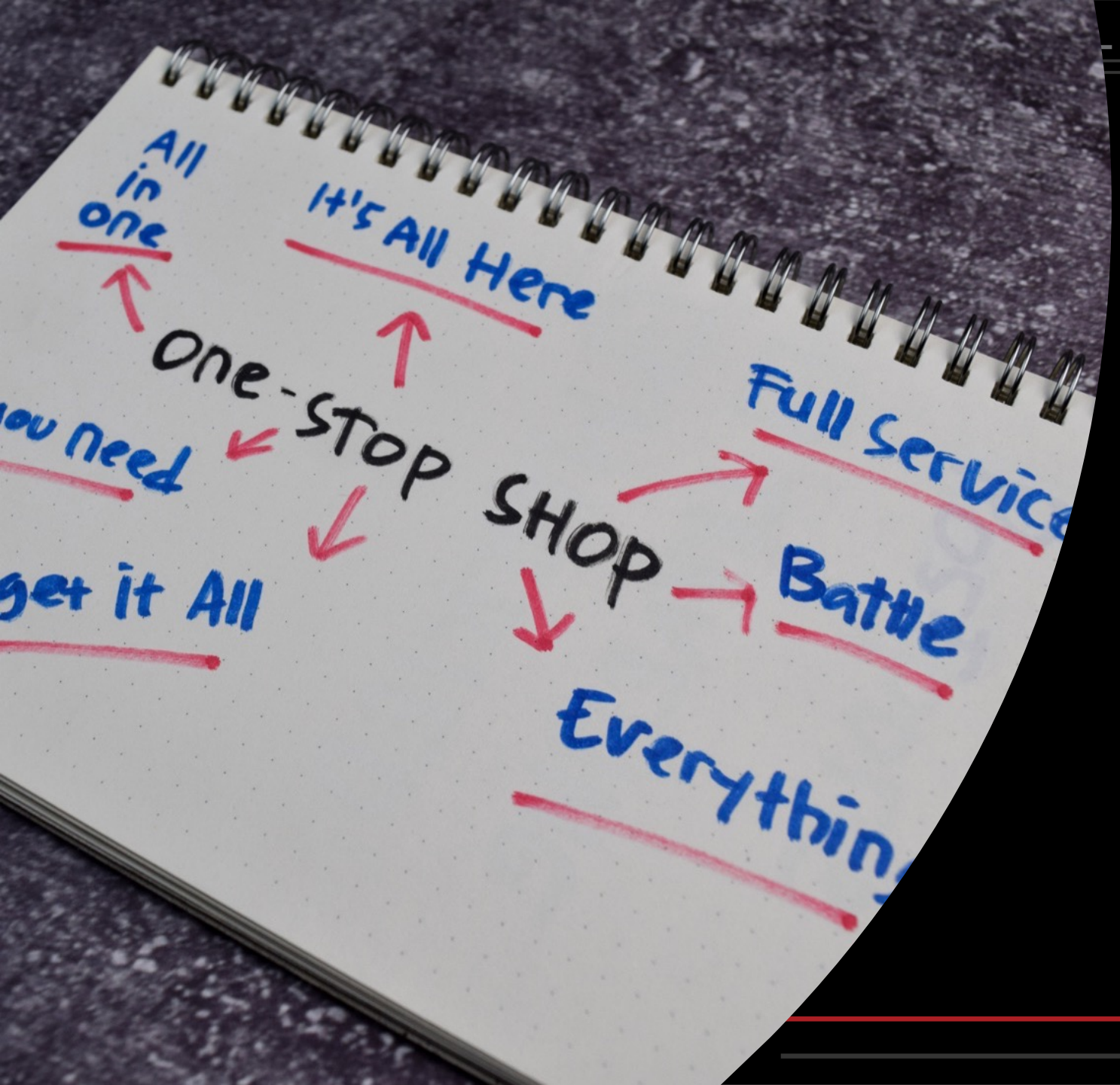
(now an open-source project)

Andrei Cotaie | Tiberiu Boros
Adobe | SCC | Security Intelligence

whoami

Adobe
Security
Intelligence
Team

- Andrei Cotaie
  - *Tech Lead, Security Intelligence*
- Tiberiu Boros
  - *Machine Learning Engineer*
- Kumar Vikramjeet
  - *Security Engineer*
- Vivek Malik
  - *Incident responder*

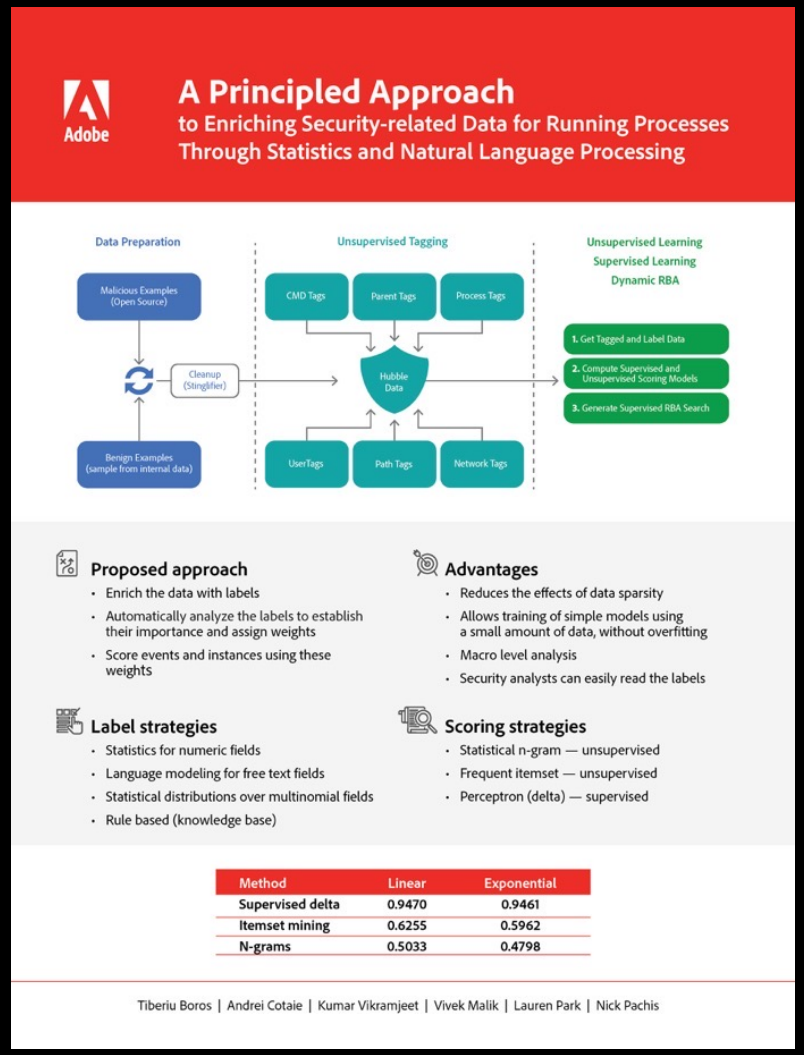


What's OSAS?!

The Anomaly Detection principle


Egg
White
Has mask
Open Eyes
Needs Eye
surgery
Egg
White
No mask
Open Eyes
Day
Dreaming
Egg
White
Has mask
Open Eyes
Cries
Desperation
Egg
White
No mask
Open Eyes
Arogant
Egg
White
Has mask
Open Eyes
Scared
Egg
White
Has mask
Open Eyes
Angry
Egg
White
Has mask
Open Eyes
In Love OR
Chicken Pox?
Looks Surprised
Egg
White
Has mask
Open Eyes
In Love OR
Chicken Pox?
Egg
White
Sneezes
No mask
Closed Eyes
Egg
Brown
Tick Eyebrowns
Has mask
Open Eyes

So… it’s all about the TAGS (aka Label Generators)



Adobe

A Principled Approach
to Enriching Security-related Data for Running Processes Through Statistics and Natural Language Processing

Proposed approach
- Enrich the data with labels
- Automatically analyze the labels to establish their importance and assign weights
- Score events and instances using these weights

Advantages
- Reduces the effects of data sparsity
- Allows training of simple models using a small amount of data, without overfitting
- Macro level analysis
- Security analysts can easily read the labels

Label strategies
- Statistics for numeric fields
- Language modeling for free text fields
- Statistical distributions over multinomial fields
- Rule based (knowledge base)

Scoring strategies
- Statistical n-gram — unsupervised
- Frequent itemset — unsupervised
- Perceptron (delta) — supervised

Method	Linear	Exponential
Supervised delta	0.9470	0.9461
Itemset mining	0.6255	0.5962
N-grams	0.5033	0.4798

Tiberiu Boros | Andrei Cotaie | Kumar Vikramjeet | Vivek Malik | Lauren Park | Nick Pachis

Standard Tags :

- Multinomial fields
- Numeric fields
- Multinomial Combiner

Text Tags

- Text/NLP processing

Expert Knowledge

- Keyword identification
- Regex matching

Tag generators 1/2

Multinomial fields

Detection:

- Count unique attribute values
- Less than 10% unique values

Model

- Statistical distribution of values
- Labels based on value frequency
- Special tags for unseen data

Linking

- Model requires training
- Stored

Numerical fields

Detection:

- Count unique attribute values
- More than 10% unique values
- All values must be numerical or None

Model

- Mean an standard deviation
- Labels based on Gaussian probability

Linking

- Model requires training
- Stored

Field combiners

Detection:

- All multinomial fields can be combined

Model

- Statistical distribution of values
- Labels based on value frequency
- Special tags for unseen data

Linking

- Model requires training
- Stored

Tag generators 2/2

Text fields

Detection:
- Non-numerical
- Count unique attribute values
- More than 10% unique values

Model
- Compute n-gram language model
- Compute perplexity for each example
- Mean an standard deviation
- Labels based on Gaussian probability

Linking
- Model requires training
- Stored

Expert knowledge

Detection:
- Manual

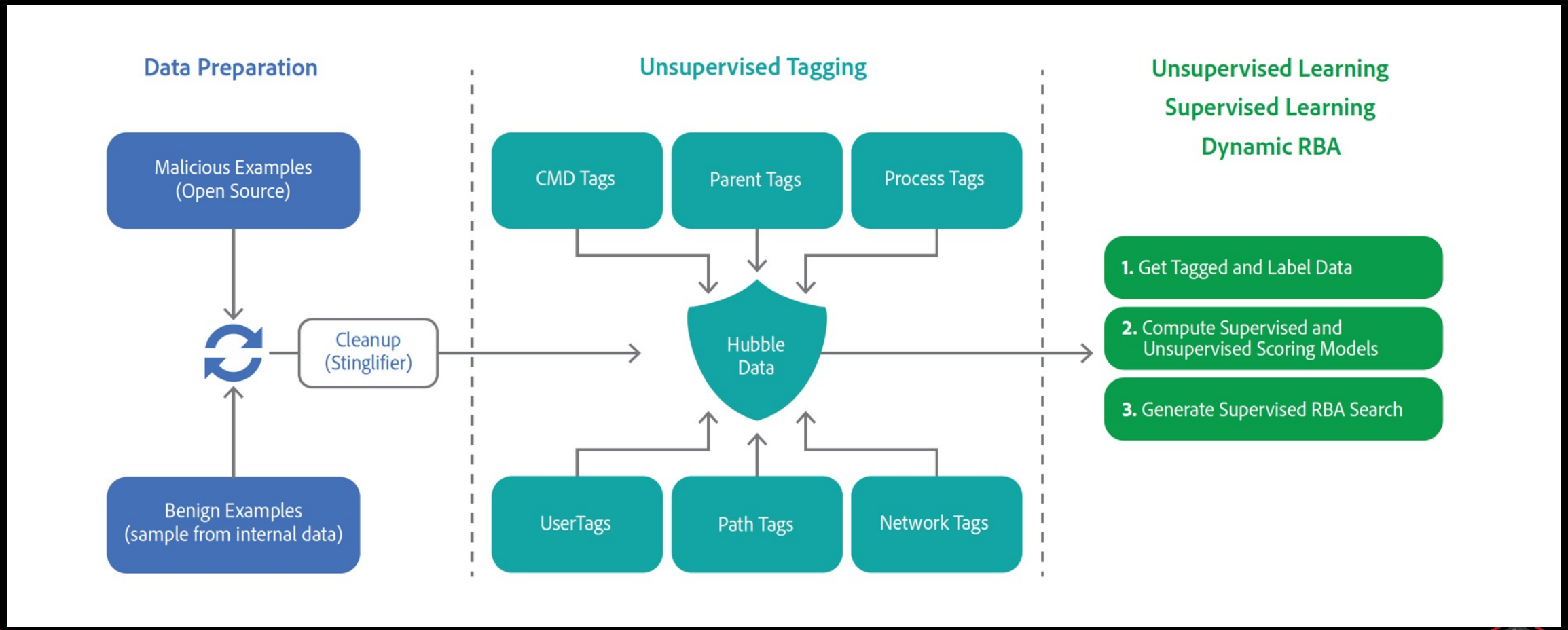
Model
- Keyword or regex-based
- Labels for matched instances

Linking
- No training/storing required

Anomaly Detection Algorithms

- **Isolation Forest**
- **Local Outlier Factor**
- **SVD based Anomaly Detector**
- **Statistical Ngram Anomaly**

All in a Nutshell



Demo Time!


A B C D E F G
H I J K L M N
O P Q R S T U
V W X Y Z
1 2 3 4 5 6
7 8 9 0
WATCH
LIVE
A B C D E F G
H I J K L M N
O P Q R S T U
V W X Y Z
1 2 3 4 5 6
7 8 9 0

1. Get **OSAS** up and running

adobe / OSAS

Watch 5 Star 11 Fork 0

Code Issues Pull requests Actions Projects Wiki Security Insights

main 1 branch 0 tags Go to file Add file Code

Tiberiu Boros update 4566fe6 24 minutes ago 13 commits

docker/osas-elastic	Switched to https	14 days ago
osas	update	24 minutes ago
resources	First commit	14 days ago
scripts	Kibana dashboard	1 hour ago
.gitignore	First commit	14 days ago
CODE_OF_CONDUCT.md	First commit	14 days ago
CONTRIBUTING.md	First commit	14 days ago
COPYRIGHT	First commit	14 days ago
LICENSE	First commit	14 days ago
README.md	Update README.md	14 days ago

About

One Stop Anomaly Shop: Anomaly detection using two-phase approach: (a) pre-labeling using statistics, Natural Language Processing and static rules; (b) anomaly scoring using supervised and unsupervised machine learning.

Readme

View license

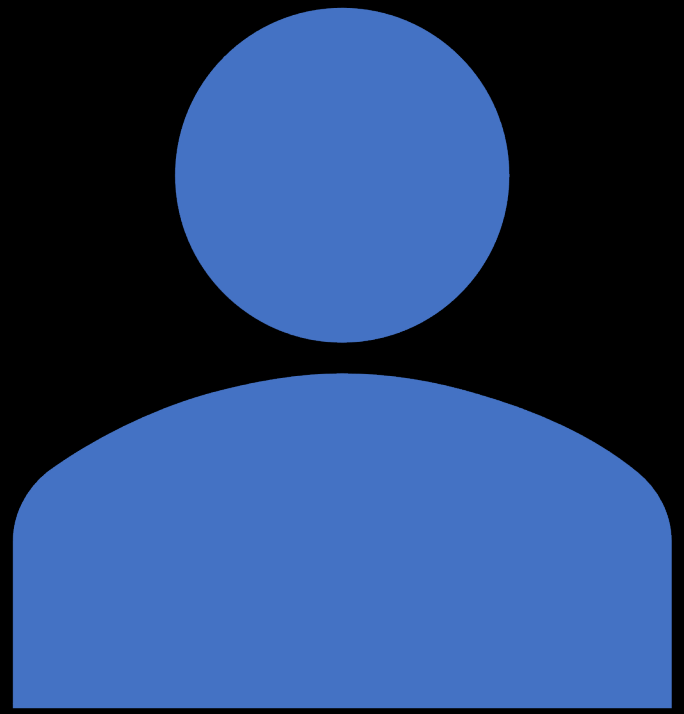
Releases

No releases published

Packages

No packages published

2. Default **OSAS** on a dataset - **Walkthrough**



Andrei Cotaie | Tiberiu Boros | Security Intelligence

adobe/OSAS: One Stop Anoma

github.com/adobe/OSAS

Code | Issues | Pull requests | Actions | Projects | Wiki | Security | Insights

main | 1 branch | 0 tags | Go to file | Add file | Code

Tiberiu Boros update — 4566fe6 17 hours ago — 13 commits

docker/osas-elastic	Switched to https	15 days ago
osas	update	17 hours ago
resources	First commit	15 days ago
scripts	Kibana dashboard	18 hours ago
.gitignore	First commit	15 days ago
CODE_OF_CONDUCT.md	First commit	15 days ago
CONTRIBUTING.md	First commit	15 days ago
COPYRIGHT	First commit	15 days ago
LICENSE	First commit	15 days ago
README.md	Update README.md	15 days ago
requirements.txt	unused class	15 days ago

README.md

One Stop Anomaly Shop (OSAS)

About

One Stop Anomaly Shop: Anomaly detection using two-phase approach: (a) pre-labeling using statistics, Natural Language Processing and static rules; (b) anomaly scoring using supervised and unsupervised machine learning.

Readme

View license

Releases

No releases published

Packages

No packages published

Languages

Python 83.5% HTML 13.3% Dockerfile 2.2% Shell 1.0%

3. **OSAS** and Expert Knowledge Labels!

IMPORTANT NOTE: Please modify the above command by adding the absolute path to your datafolder in the appropiate location

After OSAS has started (it might take 1-2 minutes) you can use your browser to access some standard endpoints:

- http://localhost:5601/app/home#/ - access to Kibana frontend (this is where you will see your data)
- http://localhost:8888/osas/console - command-line access to osas scripts and utilities

For Debug (in case you need to):

```
docker run -p 8888:8888/tcp -p 5601:5601/tcp -v <ABSOLUTE PATH TO DATA FOLDER>:/app -ti osas /bin/ba
```

Building the test pipeline

This guide will take you through all the necessary steps to configure, train and run your own pipeline on your own dataset.

Prerequisite: Add you own CSV dataset into your data-folder (the one provided in the `docker run` command)

Once you started your docker image, use the OSAS console to gain CLI access to all the tools.

In what follows, we assume that your dataset is called `dataset.csv`. Please update the commands as necessary in case you use a different name/location.

Be sure you are running scripts in the root folder of OSAS:

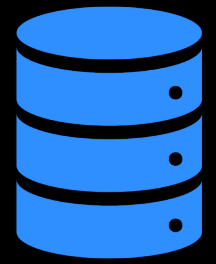
Take Aways

OSAS

Fast Deploy

Easy to use

Easy to save and repeat



In general:

Not all features provide VALUE

Each anomaly algorithm has its own ups and downs


QUESTIONS

How to Get and Use OSAS

- GitHub:
  - https://github.com/adobe/OSAS
- Original Blogpost:
  - https://medium.com/adobetech/introducing-the-one-stop-anomaly-shop-osas-c27581ee1bd3
- Docker:
  - docker pull tiberiu44/osas:latest

Resources

Adobe Security Newsletter
adobe.com/go/securitynews

Twitter
@AdobeSecurity

Trust Center
trust.adobe.com

Open Source CCF v4.0
adobe.com/go/open-source-ccf

Security @ Adobe blog
https://medium.com/adobetech/tagged/security

Thank You !

For your attention

https://github.com/adobe/OSAS

#giveOSASastar

#AdobeSecurity