CIBERSEGURIDAD

‘Bootcamp IX’

Informe Práctica Módulo Machine Learning y Ciberseguridad.

Maximiliano Dariel Altamirano.

Academia KeepCoding.

Informe

Descripción de la practica

Abordamos una problemática de clasificación con Machine Learning con datos del fichero dataset “cybersecurity\_attacks.csv”.

Trabajaremos, en primera instancia, con la columna “Action Taken” como objetivo a predecir. Luego en el análisis extendido sumaremos las columnas “Severity Level” y “Attack Type”.

Gestionaremos los datos en el entorno JupyterLab.

Desarrollo de la práctica.

En el fichero “práctica\_machine\_learning.ipynb” realizamos el primer análisis y desarrollo cubriendo las expectativas básicas de la práctica. Segmentamos el análisis en:

1. Preparación de datos
2. Análisis exploratorio
3. Preprocesamiento/Generación de variables
4. División train/test
5. Modelado/Evaluación
6. Conclusiones

Entrenando los modelos de ML:

* RandomForestClassifier
* LogisticRegression
* KNeighborsClassifier
* DecisionTreeClassifier

Este primer análisis se corresponde con los siguientes ficheros:

* práctica\_machine\_learning.ipynb
* cybersecurity\_attacks.csv
* action\_taken\_test.csv
* action\_taken\_train.csv
* consigna\_practica\_final.pdf

Práctica extendida

La entrega incluye un directorio “Extendido”, donde intentaremos evidenciar el análisis de la misma problemática con otro enfoque.

Realizamos una breve descripción de cada fichero:

* 01\_análisis exploratorio.ipynb: intentaremos ser más minusiosos en la selección de columnas, exploración de datos y transformaciones. Del análisis realizado guardaremos la información en “02\_clean\_cybersecurity\_attacks.csv” (datos ya transformados y seleccionados) y “03\_codif\_cybersecurity\_attacks.csv” (datos codificados).
* 04\_action\_taken.ipynb: entrenaremos varios modelos de ML con la columna “Action Taken” como objetivo a predecir, descartando las columnas con menor importancia.
* 05\_severity\_level.ipynb: entrenaremos varios modelos de ML con la columna “Severity Level” como objetivo a predecir, descartando las columnas con menor importancia.
* 06\_attack\_type.ipynb: entrenaremos varios modelos de ML con la columna “Action Taken” como objetivo a predecir, descartando las columnas con menor importancia.
* 07\_feature\_importances.ipynb: considerando los datos obtenidos anteriormente, en este fichero planteamos trabajar con una cantidad de columnas reducidas, ahora seleccionando solo las columnas más importantes. Guardamos este análisis en el fichero “08\_feature\_importances.csv”.

En este punto decidimos utilizar solo dos modelos a entrenar, LogisticRegression y KNeighborsClassifier, ya que obtuvimos resultados equilibrados entre los conjuntos de datos “train” y “test”.

* 09\_LogisticRegression.ipynb: entrenamos el modelo LogisticRegression y marcamos conclusiones con respecto a resultados anteriores
* 10\_KNeighborsClassifier.ipynb: entrenamos el modelo KNeighborsClassifier y marcamos conclusiones.