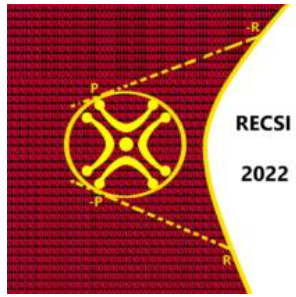


XVII Reunión Española sobre Criptología y Seguridad de la Información

Asignación multiclase de la severidad de IPs mediante aprendizaje no supervisado

Noemí DeCastro-García, *Universidad de León*

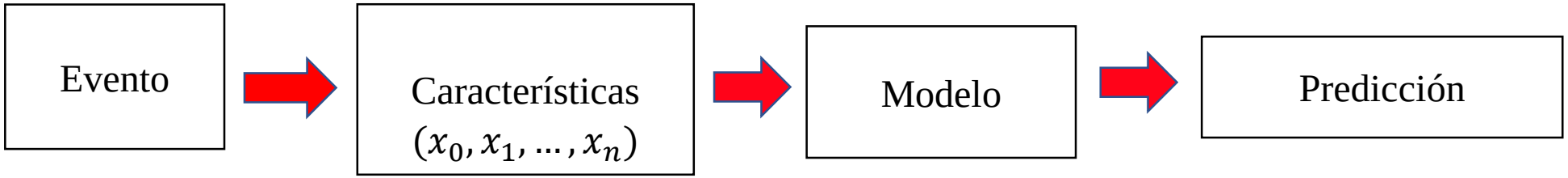
David Escudero García, *RIASC, Universidad de León*

Introducción

- Objetivo: caracterizar direcciones IP asociadas a eventos en diferentes grados de severidad
- Estimar correctamente el grado de amenaza para responder correctamente

Esquema

- Uso de machine learning no supervisado
- Selección de las características
  - Mayor tasa de acierto con mayor complejidad; más difícil de implementar



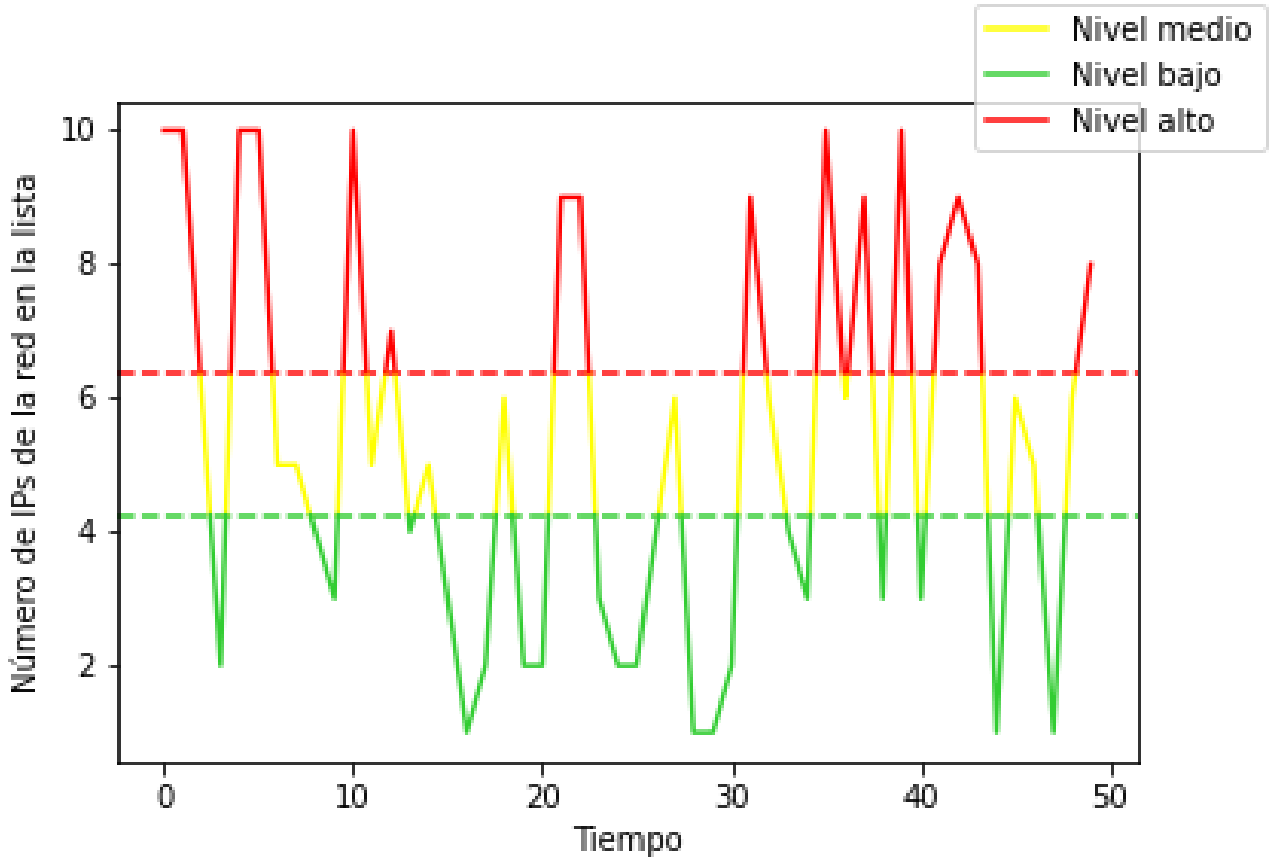
RECSI 2022

Características (I)

- Información propia de los eventos, sin apenas preprocesado
  - La propia IP, representada numéricamente.
  - Fecha de ocurrencia (timestamp)
  - Geolocalización: latitud, longitud, código ISO de país.

Características (II)

- Para cada subred, se forma una serie temporal que se divide en 3 franjas [1]:
  - Bajo: <= media * (1 - δ)
  - Alto: >= media * (1 + δ)



[1] Y. Liu, J. Zhang, A. Sarabi, M. Liu, M. Karir, and M. Bailey, "Predicting cyber security incidents using feature-based characterization of network-level malicious activities," 2015, ACM International Workshop on International Workshop on Security and Privacy Analytics

RECSI 2022

Datos

- 99720 eventos de mayo de 2021
- 4 niveles de severidad
- Modelos:
  - K-Means
  - DBSCAN

Severidad	Porcentaje
1	8.42
3	25.03
6	54.58
9	11.97

Métricas

- Al usar aprendizaje no supervisado es necesario trasladar los grupos a predicciones
- Tasa de acierto
- Coeficiente de correlación de Matthews

$$\frac{TP.TN - FP.FN}{\sqrt{(TP.FP)(TP.FN)(TN.FP)(TN.FN)}}$$

$$TP = verdadero\ positivo$$
$$TN = verdadero\ negativo$$
$$FP = falso\ positivo$$
$$FN = falso\ negativo$$

- Eficaz con datos desbalanceados

Resultados

Algoritmo	MCC	Tasa de acierto
K-Means	0.512	0.637
DBSCAN	0.488	0.594

- Características de geolocalización no muy informativas: la mayoría en España
- Insuficientes datos para el análisis de series temporales

Modificaciones (I)

- Más características[2]:
  - IPs pertenecientes a una clase presentes en la red de la IP
  - $Score(IP)_c = Número\ de\ direcciones\ con\ clase\ c\ en\ la\ red\ de\ IP$

IP	Clase	Score_1	Score_3	Score_9
123.45.23.12	1	2	1	0
123.45.23.45	1	2	1	0
187.145.34.56	9	0	0	1
123.45.23.7	3	2	1	0

[2]Moura, G.C.M., Sadre, R., Pras, A., 2014. Taking on internet bad neighborhoods, in 2014 IEEE Network Operations and Management Symposium

Modificaciones (II)

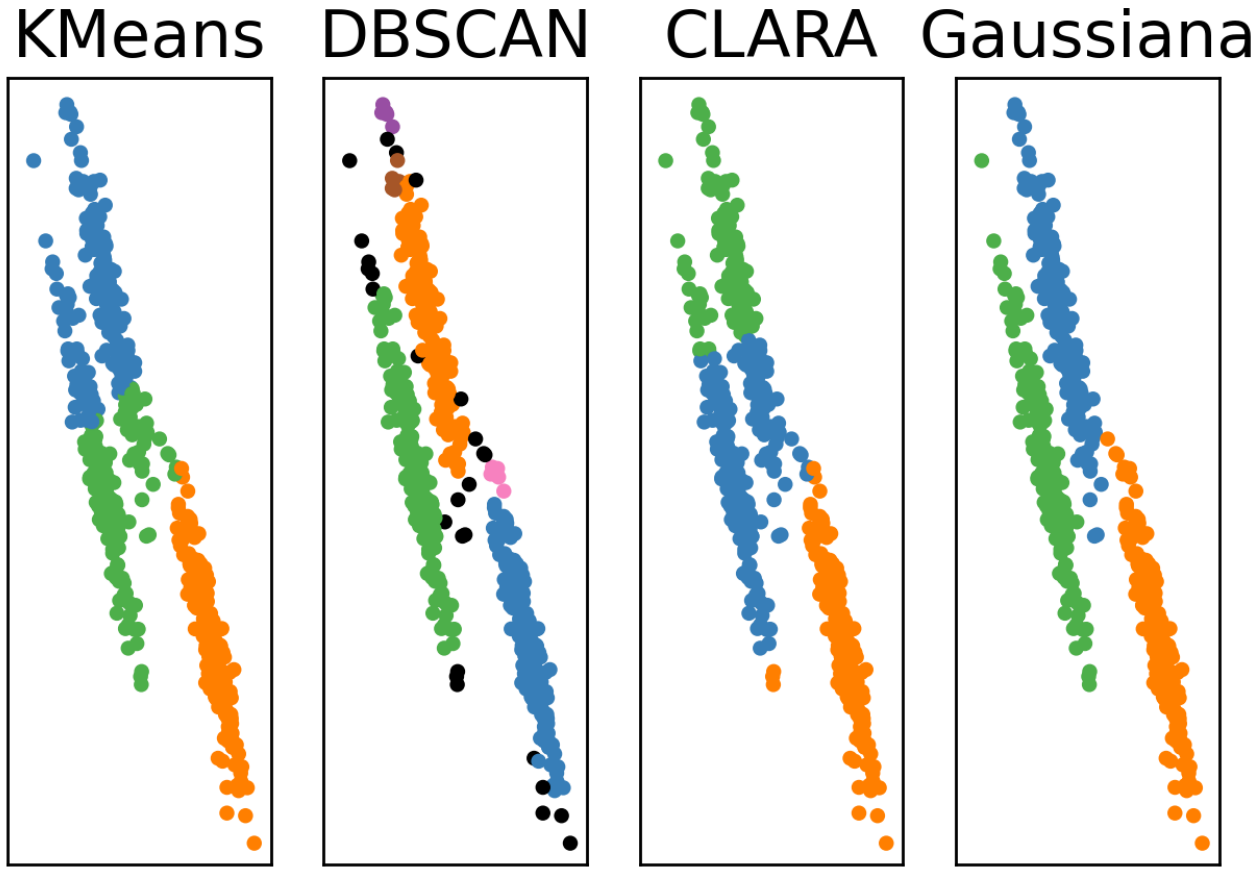
- Más características[3]:
  - Frecuencia normalizada de características categóricas respecto de la clase
  - $FN(d_{ij})_c = \frac{Número\ de\ IPs\ con\ clase\ c\ cuya\ característica\ d_i\ tiene\ valor\ j}{Número\ de\ IPs\ con\ clase\ c}$

IP	Clase	País	País_1	País_3	País_9
123.45.23.12	1	España	0.5	1	0
123.45.23.45	1	Francia	0.5	0	0
187.145.34.56	9	Alemania	0	0	1
123.45.23.7	3	España	0.5	1	0

[3] Renjan, A., Joshi, K.P., Narayanan, S.N., Joshi, A., 2018. Dabr: Dynamic attribute based reputation scoring for malicious ip address detection, in 2018 IEEE International Conference on Intelligence and Security Informatics (ISI)

Modificaciones (III)

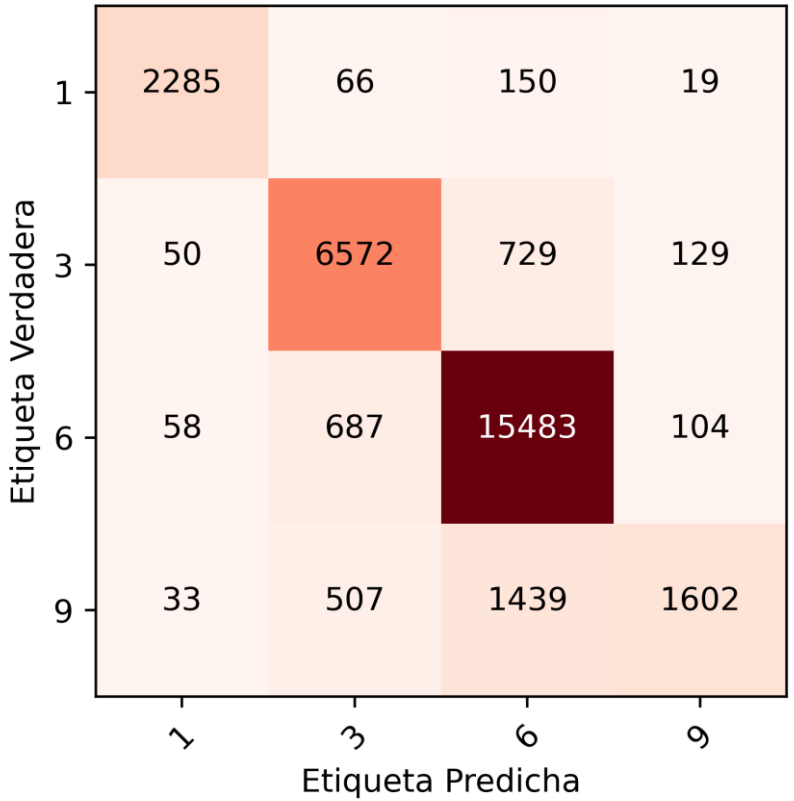
- Mayor variedad de modelos
  - Modelos de mezcla Gaussiana
  - CLARA

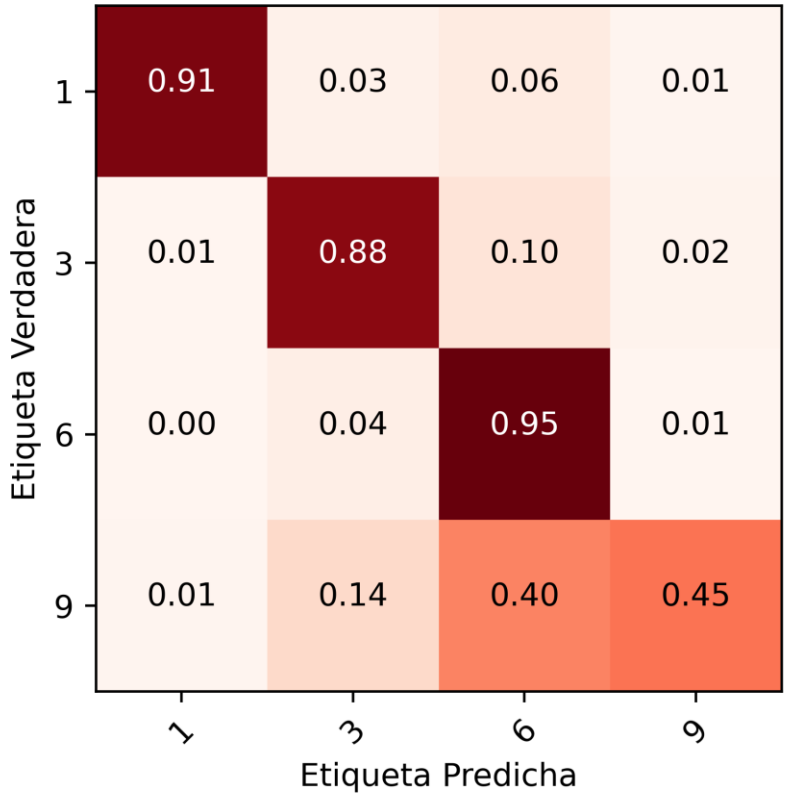


RECSI 2022

Nuevos resultados

Algoritmo	Tasa de acierto	MCC
Mezcla Gaussiana	0.867	0.781





RECSI 2022