Hunting mediante la detección de anomalías con Machine Learning y DAISY

Threat Hunting Summit & Training 2021

$whoami

Jess Garcia
@j3ssgarcia

- Fundador y CEO de One eSecurity, una compañía global de Digital Forensics and Incident Response (DFIR) (~15 años)
- Líder del proyecto DS4N6.
  Visita: **www.ds4n6.io**
- Instructor Senior en el Instituto SANS (~20 años)

Problemas actuales en TH

Hunters un paso por detrás de atacantes

Hunts menos básicos son complicados con mucho volumen de datos

Hunters con menos experiencia pueden tardar mucho en encontrar eventos maliciosos

Muy difícil cuando hay que correlar fuentes de datos

Dependencia de herramientas caras (SIEMs, EDRs, etc.)

Métodos de Threat Hunting

IOCs

- El método más utilizado
- Pueden ser hashes, IPs, etc.
- La comunidad debe compartir los IOCs

TTPs

- Basado en la experiencia del analista y conocimiento de la comunidad
- Difícil a gran escala
- Lento

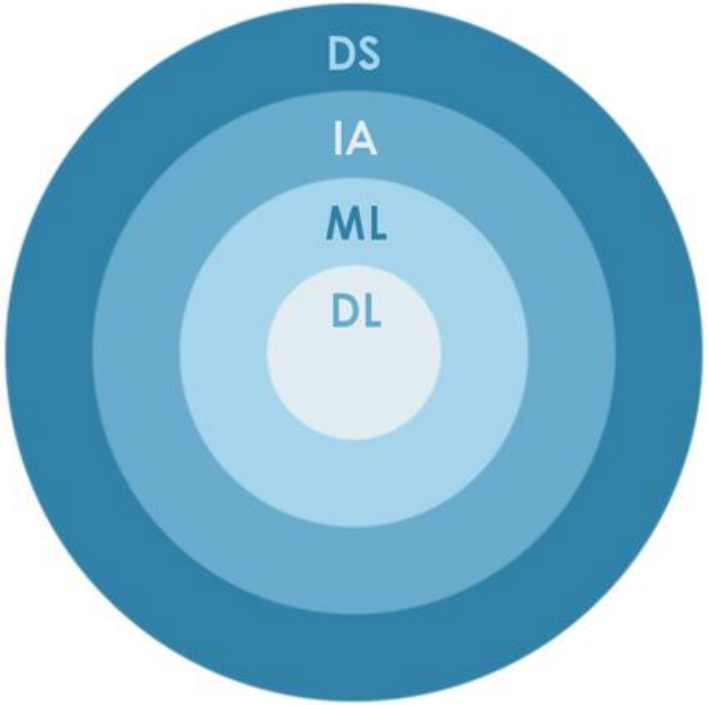
Anomalías

- Comúnmente realizado con baselines y análisis estadísticos
- Sin scoring -> Demasiados Falsos Positivos
- mitre.org/sites/default/files/publications/pr-19-3892-ttp-based-hunting.pdf

Hunting de anomalías con ML

- Búsqueda proactiva de anomalías en sistemas
- Proporciona métricas de la anomalía de cada artefacto
- Detección de ataques de actores no conocidos -> No ciberinteligencia
- Junto a DS permite TH en grandes datasets de forma más rápida
- Hacer TH en datos muy complejos de analizar

Pero, ¿qué es el ML?



- Útil para entender la *big picture*
- Escenarios complejos
- Se puede usar con **TH**, DFIR, CTI

Clasificación

Predicción

Detección de anomalías

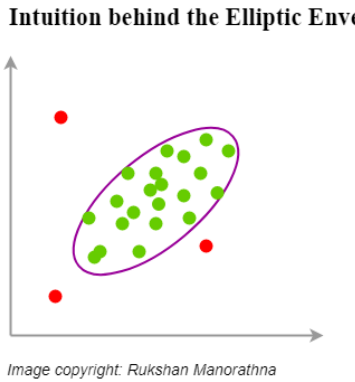
Clustering

Filtrado de ruido

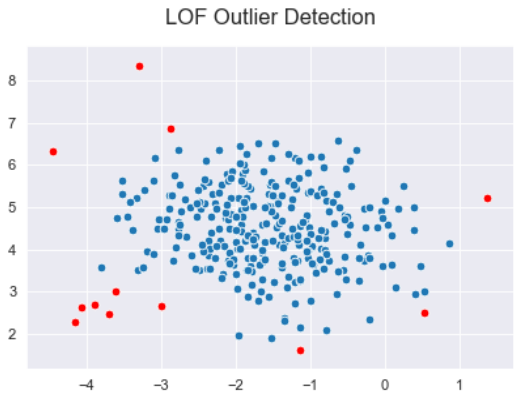
Detección de anomalías con DS

Métodos estadísticos

Elliptic Envelope Algorithm

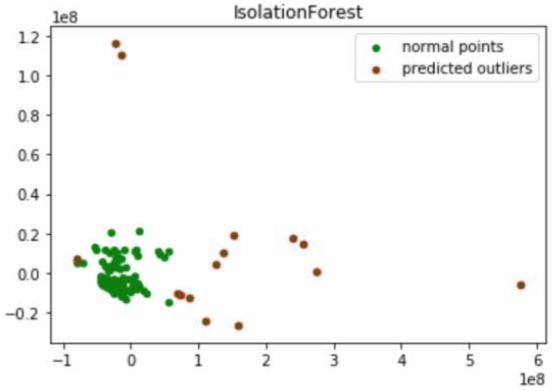


Local Outlier Factor

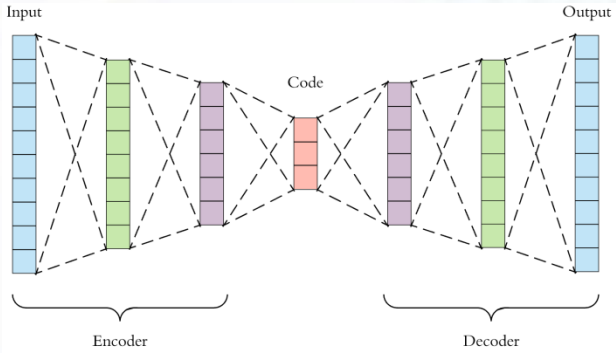


Machine Learning

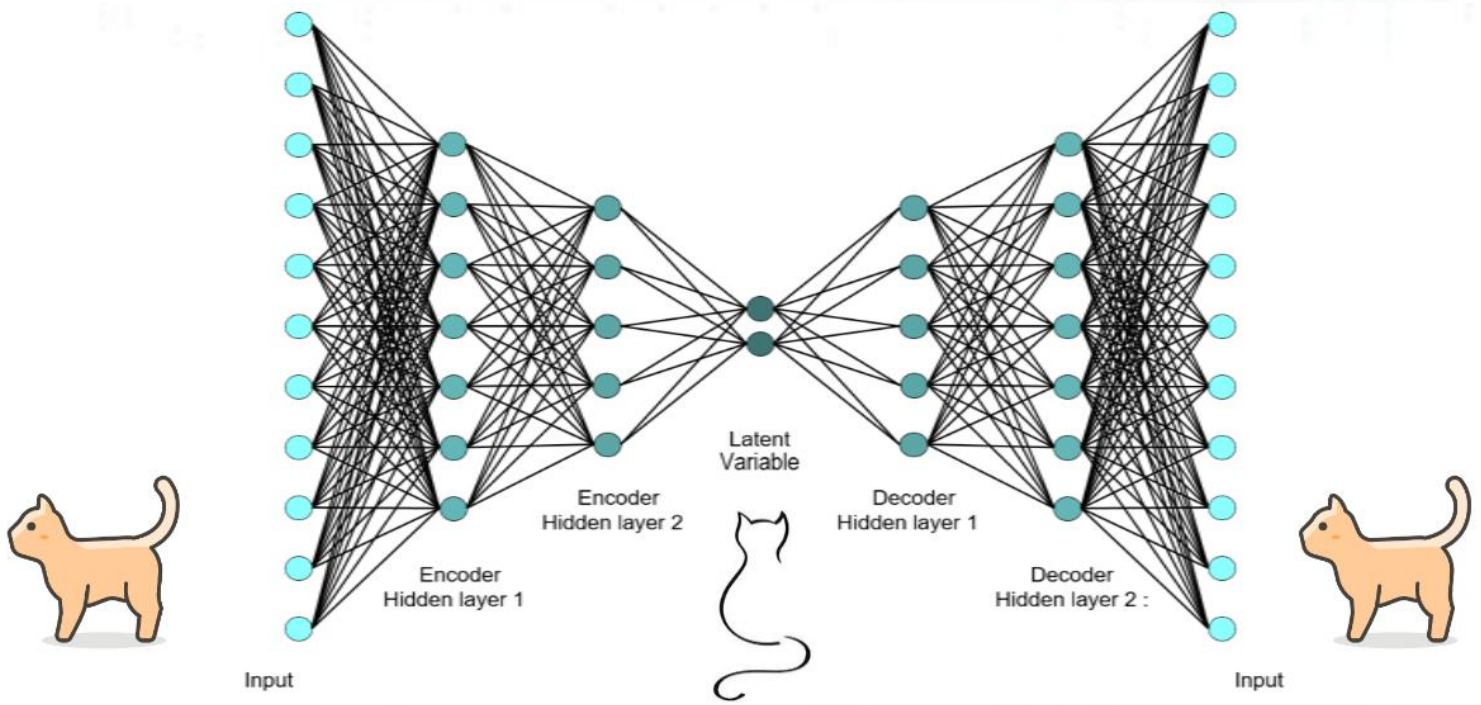
Isolation Forest Algorithm



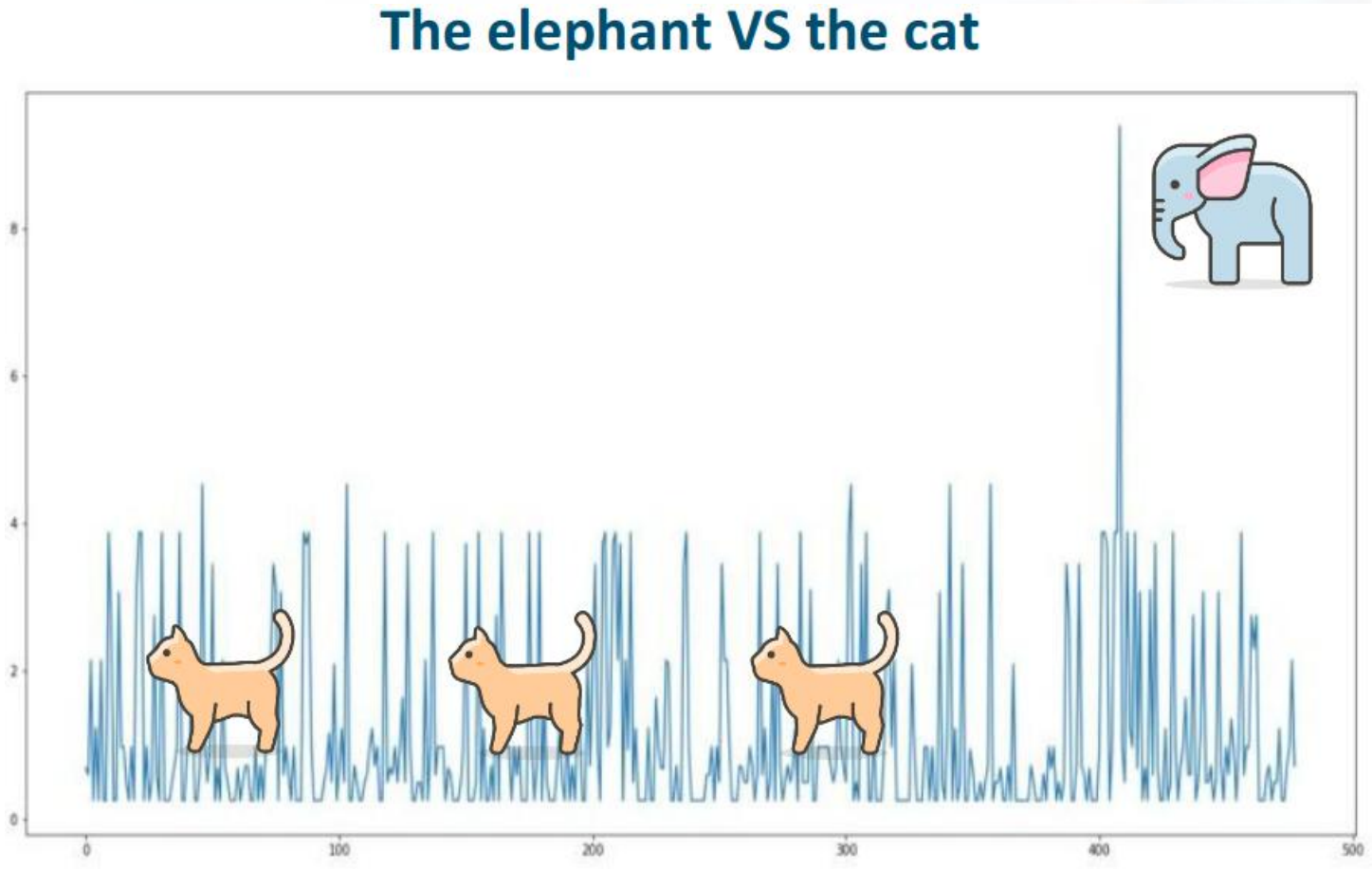
Autoencoder



Autoencoder



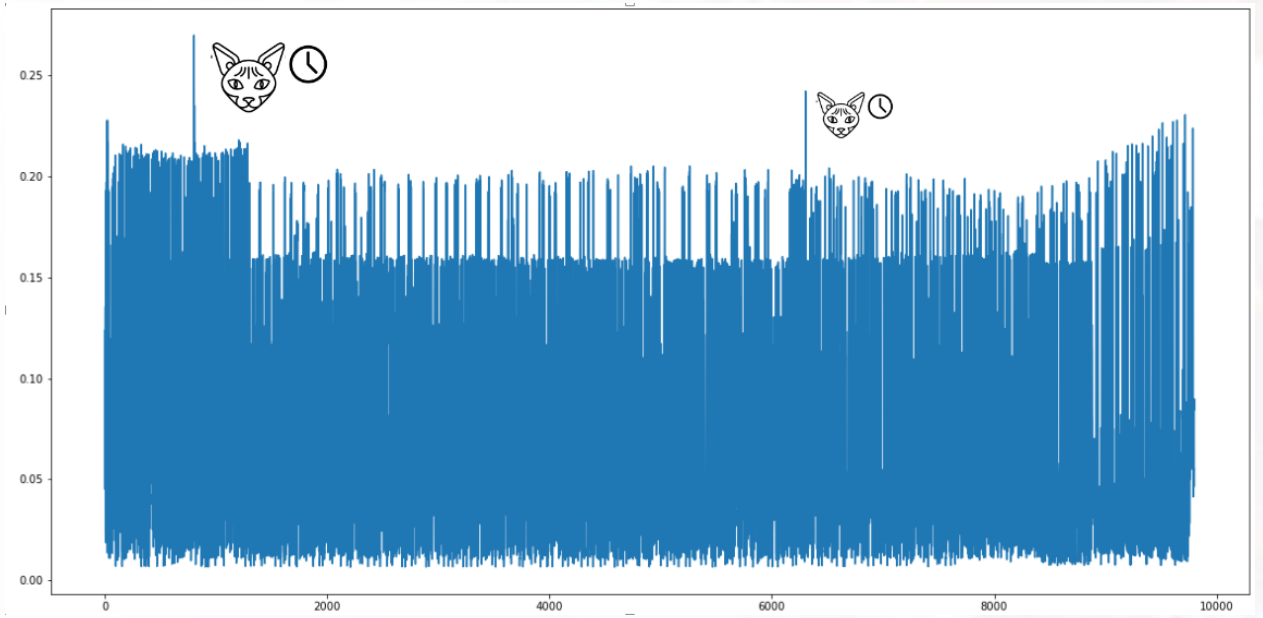
Autoencoder



Añadiendo la variable temporal

Un artefacto puede ser anómalo no solo
por la naturaleza del dato
También por el momento en el que ocurre

LSTM
Autoencoder



Autoencoder: métricas

Error de reconstrucción

lossdf

```
0     0.056931
1     0.028769
2     0.103462
3     0.123326
4     0.012045
5     0.014907
6     0.155894
7     0.023480
8     0.018195
9     0.042152
        ...
89    0.043169
90    0.075675
91    0.041373
92    0.013857
93    0.178709
94    0.042094
95    0.029611
96    0.014249
97    0.007125
98    0.007935
Length: 99, dtype: float64
```

Ordenando de mayor a menor error

Tabla de anomalías ordenadas

	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

Investigar % de anomalías TOP

Feature selection & engineering

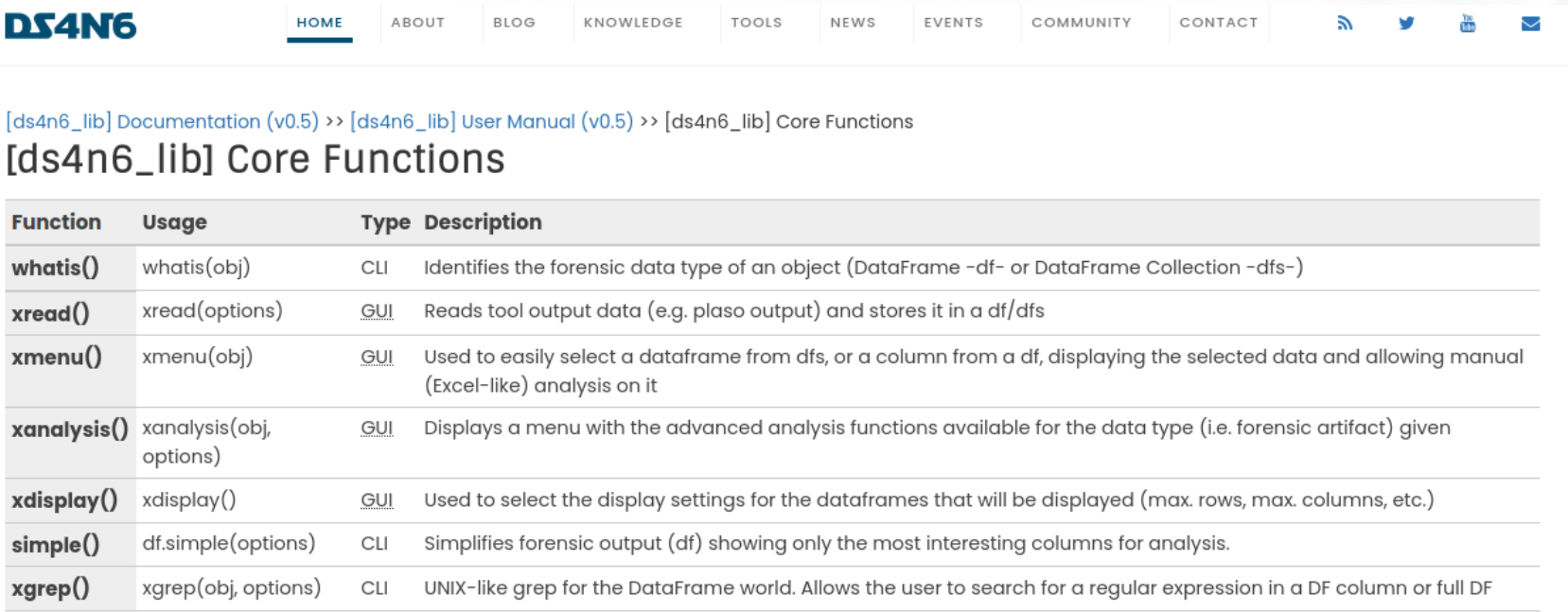
	Timestamp	EventID_	Computer	@Name	TaskName	UserName	UserContext	@UserID	ActionName	ResultCode	D4_DataType_	D4_Tool_	D4_Plugin_	evtxFileName_
0	2021-01-20 13:04:35	201	mc80-sc-2588	ActionSuccess	\Microsoft \Windows \ExploitGuard \ExploitGuard M...	NaN	NaN	S-1-5-18	NaN	0	evtx	plaso	windows_evtx_record	Microsoft-Windows-TaskScheduler%4Operational.evtx
1	2020-12-21 18:41:24	200	xwt70-sf-7556	ActionStart	\Microsoft \Windows \TPM\Tpm-Maintenance	NaN	NaN	S-1-5-18	TPM Maintenance Task Handler	NaN	evtx	plaso	windows_evtx_record	Microsoft-Windows-TaskScheduler%4Operational.evtx
2	2020-12-22 07:26:31	200	xwt70-sf-5375	ActionStart	\Microsoft \Windows \TPM\Tpm-Maintenance	NaN	NaN	S-1-5-18	TPM Maintenance Task Handler	NaN	evtx	plaso	windows_evtx_record	Microsoft-Windows-TaskScheduler%4Operational.evtx

Feature Selection
+
Feature Engineering

	index	EventID_	AtName_	TaskName_	AtUserID_	ResultCode_	ActionName_	UserNC_	Hostname_	D4_DataType_
0	0	201	ActionSuccess	\Microsoft\Windows\ExploitGuard\ExploitGuard M...	S-1-5-18	0	d4_null	d4_null	mc80-sc-2588	evtx-hml
1	1	200	ActionStart	\Microsoft\Windows\TPM\Tpm-Maintenance	S-1-5-18	-64646464	TPM Maintenance Task Handler	d4_null	xwt70-sf-7556	evtx-hml
2	2	200	ActionStart	\Microsoft\Windows\TPM\Tpm-Maintenance	S-1-5-18	-64646464	TPM Maintenance Task Handler	d4_null	xwt70-sf-5375	evtx-hml
3	3	200	ActionStart	\Microsoft\Windows\TPM\Tpm-Maintenance	S-1-5-18	-64646464	TPM Maintenance Task Handler	d4_null	xwt70-sf-1958	evtx-hml
4	4	201	ActionSuccess	\Microsoft\Windows\TPM\Tpm-Maintenance	S-1-5-18	0	TPM Maintenance Task Handler	d4_null	xwt70-sf-7556	evtx-hml

CHRYSALIS

- **Utiliza DS sin tener que aprender DS**
- Realiza tus investigaciones conociendo solamente 7 funciones



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

- Otros proyectos →

Toda la información en

www.ds4n6.io/tools/ds4n6_lib.html

Otros proyectos

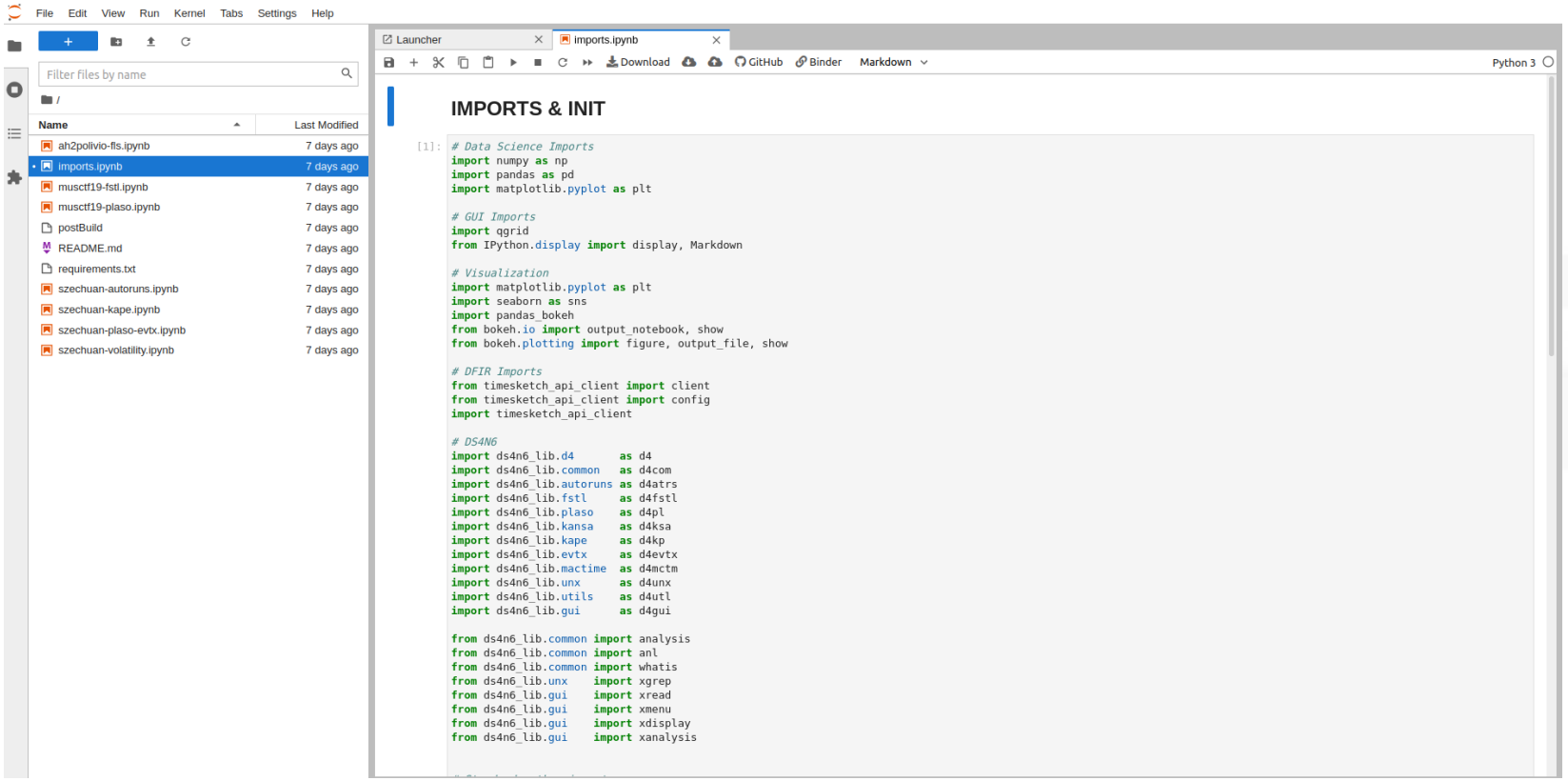
elastic

elastic.co

timesketch
Digital Forensics Timeline Analysis

github.com/google/timesketch

Meet JupyterLab



Introducción a Jupyter por Roberto Rodriguez:
https://posts.specterops.io/threat-hunting-with-jupyter-notebooks-part-1-your-first-notebook-9a99a781fde7

Objetivo: hacer TH con ML sin saber ML

find_anomalies()

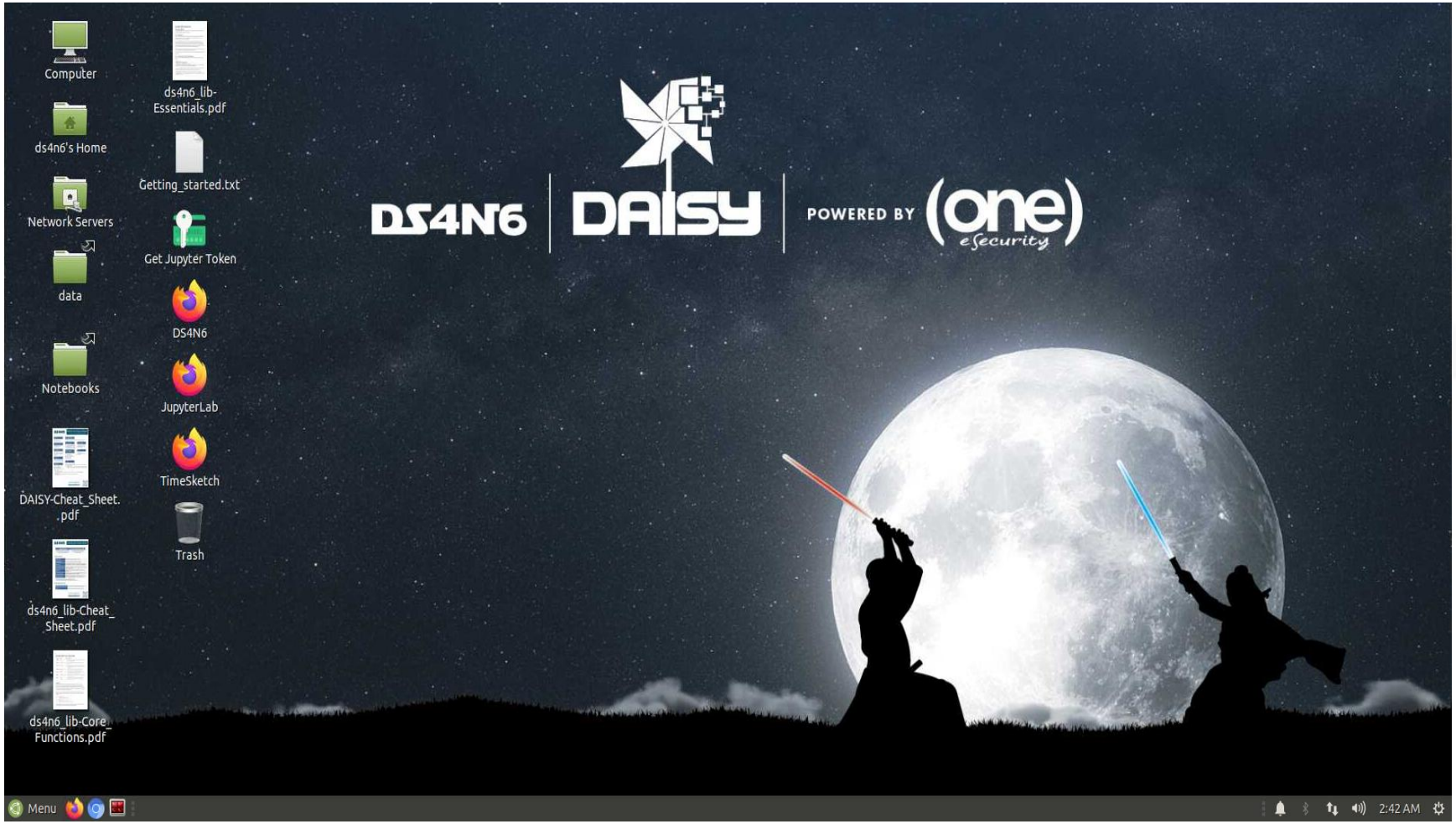
- Automática. **Aplica algoritmos de ML a tus datos sin saber sobre ML**
- Facilita una tabla con los eventos ordenados de mayor a menos anomalía
- Compatible con eventos de tareas programadas

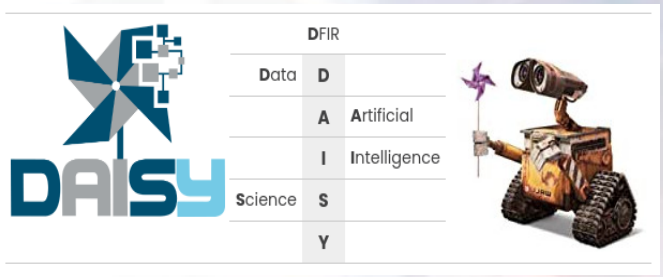
```
[40]: # Ejecución de la función find_anomalies().
      anomdf, lossdf = find_anomalies(schtsksevtxdf, model_type='lstm_autoencoder', epochss=[1])
```

1 epochs

```
- General:
  + Verbosity:              0
- Input Data:
  + Train DF:               (676476, 9)
  + Prediction DF:          (676476, 9)
- Data Preparation:
  + Transform Method:       label_encoder
  + Data Scaling Method:    normalize
- Model Parameters:
  + Model Type:             lstm_autoencoder
  + Encoding Dimension:     3
  + Activation Function:    relu
  + lstm_units:             50
  + lstm_time_steps:        200
- Training Parameters:
  + Training Loops:         1
  + epochs:                 1
  + batch_size:             32
```

DAISY





www.ds4n6.io/daisy

Presentación completa en

www.ds4n6.io/rsac21

Caso práctico

120+ servers

600.000+ events
from real world

¿Detectaremos el evento malicioso como anómalo?

	EventID_	AtName_	TaskName_	AtUserID_	ActionName_	ResultCode_	UserNC_	Hostname_
202569	106	TaskRegisteredEvent	\Microsoft\Windows\SoftwareProtectionPlatform \EventCacheManager	S-1-5-18	d4_null	None	d4_null\cssvrtmp	compromised_host
202570	140	TaskUpdated	\Microsoft\Windows\SoftwareProtectionPlatform \EventCacheManager	S-1-5-18	d4_null	None	d4_null\system	compromised_host
215047	200	ActionStart	\Microsoft\Windows\SoftwareProtectionPlatform \EventCacheManager	S-1-5-18	C:\Windows\SoftwareProtectionPlatform \EventCacheManager.exe	None	d4_null\none	compromised_host
215048	201	ActionSuccess	\Microsoft\Windows\SoftwareProtectionPlatform \EventCacheManager	S-1-5-18	C:\Windows\SoftwareProtectionPlatform \EventCacheManager.exe	0	d4_null\none	compromised_host


DEMO
TIME
(one)
eSecurity

Trabajos futuros

Próximamente en DS4N6

- Más artefactos compatibles con find_anomalies()
- Pivoting entre artefactos

Algunos usos:

- **Identificación de LOLBins**: detección de outliers usados de forma maliciosa
- **Identificación de inyecciones DLL:** inyecciones de assemblies
- **Identificación de drivers maliciosos**: sin uso de reglas estáticas ni *whitelist*

Próximamente

Nueva versión de DAISY: D4ML

Nuevo blogpost: ¿Cómo puedes probarlo tú?

Mientras tanto…

¡Puedes probar la versión actual de CHRYSALIS en la nube!

Visita bit.ly/3Ff2VOm o escanea el código QR para probar el repositorio de DS4N6

All the details about this talk:

ds4n6.io/sansth21

Threat Hunting Summit & Training 2021

DS4N6

www ds4n6.io

@ds4n6_io

DS4N6

Jess Garcia

@j3ssgarcia

Thanks!

(one) eSecurity

www one-esecurity.com

One_eSecurity

One eSecurity