# DS/AI for Incident Response & Threat Hunting with CHRYSALIS & DAISY

**Jess Garcia**

One eSecurity – Founder
SANS – Senior Instructor

jess@one-esecurity.com - @j3ssgarcia

ODSC WEST 2022

# $ Whoami

**Jess Garcia**
@j3ssgarcia

- Founder and CEO of One eSecurity
  +25 years of experience in CybSec / DFIR
- Global DFIR company for 15 years
  www.one-esecurity.com
- Leader of the DS4N6 Project since 2020
  www.ds4n6.io
- Senior Instructor at the SANS Institute
  ~ 20 years

# Our Objective

**Transform you into AI-Enhanced Threat Hunters/Forensicators to bring the power of AI in your day to day investigations.**

You do not need to be an AI expert, you will need **to learn what AI can do for you**, becoming familiar with the tools available and how to use them to suit their needs.

# The Big Question

AI is great. But, what can it do for a Threat Hunter / Forensicator?

Would it be able to detect

**Cobalt Strike?**

What else can it do?

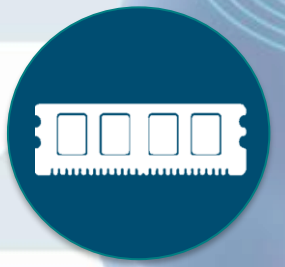
# ML FOR DFIR USE CASES

# Where Can We Use Machine Learning in DFIR?

DF

TH - CHRYSALIS

CTI



Memory Analysis – Columbo

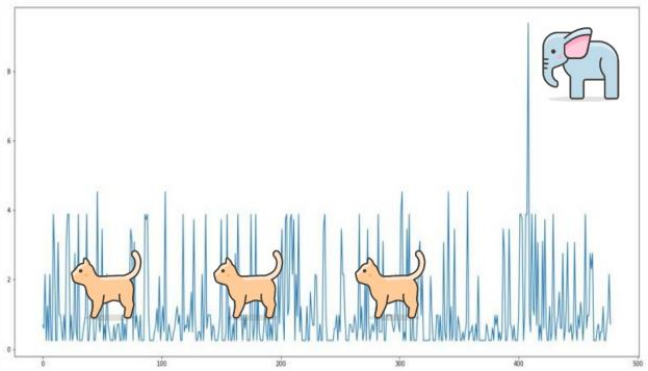
Malware – Malware Revealer

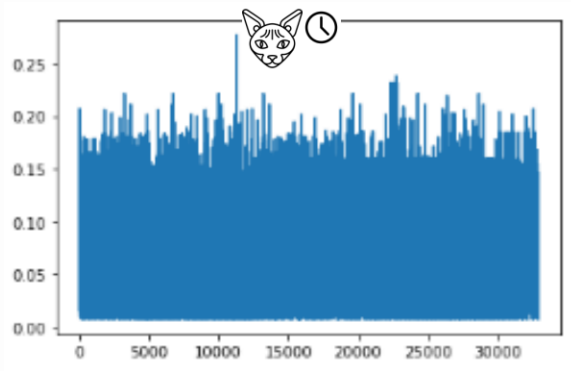
Logs - Deeplog

Network - Zeek

# ML & TH: Artifact Anomalies

| Scheduled Tasks | Scheduled Tasks |
|---|---|
| No time sequence | Time sequence is important |
| Vanilla Autoencoder | LSTM Autoencoder |







ds4n6.io/rsac21

| | level_0 | Orig_Index | EventID_ | AtName_ | TaskName_ | AtUserID_ | ResultCode_ | ActionName_ | UserNC_ | Hostname_ |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 676274 | 676473 | 140 | TaskUpdated | \Microsoft\Windows\SoftwareProtectionPlatform\... | S-1-5-18 | None | None | d4_null\system$ | mc80-sc-7813 |
| 1 | 676273 | 676472 | 106 | TaskRegisteredEvent | \Microsoft\Windows\SoftwareProtectionPlatform\... | S-1-5-18 | None | None | d4_null\rice.berav$ | mc80-sc-7813 |
| 2 | 670275 | 670474 | 106 | TaskRegisteredEvent | \TratarTrazas | S-1-5-18 | -64646464 | d4_null | scpd02mq01\adm_sna | xwt70-sf-2560 |
| 3 | 670273 | 670472 | 106 | TaskRegisteredEvent | \SyncFolder | S-1-5-18 | -64646464 | d4_null | scpd02mq01\adm_sna | xwt70-sf-2560 |
| 4 | 670271 | 670470 | 106 | TaskRegisteredEvent | \RestartDocpath | S-1-5-18 | -64646464 | d4_null | scpd02mq01\adm_sna | xwt70-sf-2560 |
| 5 | 676275 | 676474 | 200 | ActionStart | \Microsoft\Windows\SoftwareProtectionPlatform\... | S-1-5-18 | None | C:\Windows\SoftwareProtectionPlatform\EventCac... | d4_null\rice.berav$ | mc80-sc-7813 |
| 6 | 666222 | 666421 | 140 | TaskUpdated | \Microsoft\Windows\Customer Experience Improve... | S-1-5-18 | -64646464 | d4_null | d4_null\xwt70-sf-9087$ | mc80-sc-6106 |
| 7 | 665394 | 665593 | 140 | TaskUpdated | \Microsoft\Windows\Customer Experience Improve... | S-1-5-18 | -64646464 | d4_null | d4_null\xwt70-sf-9087$ | mc80-sc-6106 |

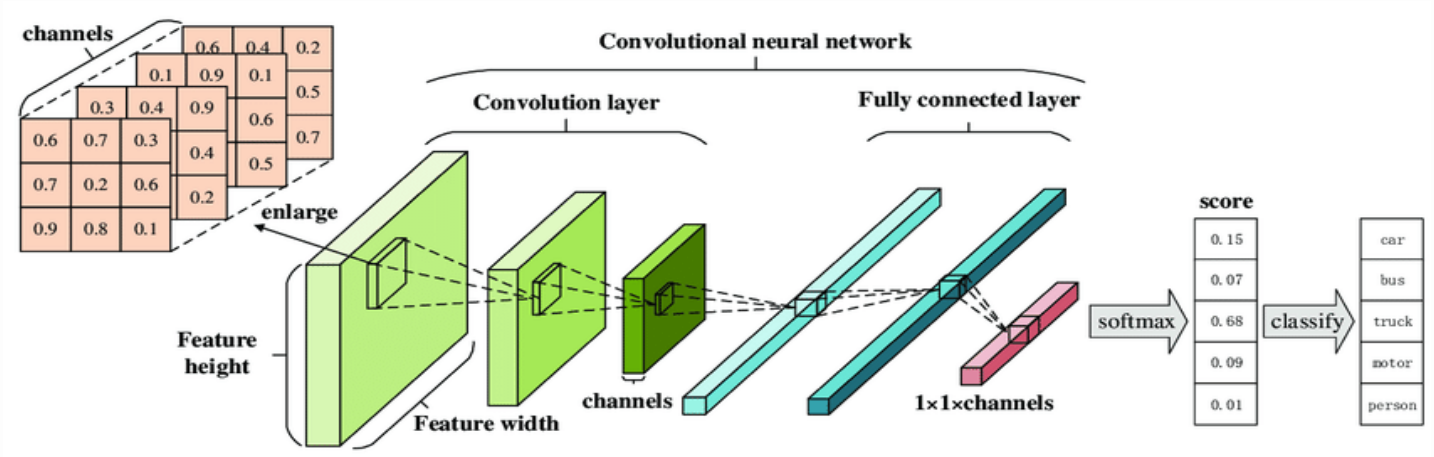
# ML & Malware: Detection and Classification

## Convolutional Neural Networks (CNN)

Raw bytes

Interesting patterns

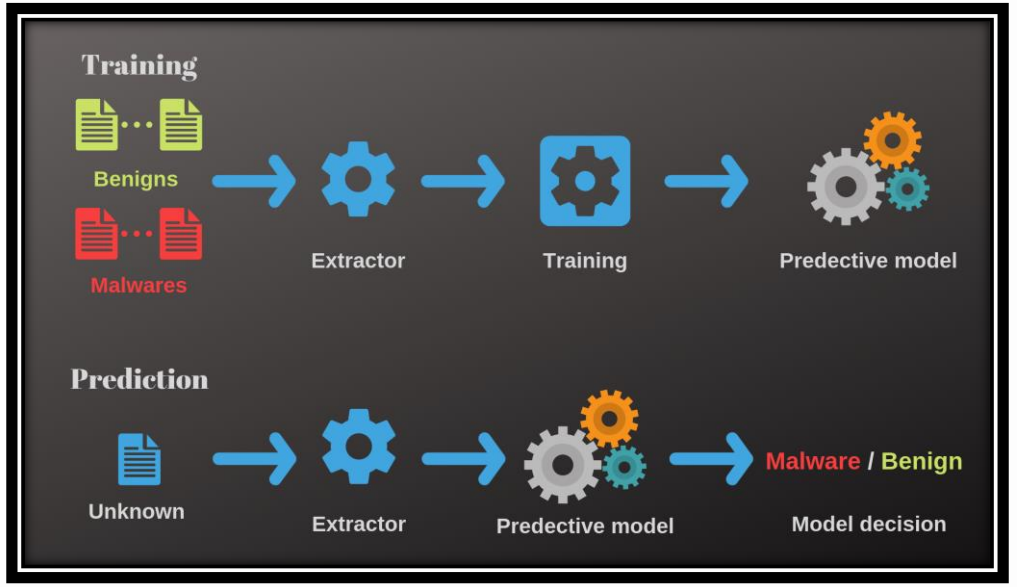
Classification




https://www.mcafee.com/blogs/other-blogs/mcafee-labs/the-rise-of-deep-learning-for-detection-and-classification-of-malware/

# ML & Malware: Malware Revealer

- Malware detection using ML with pre-trained models
- Uses SqueezeNet and Logistic Regression models
- Extracts features using convolutional filters to classify them as malware




https://www.ayoub-benaissa.com/blog/malware-revealer/

# ML & Memory Forensics: Columbo

- Used to identify specific patterns in compromised datasets
- It uses Volatility 3 outputs applying ML algorithms to look for suspicious
- You can use it with pslist, psscan, pstree, malfind, netscan, etc.


```
Information about process Number 3496

Possible process path or execution: C:\Users\Bob\AppData\Local\Temp\rad93398.tmp\UWkpjFjDzM.exe

Machine Learning model classifies  C:\Users\Bob\AppData\Local\Temp\rad93398.tmp\UWkpjFjDzM.exe  to be  suspicious. Please consider its percentage scores shown below:
   0    1
15.1 84.9

Process traceability coupled with time executions of each process

process UWkpjFjDzM.exe(3496)/2019-03-22-05:35:33.000000 executed by
 wscript.exe(5116)/2019-03-22-05:35:32.000000 <- hfs.exe(3952)/2019-03-22-05:34:51.000000 <- explorer.exe(1432)/2019-03-22-05:32:07.000000 root process is  1308


3496 is a parent process of the following process(es):

ImageFileName PPID  PID
      cmd.exe 3496 4660
```

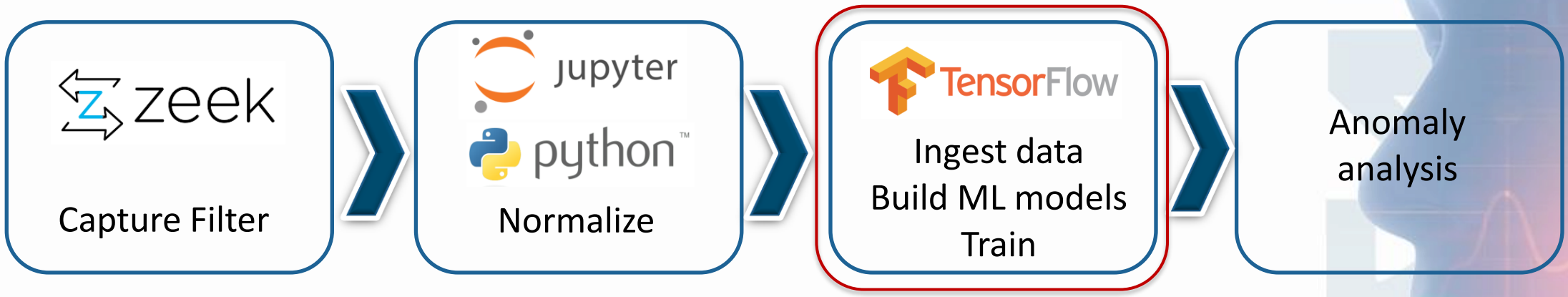
# ML & Logs: Deeplog

- It learns from tagged data to classify as anomaly or normal entry
- It helps to identify anomalies, using LSTM in large volumes of system logs
- Used in IDS/Firewall logs to detect DDoS and Port scans

```
(deeplog_env) ds4n6@daisy:~/Downloads/deeplog_tests$ sh train.sh
[Epoch  1/10] average loss = 8.0148 ######################################## (100.00%) runtime 0:00:04.2
[Epoch  2/10] average loss = 8.0144 ######################################## (100.00%) runtime 0:00:03.6
[Epoch  3/10] average loss = 8.0140 ######################################## (100.00%) runtime 0:00:03.8
[Epoch  4/10] average loss = 8.0136 ######################################## (100.00%) runtime 0:00:03.0
[Epoch  5/10] average loss = 8.0132 ######################################## (100.00%) runtime 0:00:02.6
[Epoch  6/10] average loss = 8.0128 ######################################## (100.00%) runtime 0:00:02.5
[Epoch  7/10] average loss = 8.0124 ######################################## (100.00%) runtime 0:00:02.8
[Epoch  8/10] average loss = 8.0120 ######################################## (100.00%) runtime 0:00:02.8
[Epoch  9/10] average loss = 8.0116 ######################################## (100.00%) runtime 0:00:04.8
[Epoch 10/10] average loss = 8.0112 ######################################## (100.00%) runtime 0:00:02.9
```

# ML & Network Traffic: Zeek

- Customized in-depth monitoring far beyond the capabilities of traditional systems
- Perform clustering to find anomalies, setting apart outliers
- We can find threats in large data sets even when they're unknown




**David Hoelzer.** *Author of:*

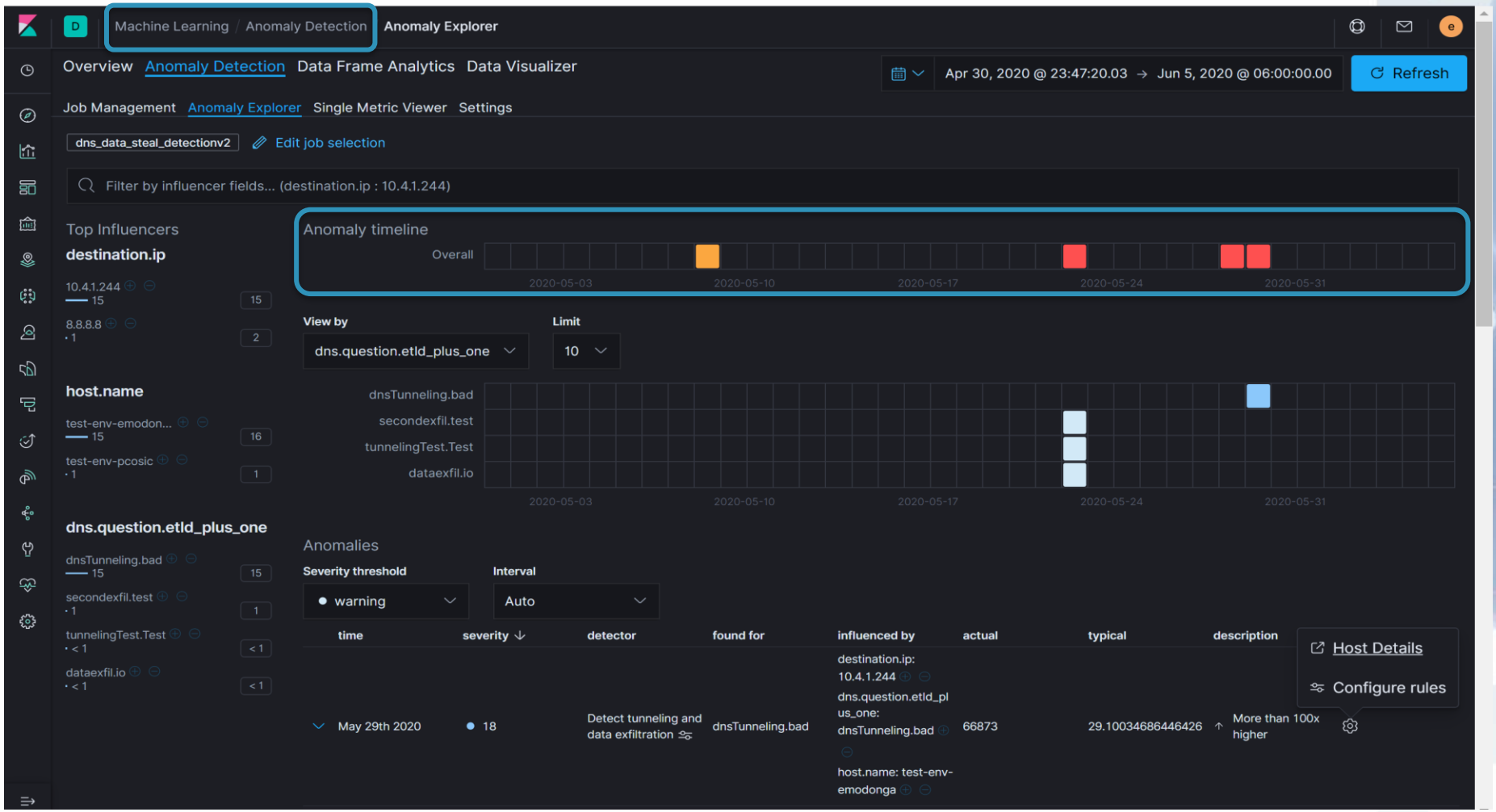
- ***SEC503****: Intrusion Detection In-Depth.*
- ***SEC595****: Applied Data Science and AI/Machine Learning for Cybersecurity Professionals.*

**Threat Hunting: Old Data New Tricks!**
https://www.youtube.com/watch?v=OCTz62fN8OA

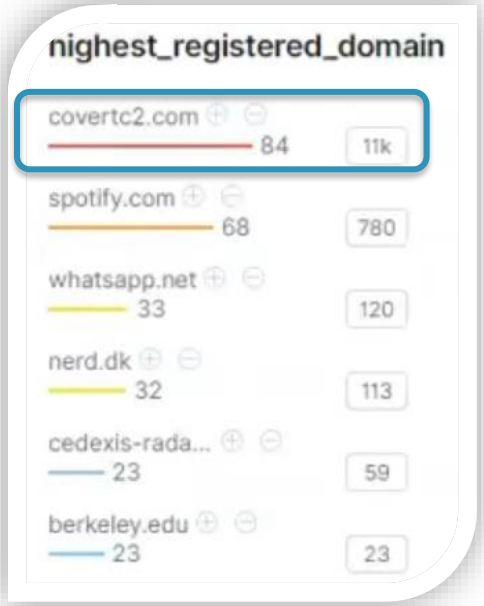
**Applying Machine Learning to Network Anomalies:**
https://www.youtube.com/watch?v=qOfgNd-qijI

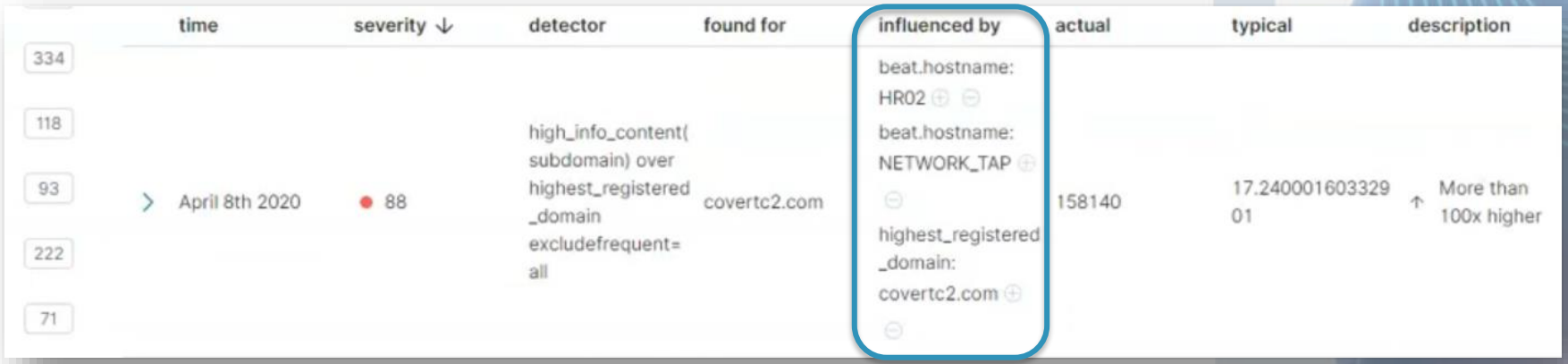
# ML & DF: Elastic

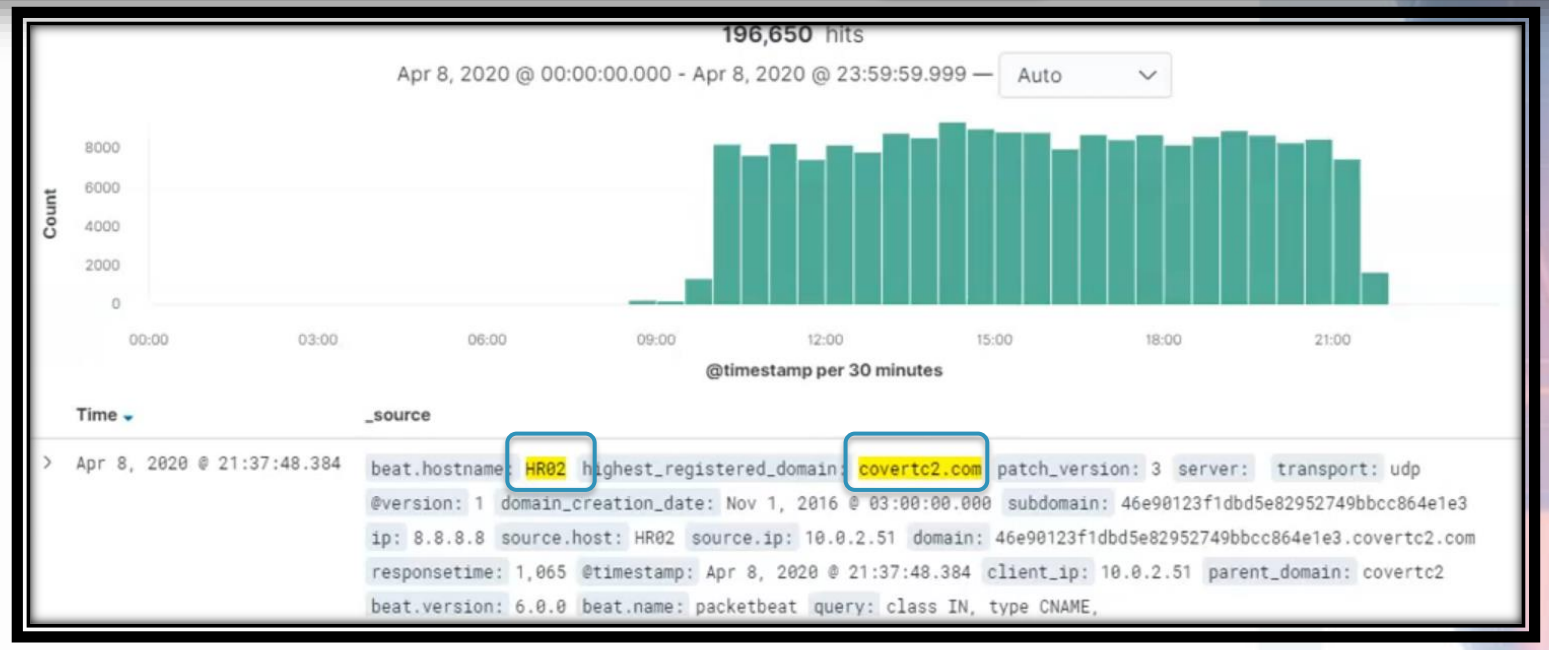
**The Elastic Observability and Security solutions have preconfigured machine learning models**



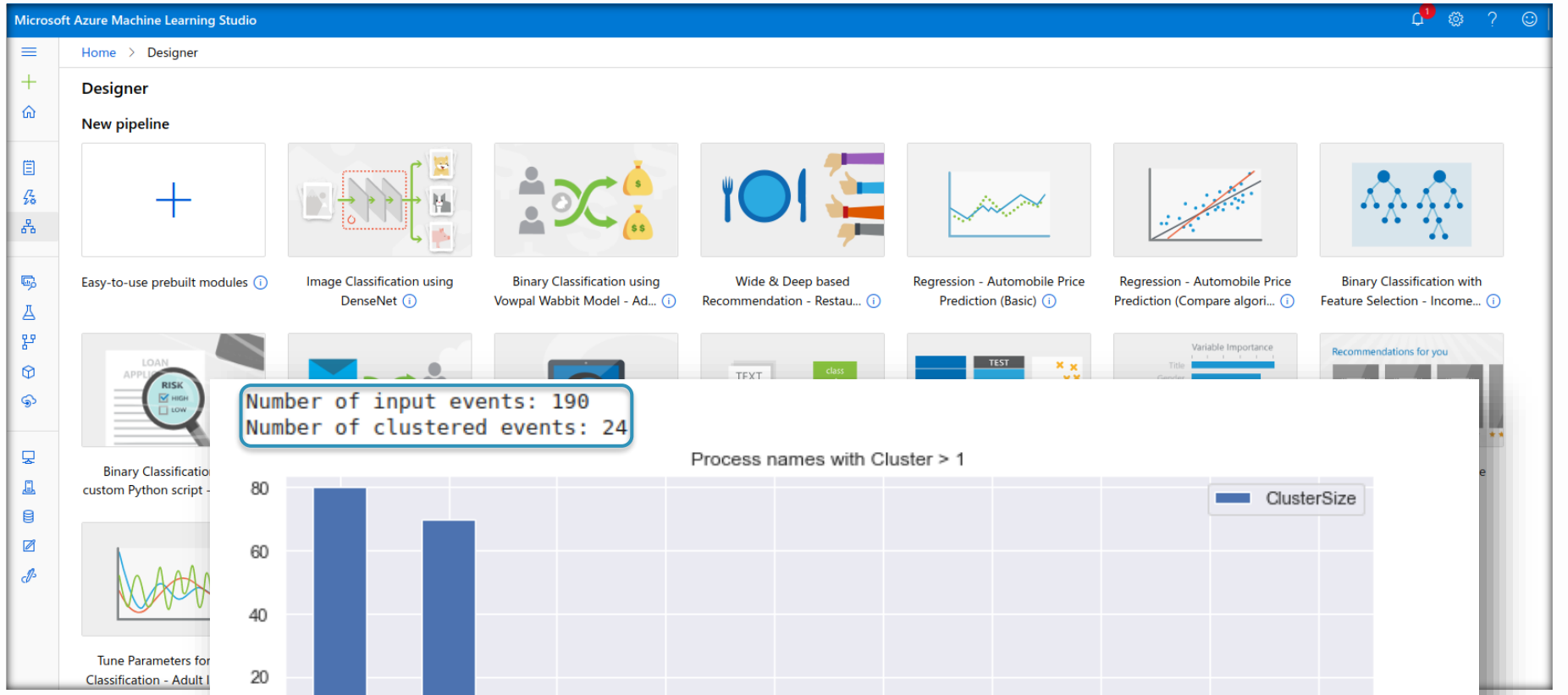
# ML & DF: Elastic – Use Case: DNS Exfiltration



# ML on the Cloud: MSTICPy and Azure







https://github.com/microsoft/msticpy

https://github.com/Azure/Azure-Sentinel


DS4N6

# Putting All Together: DS4N6

**Mission:** **Bring Data Science & Artificial Intelligence to the fingerprints of the average Forensicator and promote advances in the field**

**ds4n6.io**

**Since 2020**

**Presented in**

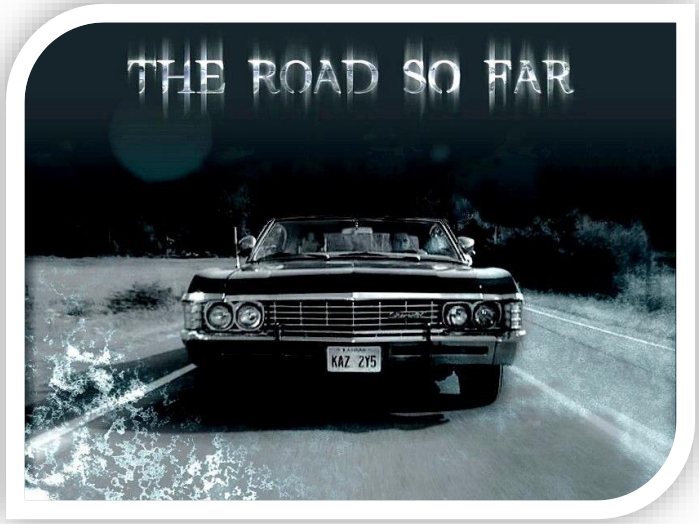





# DS4N6: The Road So Far

DS4N6

ds4n6.io




CHRYSALIS

D4ML

EVTX
PF
MFT
REG
WEB

HAM

ADAM

ADversAry eMulator

DAISY

Daisy VM

# CHRYSALIS

## Python framework that provides DS/ML functions to use without any specific DS/ML knowledge

**More information in:**
**www.ds4n6.io/chrysalis**

## Complete your investigations with only 7 functions!

CORE FUNCTIONS

| Function | Usage | Type | Description |
|---|---|---|---|
| **whatis()** | whatis(obj) | CLI | Identifies the forensic data type of an object (DataFrame -df- or DataFrame Collection -dfs-) |
| **xread()** | xread(options) | GUI | Reads tool output data (e.g. plaso output) and stores it in a df/dfs |
| **xmenu()** | xmenu(obj) | GUI | Used to easily select a dataframe from dfs, or a column from a df, displaying the selected data and allowing manual (Excel-like) analysis on it |
| **xanalysis()** | xanalysis(obj, options) | GUI | Displays a menu with the advanced analysis functions available for the data type (i.e. forensic artifact) given |
| **xdisplay()** | xdisplay() | GUI | Used to select the display settings for the dataframes that will be displayed (max. rows, max. columns, etc.) |
| **simple()** | df.simple(options) | CLI | Simplifies forensic output (df) showing only the most interesting columns for analysis. |
| **xgrep()** | xgrep(obj, options) | CLI | UNIX-like grep for the DataFrame world. Allows the user to search for a regular expression in a DF column or full DF |

# Try CHRYSALIS on the Cloud: Colab & Binder

ODSC_TheStolenSzechuanSauceCase.ipynb

File Edit View Insert Runtime Tools Help Cannot save changes

Share DS

Table of contents



+ Code + Text Copy to Drive

RAM Disk Editing

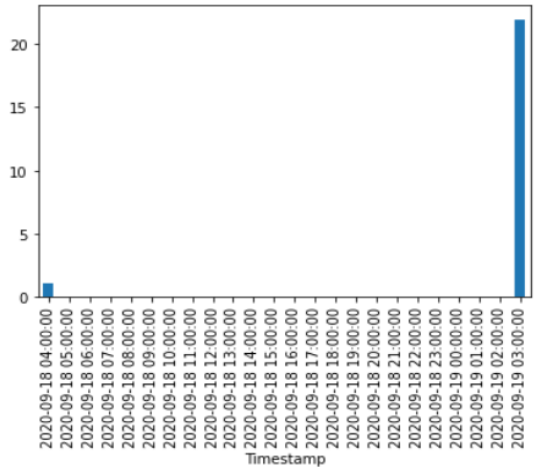
- Select 4625 as DF to analyze
- Select Failed Logons info as the available analysis

```
xanalysis(secevtxdfs_srv)
```

**Analisys explorer:**

Analysis object: DataFrame Analysis type: evtx DF to analyze: 4625

Available analysis types: Failed Logons info Export Result to d4.out




**Failed Logons**

| Timestamp | |
|---|---|
| 2020-09-18 | 1 |

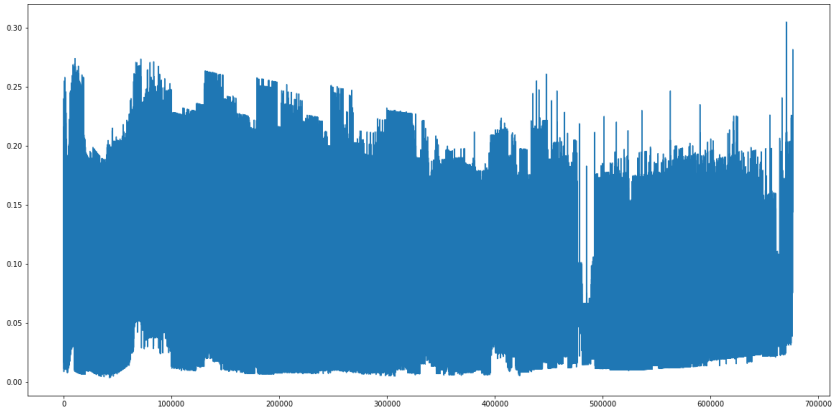
Try Colab now:
bit.ly/3Ff2VOm

Try Binder now:
bit.ly/3Ff2VOm

0s completed at 1:48 PM

# D4ML

**Easy-to-use ML functions that you can apply to your artifact dataframes.**
**It can be implemented stand-alone or via xanalysis()**



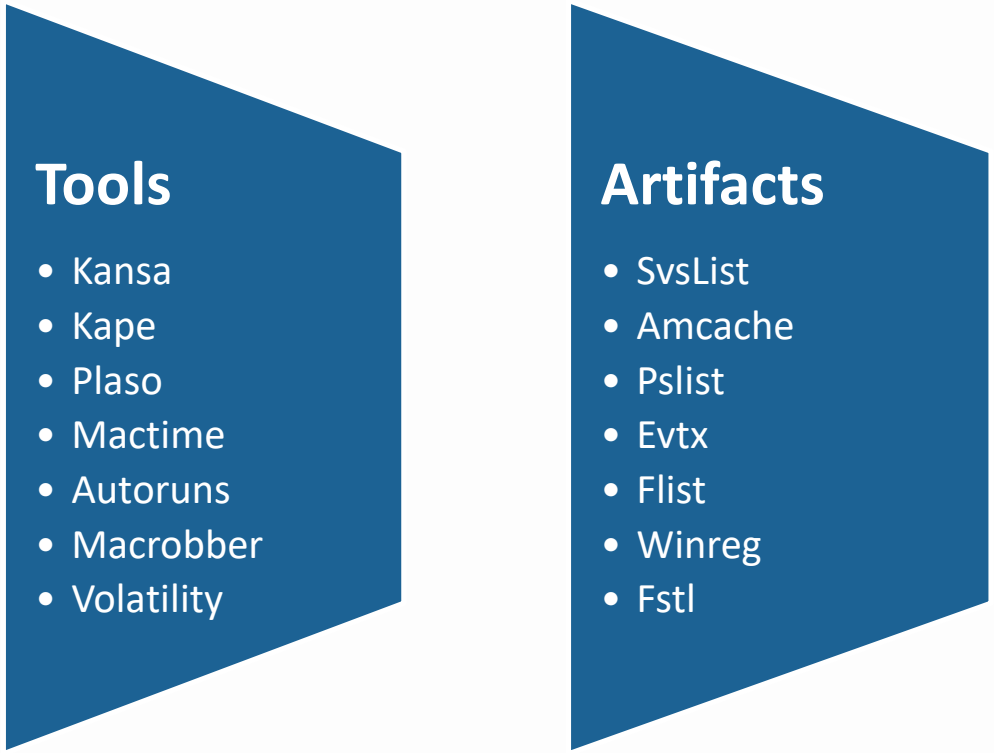

**find_anomalies()**
D4ML function to find anomalies via ML without knowing ML

**More information in:**
**www.ds4n6.io/d4ml**

| | level_0 | Orig_Index | EventID_ | AtName_ | TaskName_ | AtUserID_ | ResultCode_ | ActionName_ | UserNC_ | Hostname_ |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 676274 | 676473 | 140 | TaskUpdated | \Microsoft\Windows\SoftwareProtectionPlatform\... | S-1-5-18 | None | None | d4_null\system$ | mc80-sc-7813 |
| 1 | 676273 | 676472 | 106 | TaskRegisteredEvent | \Microsoft\Windows\SoftwareProtectionPlatform\... | S-1-5-18 | None | None | d4_null\rice.berav$ | mc80-sc-7813 |
| 2 | 670275 | 670474 | 106 | TaskRegisteredEvent | \TratarTrazas | S-1-5-18 | -64646464 | d4_null | scpd02mq01\adm_sna | xwt70-sf-2560 |
| 3 | 670273 | 670472 | 106 | TaskRegisteredEvent | \SyncFolder | S-1-5-18 | -64646464 | d4_null | scpd02mq01\adm_sna | xwt70-sf-2560 |
| 4 | 670271 | 670470 | 106 | TaskRegisteredEvent | \RestartDocpath | S-1-5-18 | -64646464 | d4_null | scpd02mq01\adm_sna | xwt70-sf-2560 |
| 5 | 676275 | 676474 | 200 | ActionStart | \Microsoft\Windows\SoftwareProtectionPlatform\... | S-1-5-18 | None | C:\Windows\SoftwareProtectionPlatform\EventCac... | d4_null\rice.berav$ | mc80-sc-7813 |
| 6 | 666222 | 666421 | 140 | TaskUpdated | \Microsoft\Windows\Customer Experience Improve... | S-1-5-18 | -64646464 | d4_null | d4_null\xwt70-sf-9087$ | mc80-sc-6106 |
| 7 | 665394 | 665593 | 140 | TaskUpdated | \Microsoft\Windows\Customer Experience Improve... | S-1-5-18 | -64646464 | d4_null | d4_null\xwt70-sf-9087$ | mc80-sc-6106 |

# HAM / HAMML

**Model that harmonizes the output of different tools so the underlying artifact data has the same format regardless of the tool that generated it**




**More information in:**
**www.ds4n6.io/ham**

**HAMML: HAM + Feature Selection + Feature Engineering**

# HAM / HAMML

**Unharmonized DataFrame**

**xread()**

**Harmonized DataFrame**

```
[10]: plaso_JSON.head()
```

[10]:

| | event_0 | event_1 | event_2 | event_3 | event_4 | event_5 | event_6 | |
|---|---|---|---|---|---|---|---|---|
| __container_type__ | event | event | event | event | event | event | event | |
| __type__ | AttributeContainer | AttributeContainer | AttributeContainer | AttributeContainer | AttributeContainer | AttributeContainer | AttributeContainer | |
| build_number | 9600 | NaN | NaN | NaN | NaN | NaN | NaN | |
| data_type | windows:registry:installation | windows:shell_item:file_entry | windows:shell_item:file_entry | windows:shell_item:file_entry | windows:shell_item:file_entry | windows:shell_item:file_entry | windows:shell_item:file_entry | windows |
| date_time | {'__class_name__': 'PosixTime', '__type__': 'DateTimeValues', 'timestamp': 0} | {'__class_name__': 'FATDateTime', '__type__': 'DateTimeValues'} | {'__class_name__': 'FATDateTime', '__type__': 'DateTimeValues'} | {'__class_name__': 'FATDateTime', '__type__': 'DateTimeValues'} | {'__class_name__': 'FATDateTime', '__type__': 'DateTimeValues'} | {'__class_name__': 'FATDateTime', '__type__': 'DateTimeValues'} | {'__class_name__': 'FATDateTime', '__type__': 'DateTimeValues'} | 'FATD |

**Statistics:**

No. Entries: 72

HIDDEN COLUMNS

| | 0 |
|---|---|
| 0 | __container_type__ |
| 1 | __type__ |
| 2 | data_type |
| 3 | inode |
| 4 | parser |
| 5 | pevtnum |
| 6 | message |
| 7 | sha256_hash |
| 8 | pathspec |

CONSTANT COLUMNS

| | Column | Value |
|---|---|---|
| 0 | D4_DataType_ | nan |
| 1 | D4_Orchestrator_ | nan |
| 2 | D4_Tool_ | plaso |
| 3 | D4_Plugin_ | windows_shell_item_file_entry |
| 4 | D4_Hostname_ | |
| 5 | date_time | {'__class_name__': 'FATDateTime', '__type__': 'DateTimeValues'} |
| 6 | hostname | DESKTOP-SDN1RPT |

| | Timestamp_ | timestamp_desc | display_name | file_reference | filename | long_name | name | origin | shell_item_path | timestamp | localized_name | pathspec_simple_ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 2019-12-07 09:03:46 | Creation | NTFS:\Users \Administrator \AppData \Local \Microsoft \Windows \UsrClass.dat | 1532-1 | \Users \Administrator \AppData \Local \Microsoft \Windows \UsrClass.dat | Windows | Windows | HKEY_CURRENT_USER\Software \Classes\Local Settings\Software \Microsoft\Windows\Shell \BagMRU\1\6 | <My Computer> C:\Windows | 1575709426000000 | <NA> | [p3]\Users \Administrator \AppData\Local \Microsoft\Windows \UsrClass.dat |

# ADAM

**ADAM allows you to define a sequence of malicious artifact data and inject it in a dataframe to test the detection capabilities**

- The DS ADversAry eMulator
- Mimick real attacks
- Inject events in multiple Artifact-specific Dataframes
- Creates a “Virtual” DataFrame environment


**More information in:**
**www.ds4n6.io/adam**

# DAISY

**Ready-to-use DS Virtual Machine designed to carry out Data Science and Machine/Deep Learning Analysis on DFIR data**

DAISY

| | DFIR | |
|---|---|---|
| Data | D | |
| | A | Artificial |
| | I | Intelligence |
| Science | S | |
| | Y | |

**More information in:**
**www.ds4n6.io/daisy**

# DAISY

## Forensics tools

RegRipper
http://github.com/keydet89

Volatility

The Sleuth Kit
DIGITAL FORENSIC TOOL

timesketch
Digital Forensics Timeline Analysis

## DS4N6

CHRYSALIS

## ML/DS tools

eland

jupyterlab TNG

pandas

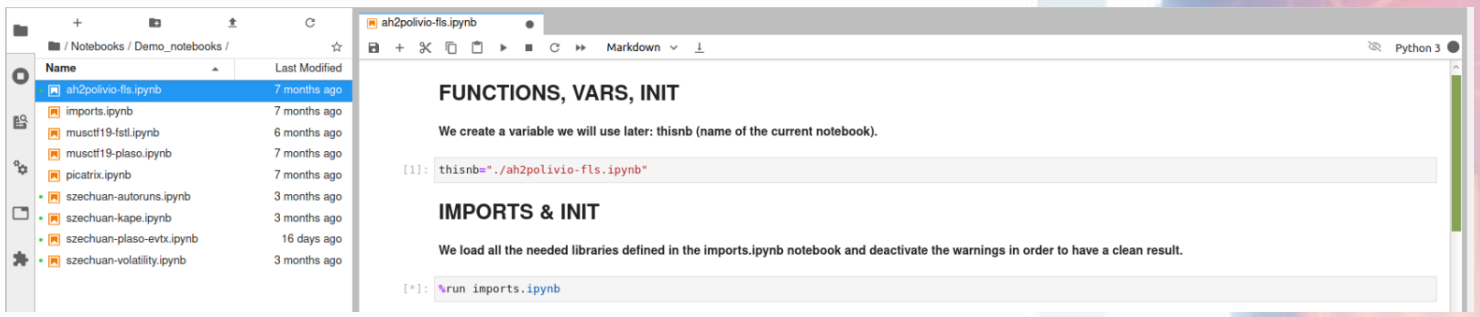
## Forensic demo data

I WANT THAT MULAN SZECHUAN NUGGET SAUCE

MAGNET FORENSICS

Ali Hadi

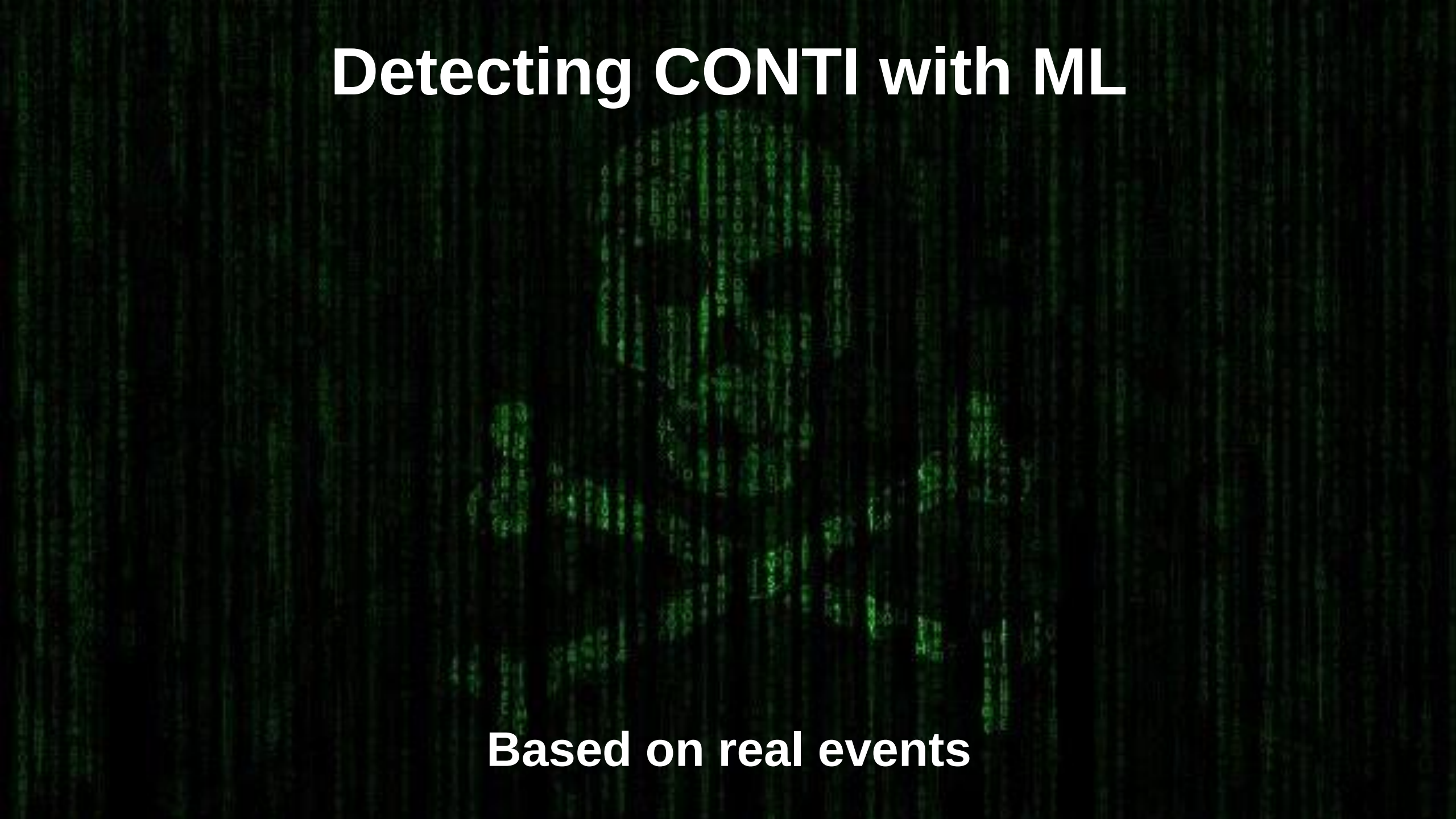
## Ready to use notebooks



# DAISY

ds4n6.io/daisy

# Detecting CONTI with ML



**Based on real events**

# The Big Challenge

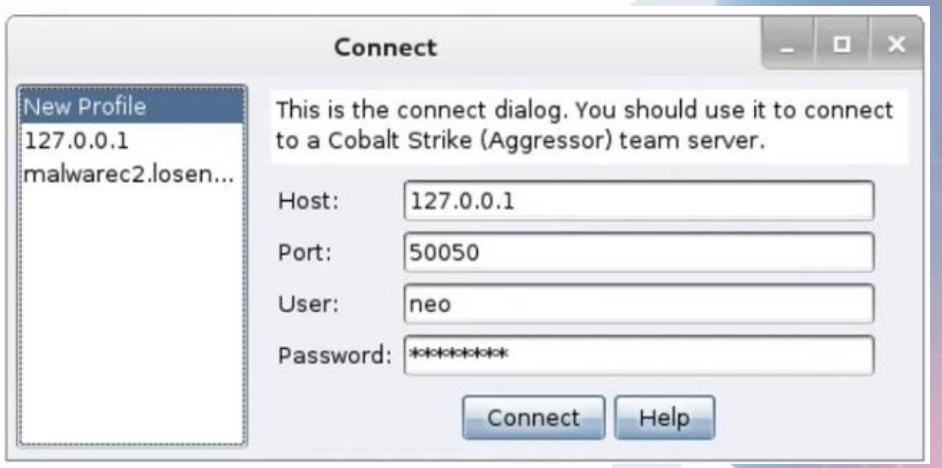
Would we be able to detect

**Cobalt Strike**

by just using

**Machine Learning?**

**Let's try!**



# Use Case: Cobalt Strike Detection

## Platform for Red Teams operations and adversary simulations

- 3rd most common threat (Red Canary)
- Beacons: Post exploitation payloads
- Malleable C2: language to give control over the indicators in the Beacon payload

THREAT

**Cobalt Strike**

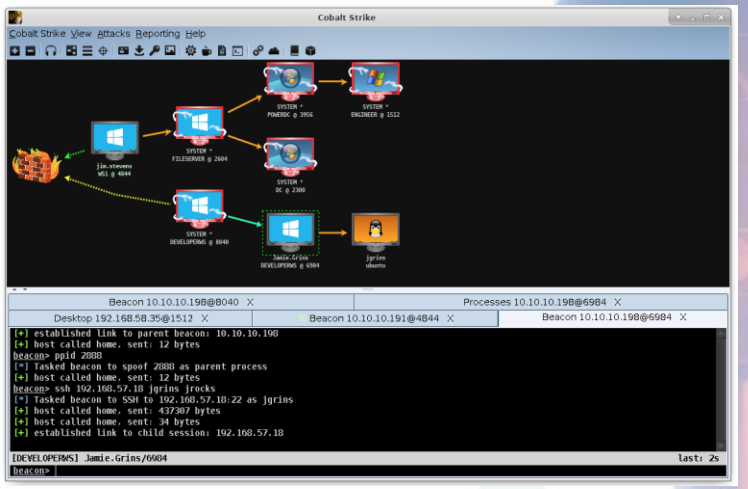
Cobalt Strike continues to be a favorite C2 tool among adversaries, as many rely on its functionality to maintain a foothold into victim organizations.

**#3**

OVERALL RANK

**7.9%**

CUSTOMERS AFFECTED



# Demo Data

30 days of **real world production server data**

**+100** servers

**+200K** events

**Cobalt Strike** real events injected with ADAM

**ML analysis** performed with CHRYSALIS

# The Ransomware Attack

## Global Company

**The attack could spread**

## CONTI

**TOP Threat Actor from Russia using Cobalt Strike**

## Worldwide Scope

**5k Servers + 350 DCs + 12k Laptops**

# The Breach. Day 0

**SOC Alert!**

**Pre-Ransomware tools found**

**5 infected hosts**

**5 days since intrusion**

**Possibly spread**

# Detecting The Enemy

## We will detect the intrusion in different phases

MITRE ATT&CK™

NEW

**Detecting Cobalt Strike with prefetch**

### TA0001: Initial Access
### T1078.003: Malicious Logons

ID: T1078.003

Sub-technique of: T1078

ⓘ Tactics: Defense Evasion, Persistence, Privilege Escalation, Initial Access

ⓘ Platforms: Containers, Linux, Windows, macOS

ⓘ Permissions Required: Administrator, User

Version: 1.2

Created: 13 March 2020

Last Modified: 18 October 2021

### TA0003: Persistence
### T1053.005: Scheduled Tasks

ID: T1053.005

Sub-technique of: T1053

ⓘ Tactics: Execution, Persistence, Privilege Escalation

ⓘ Platforms: Windows

ⓘ Permissions Required: Administrator

ⓘ Supports Remote: Yes

Contributors: Andrew Northern, @ex_raritas; Bryan Campbell, @bry_campbell; Selena Larson, @selenalarson; Zachary Abzug, @ZackDoesML

Version: 1.1

Created: 27 November 2019

Last Modified: 14 April 2022

### TA0005: Defense Evasion
### T1218: System Binary Proxy Execution

ID: T1218

Sub-techniques: T1218.001, T1218.002, T1218.003, T1218.004, T1218.005, T1218.007, T1218.008, T1218.009, T1218.010, T1218.011, T1218.012, T1218.013, T1218.014

ⓘ Tactic: Defense Evasion

ⓘ Platforms: Linux, Windows, macOS

ⓘ Defense Bypassed: Anti-virus, Application control, Digital Certificate Validation

Contributors: Hans Christoffer Gaardløs; Nishan Maharjan, @loki248; Praetorian; Wes Hurd

Version: 3.0

Created: 18 April 2018

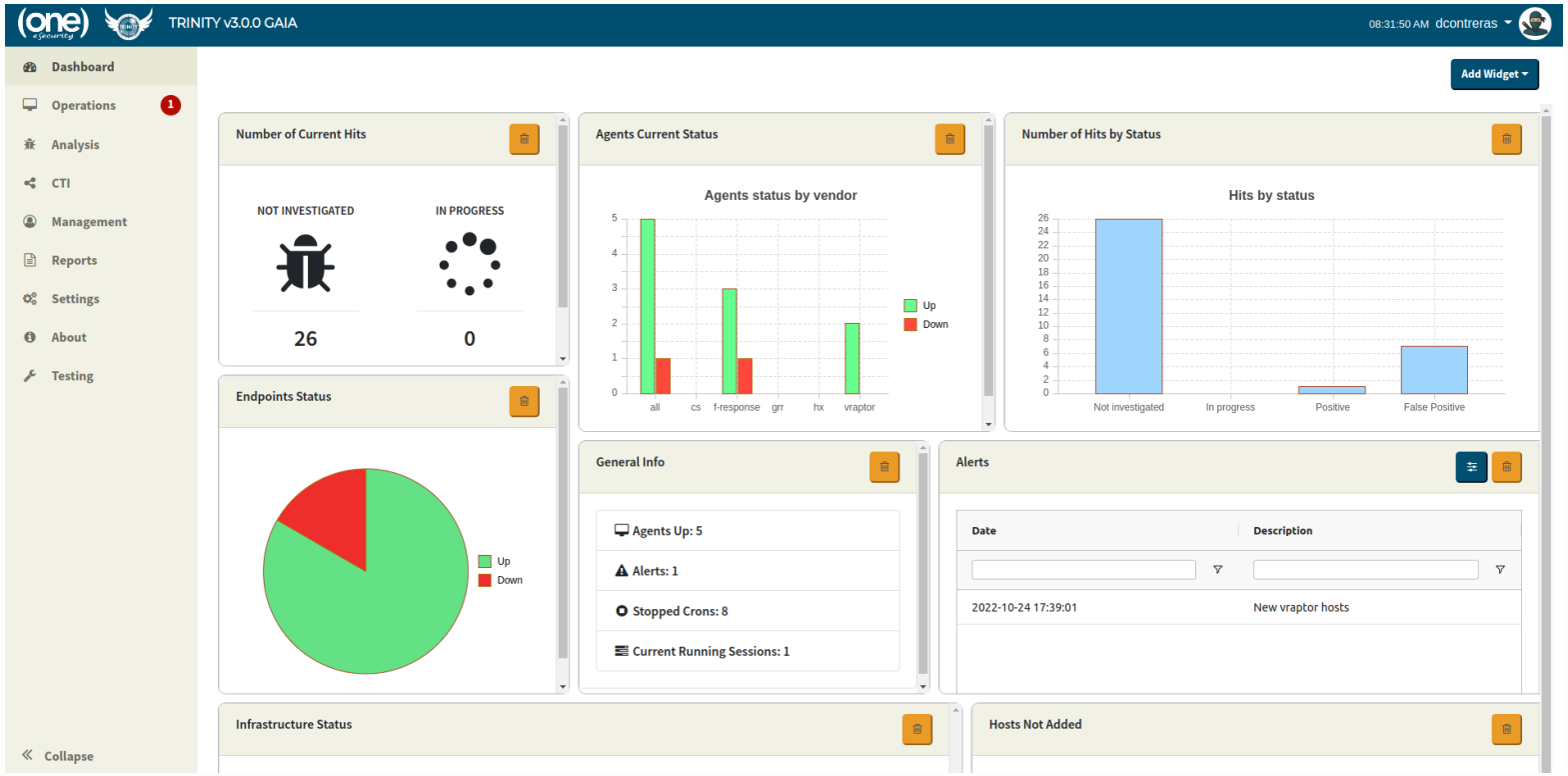
Last Modified: 18 April 2022

***New Lethal Forensicator Technique!***


DEMO
TIME
(one)
eSecurity

# Trinity

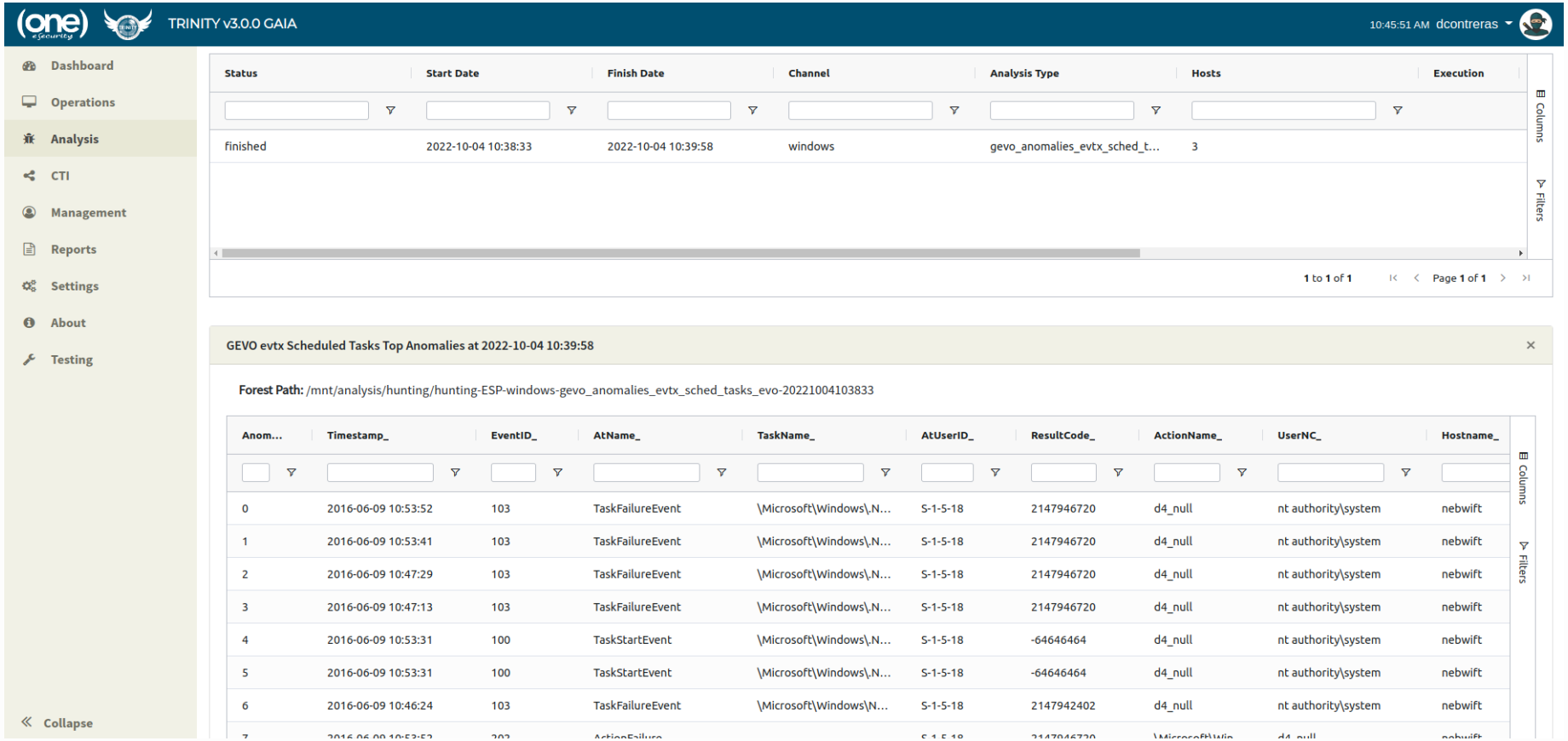
TRINITY is One eSecurity's Open XDR, developed by our DFIR experts and used to offer Threat Hunting services to our customers.




**Powered with AI by CHRYSALIS**

# Trinity

We are introducing CHRYSALIS AI power into TRINITY for anomalies detection



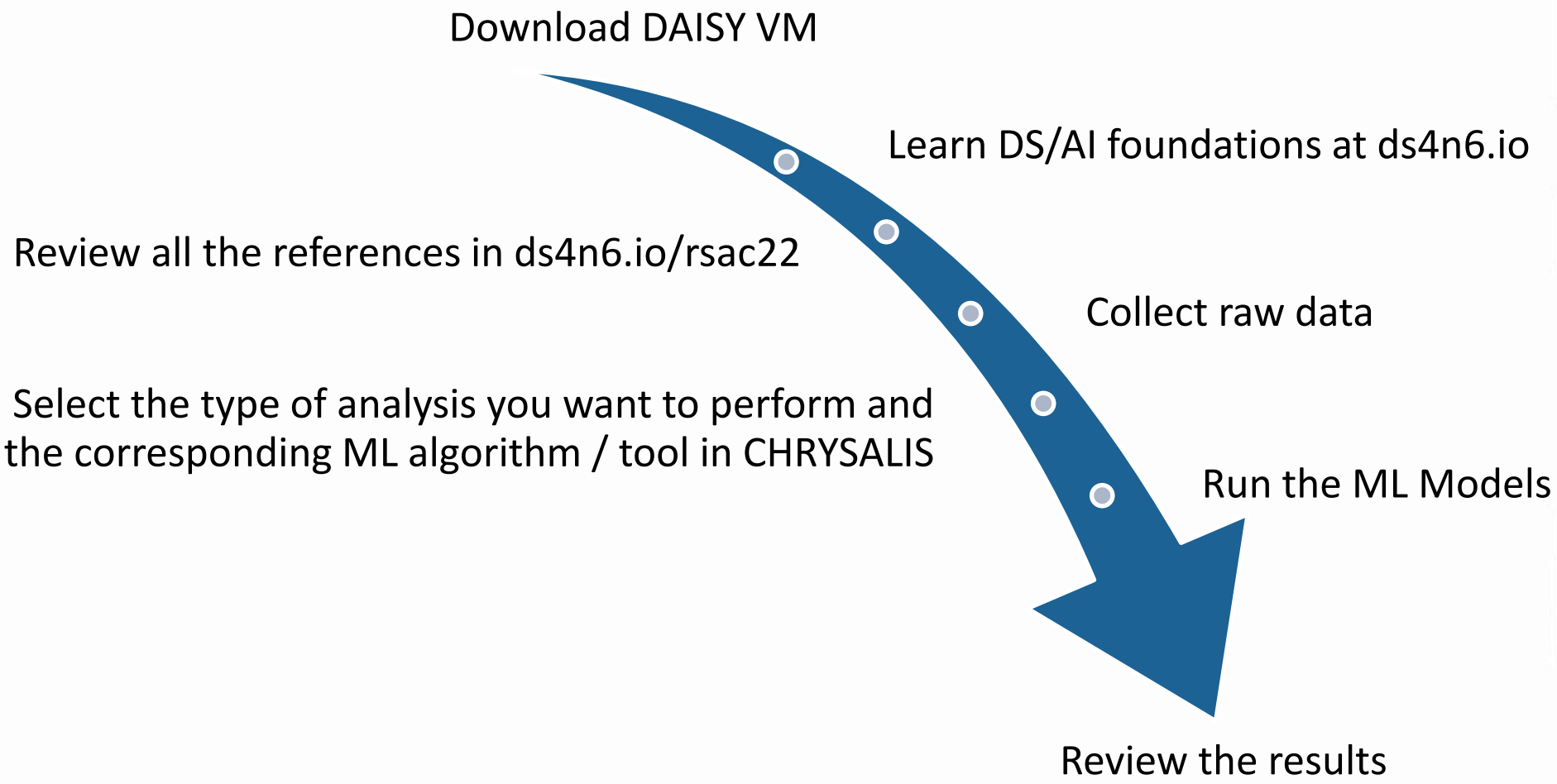

**Graph Analysis and many other AI applications are coming to TRINITY**

TRINITY shows the most anomalous scheduled tasks related events for a host

# Summary

- Machine Learning could enhance the analysis, detection and responses typically performed by forensicators
- There are not many open source tools using ML in DF
- DS4N6 is an open source project to bring the power of DS and ML to the community: CHRYSALIS, DAISY, etc.
- CHRYSALIS and the analysis presented have been used in real world incidents and with FORTUNE 500 customers

# Apply


Download DAISY VM
Learn DS/AI foundations at ds4n6.io
Review all the references in ds4n6.io/rsac22
Collect raw data
Select the type of analysis you want to perform and the corresponding ML algorithm / tool in CHRYSALIS
Run the ML Models
Review the results

All the details about this talk:

**ds4n6.io/odscwest22**

DS4N6

ds4n6.io

@ds4n6_io

DS4N6

**Jess Garcia**

@j3ssgarcia

**Thanks!**

one eSecurity

one-esecurity.com

One_eSecurity

One eSecurity