#RSAC

RSAConference2022

San Francisco & Digital | June 6 – 9

TRANSFORM

SESSION ID: **OST-T08**

# CHRYSALIS: Age of the AI-Enhanced Threat Hunters & Forensicators


**Jess Garcia**

**Founder of One eSecurity | Senior SANS Instructor**

@j3ssgarcia

# Disclaimer

Presentations are intended for educational purposes only and do not replace independent professional judgment. Statements of fact and opinions expressed are those of the presenters individually and, unless expressly stated to the contrary, are not the opinion or position of RSA Conference LLC or any other co-sponsors. RSA Conference does not endorse or approve, and assumes no responsibility for, the content, accuracy or completeness of the information presented.

Attendees should note that sessions may be audio- or video-recorded and may be published in various media, including print, audio and video formats without further notice. The presentation template and any media capture are subject to copyright protection.

# Our Objective

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

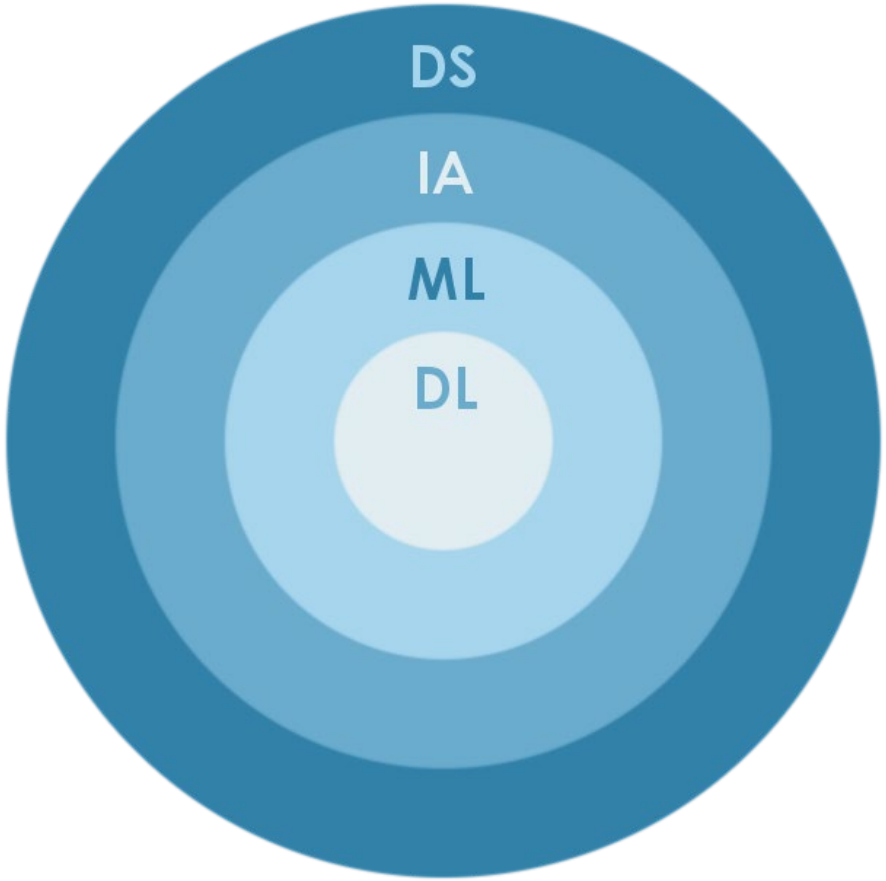
You do not need to be an AI expert, you will need **to learn what AI can do for you**, becoming familiar with the tools available and how to use them to suit their needs.

# 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 OVERVIEW

# The Age of Data Science & Machine Learning




Helps us understand the big picture

It helps working with complex scenarios

- Classification
- Prediction
- Anomaly Detection
- Noise Filtering
- Clustering

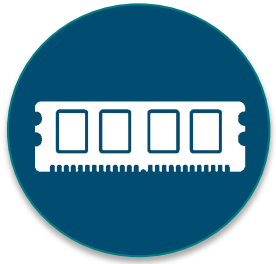
# ML FOR DFIR USE CASES

# Where Can We Use Machine Learning in DFIR?

DF

TH - CHRYSALIS

CTI



Memory Analysis – Columbo

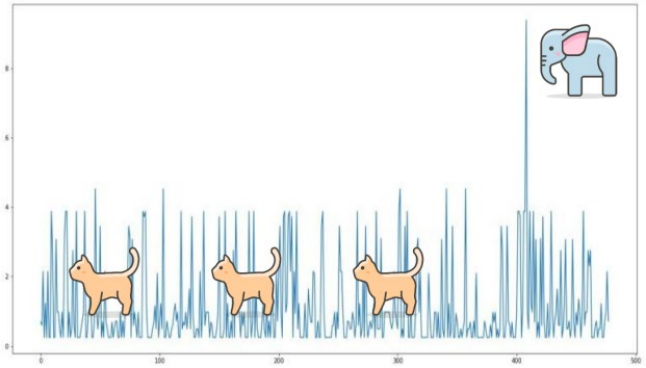
Malware – Malware Revealer

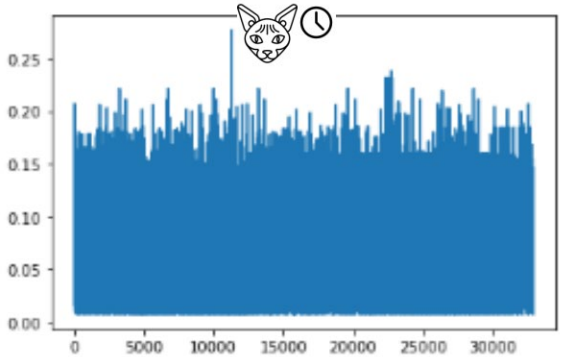
Logs - Deeplog

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 |

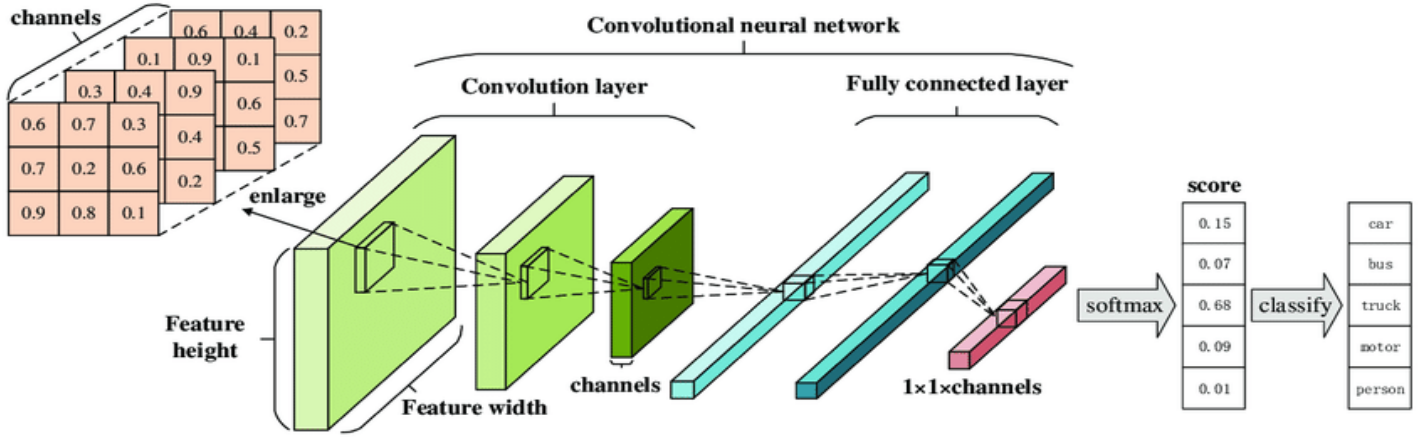
# ML & Malware: Detection and Classification

## Convolutional Neural Networks (CNN)

Raw bytes

Interesting patterns

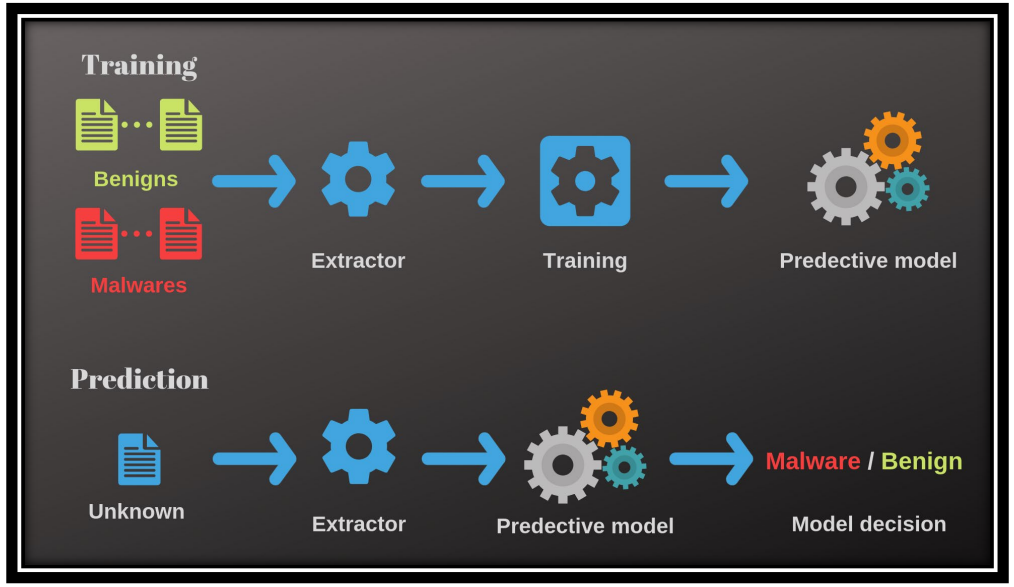
Classification




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.** ***Applied ML to Zeek****. 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

Machine Learning / Anomaly Detection Anomaly Explorer

Overview Anomaly Detection Data Frame Analytics Data Visualizer

Apr 30, 2020 @ 23:47:20.03 → Jun 5, 2020 @ 06:00:00.00 Refresh

Job Management Anomaly Explorer Single Metric Viewer Settings

dns_data_steal_detectionv2 Edit job selection

Filter by influencer fields... (destination.ip : 10.4.1.244)

Top Influencers

destination.ip

10.4.1.244 — 15 15

8.8.8.8 · 1 2

host.name

test-env-emodon... — 15 16

test-env-pcosic · 1 1

dns.question.etld_plus_one

dnsTunneling.bad — 15 15

secondexfil.test · 1 1

tunnelingTest.Test · < 1 < 1

dataexfil.io · < 1 < 1

Anomaly timeline

Overall

2020-05-03 2020-05-10 2020-05-17 2020-05-24 2020-05-31

View by: dns.question.etld_plus_one

Limit: 10

dnsTunneling.bad

secondexfil.test

tunnelingTest.Test

dataexfil.io

2020-05-03 2020-05-10 2020-05-17 2020-05-24 2020-05-31

Anomalies

Severity threshold: warning

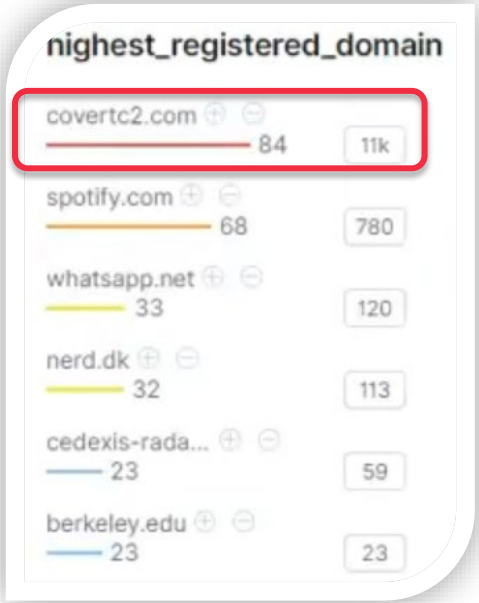
Interval: Auto

| time | severity | detector | found for | influenced by | actual | typical | description |
|---|---|---|---|---|---|---|---|
| May 29th 2020 | 18 | Detect tunneling and data exfiltration | dnsTunneling.bad | destination.ip: 10.4.1.244<br>dns.question.etld_plus_one: dnsTunneling.bad<br>host.name: test-env-emodonga | 66873 | 29.10034686446426 | More than 100x higher |

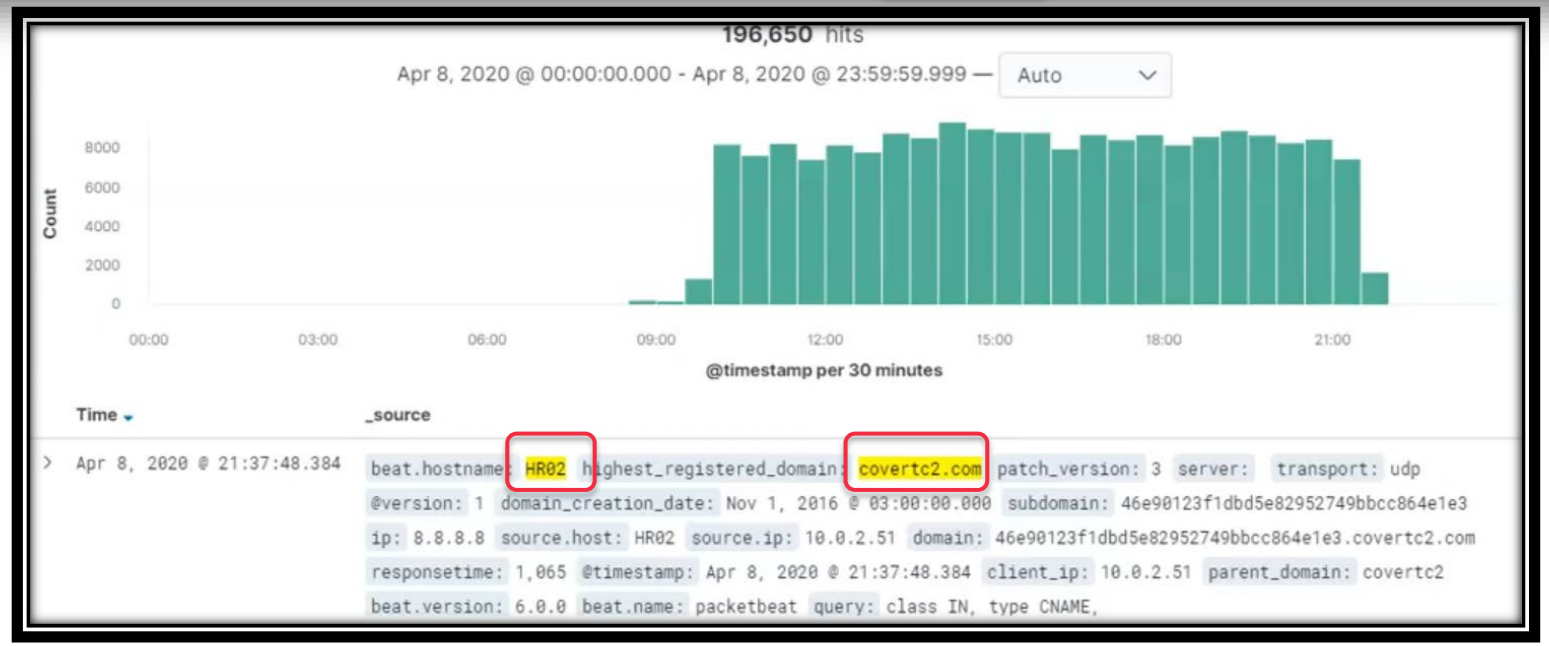
Host Details

Configure rules

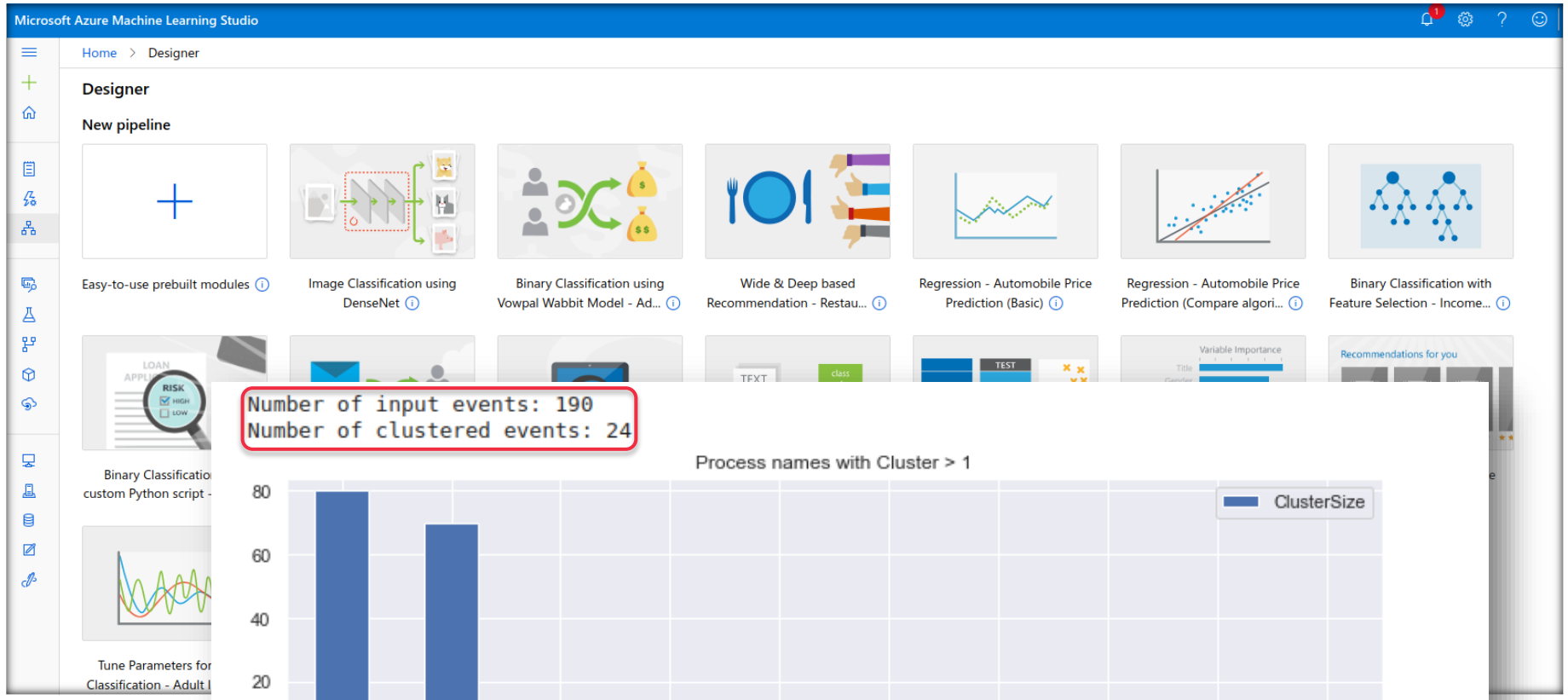
# ML & DF: Elastic – Use Case: DNS Exfiltration

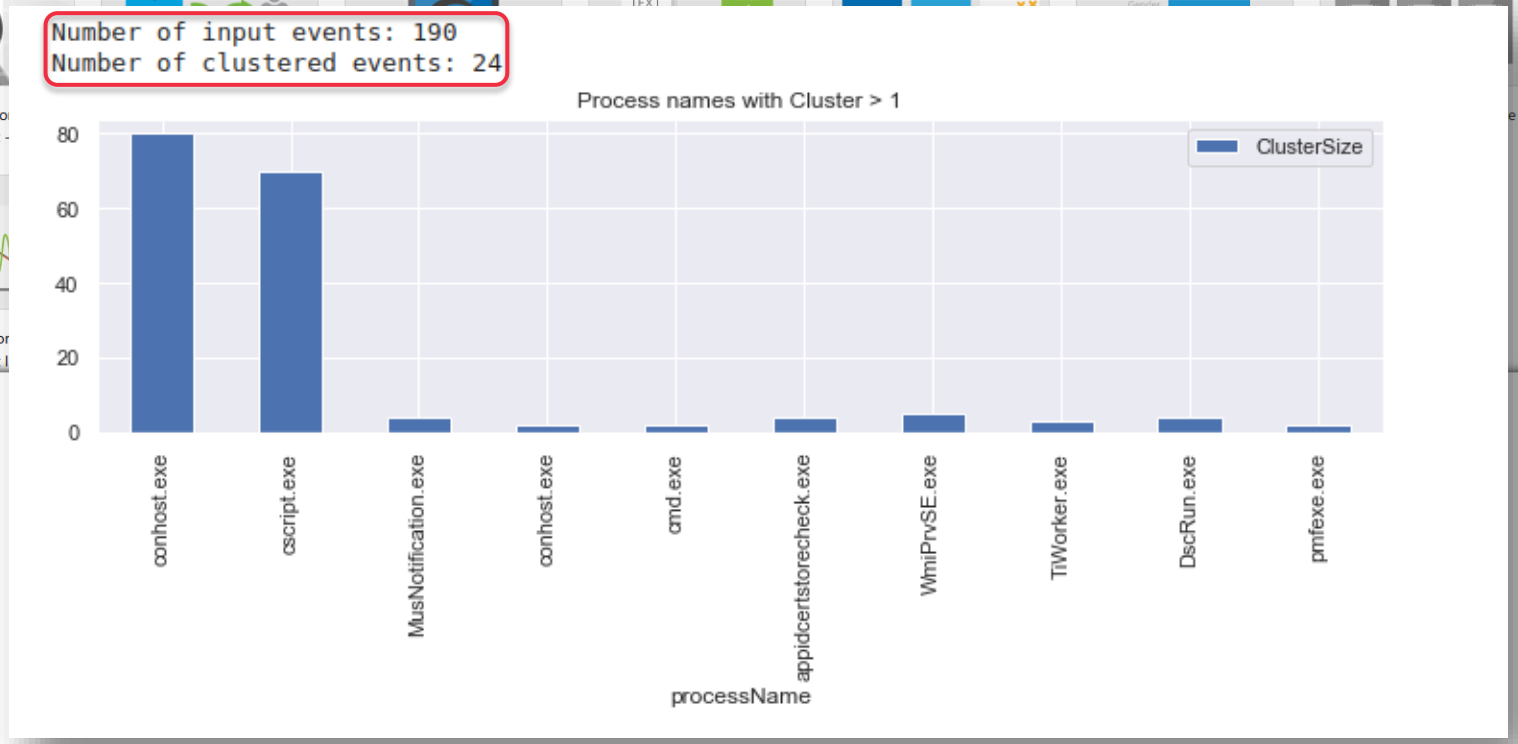



| | time | severity ↓ | detector | found for | influenced by | actual | typical | description |
|---|---|---|---|---|---|---|---|---|
| 334<br>118<br>93<br>222<br>71 | April 8th 2020 | 88 | high_info_content(subdomain) over highest_registered_domain excludefrequent=all | covertc2.com | beat.hostname: HR02<br>beat.hostname: NETWORK_TAP<br>highest_registered_domain: covertc2.com | 158140 | 17.24000160332901 | ↑ More than 100x higher |



# 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**

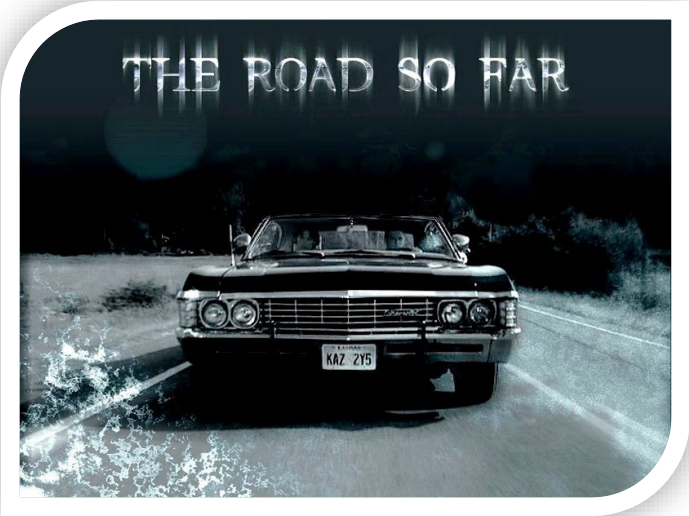
**Presented in**




# 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:**
**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

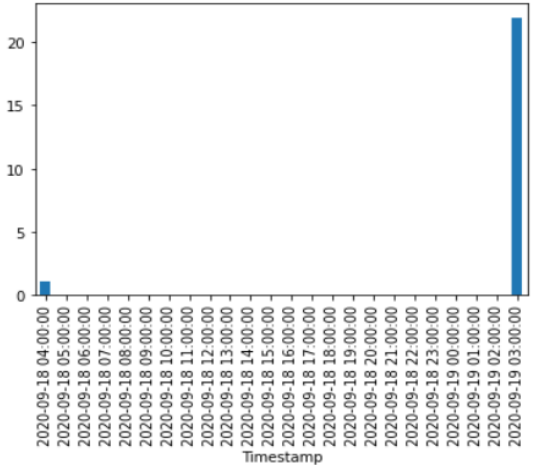
- Select 4625 as DF to analyze
- Select Failed Logons info as the available analysis

```
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 |

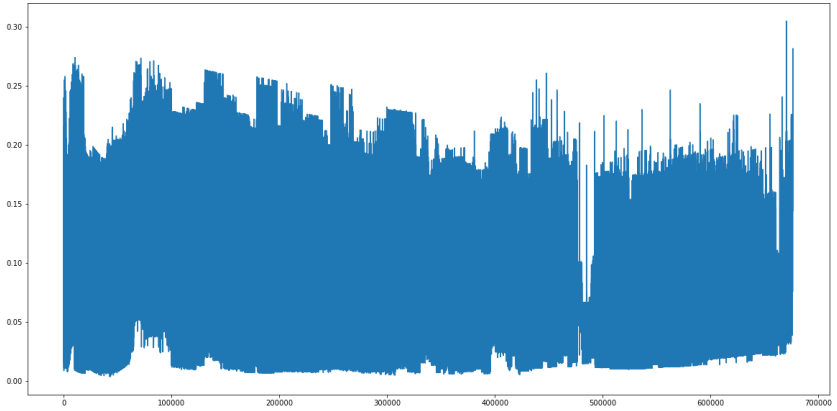
0s completed at 1:48 PM

**Try Colab now:**
**bit.ly/3Ff2VOm**

**Try Binder now:**
**bit.ly/3Ff2VOm**

# 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

| | 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**

**Tools**

- Kansa
- Kape
- Plaso
- Mactime
- Autoruns
- Macrobber
- Volatility

**Artifacts**

- SvsList
- Amcache
- Pslist
- Evtx
- Flist
- Winreg
- Fstl

**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

# DAISY

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

DFIR

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

More information in:
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™

pandas

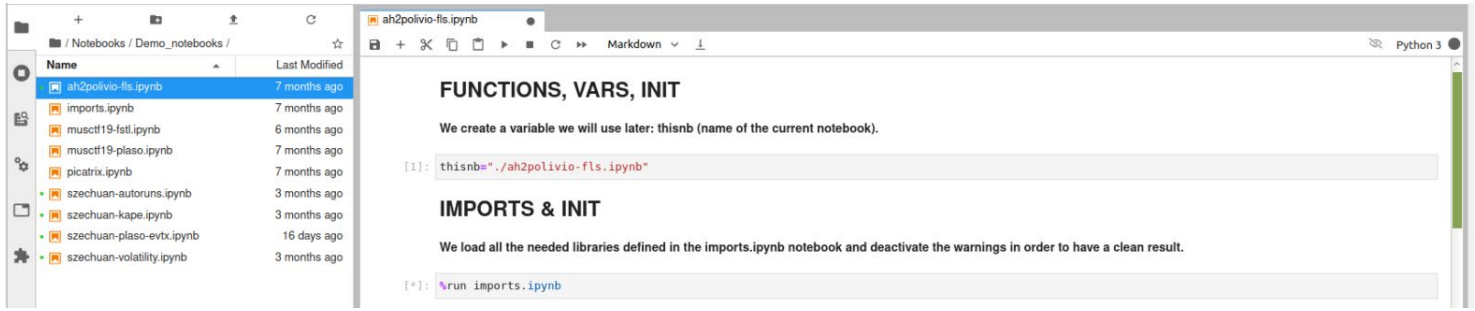
## Forensic demo data

I WANT THAT MULAN SZECHUAN NUGGET SAUCE

MAGNET FORENSICS

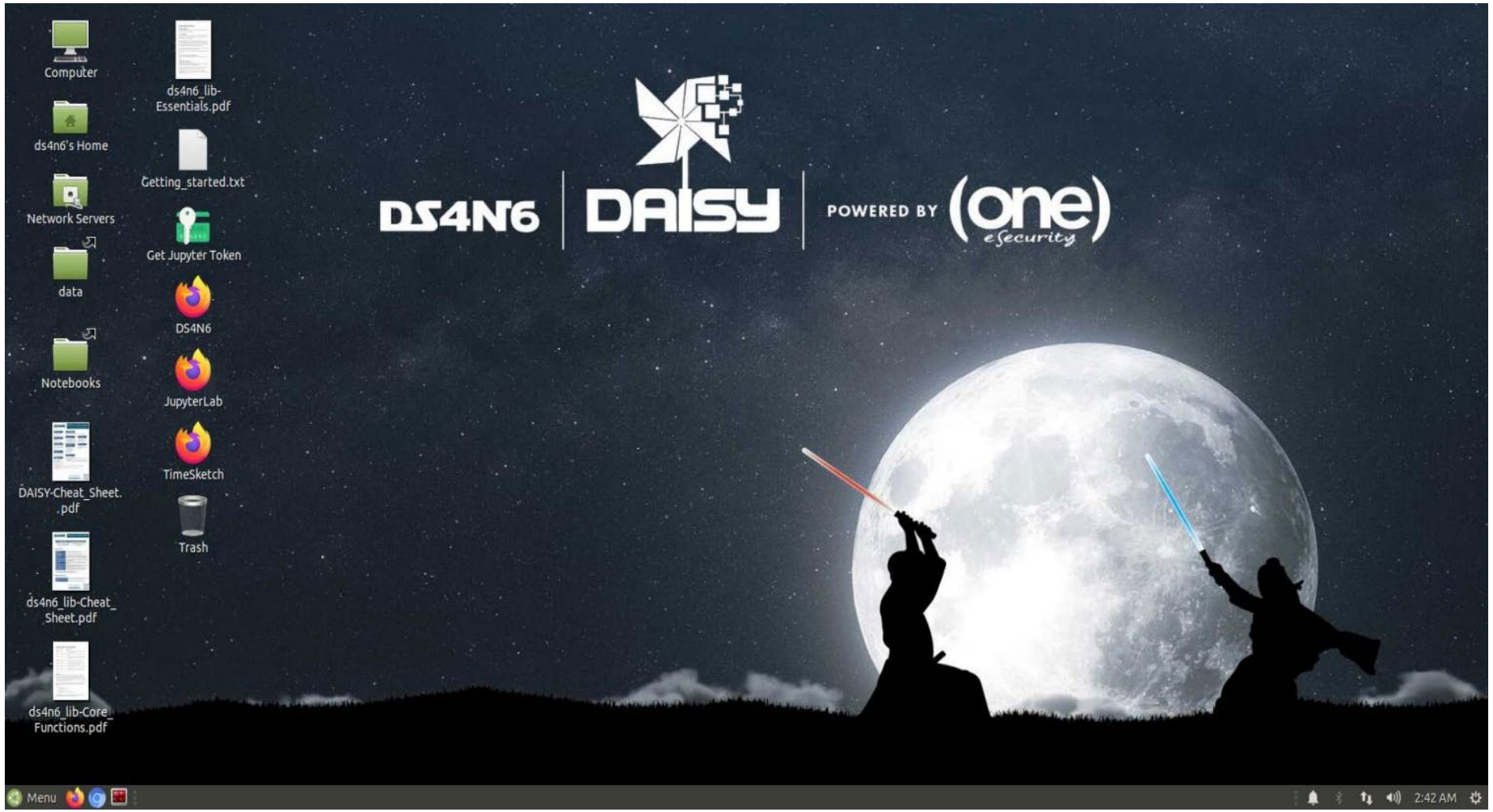
Ali Hadi

## Ready to use notebooks



# DAISY

ds4n6.io/daisy



# 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

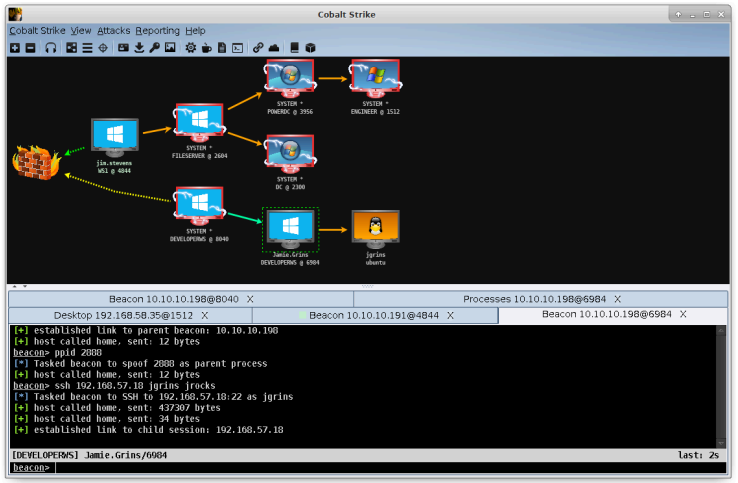
**Cobalt Strike**

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

red canary



# 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

# DEMO TIME

# 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/rsac22**

RSAConference2022

**DS4N6**

ds4n6.io

@ds4n6_io

DS4N6

**Jess Garcia**

@j3ssgarcia

**Thanks!**

one eSecurity

one-esecurity.com

One_eSecurity

One eSecurity