



Universidad  
Rey Juan Carlos

ESCUELA TÉCNICA SUPERIOR DE  
INGENIERÍA INFORMÁTICA  
GRADO EN INGENIERÍA DE LA CIBERSEGURIDAD

## **PRÁCTICA 2**

SISTEMAS DE INFORMACIÓN

CURSO 2021 - 2022

**Redactado por: Óscar del Río Jiménez**

**Daniel Paciencia Miguel**

**Rodrigo Regaliza Alonso**

## Índice

ENLACE AL REPOSITORIO DE GITHUB: .....	3
EJERCICIO 2: .....	3
EJERCICIO 3: .....	3
EJERCICIO 4: .....	4
EJERCICIO 5: .....	4
EJERCICIO 6: .....	5

## ENLACE AL REPOSITORIO DE GITHUB:

<https://github.com/Oscarmanz/SSII-GB>

### EJERCICIO 2:

Hemos utilizado los datos de la práctica anterior para recoger los usuarios más críticos en una base de datos actualizable y posteriormente hemos representado esos datos en la web utilizando Flask. Para las webs más inseguras el procedimiento ha sido el mismo, utilizando los datos previos y Flask lo hemos representado en la web.



### EJERCICIO 3:

Para este apartado hemos realizado una ampliación de la funcionalidad de la parte de usuarios críticos del ejercicio 2, añadiendo una opción en la misma página para visualizar el gráfico normal, los que han clicado más del 50% de emails y los que han clicado menos del 50%.

The screenshot shows the 'Top usuarios más críticos' web application interface. It features a dark green header with the title 'Top usuarios más críticos'. Below the header, there is a section titled 'Elija un número y una opción para mostrar el gráfico'. This section contains a text input field labeled 'Numero:' with the value '7', a dropdown menu with a downward arrow, and a blue button labeled 'Enviar'. The dropdown menu is open, showing two options: 'Menos del 50% de spam clicado' and 'Más del 50% de spam clicado'. Below the dropdown, there is a blue button labeled 'Volver'.

## EJERCICIO 4:

Para este apartado primero hacemos una petición a la API web y obtenemos sus resultados en JSON, después transformamos los resultados a HTML con la función `json2html` y luego lo incorporamos a nuestro template html utilizando Flask para que se muestren los datos en una tabla de la siguiente forma.

ID	NAME	PREREQUISITES	RELATED_WEAKNESS	SOLUTIONS	SUMMARY
105	HTTP Request Splitting	User-manipulated HTTP request headers are processed by the web server	CVE-2017-13454	Make sure to install the latest vendor security patches available for the web server. If possible, make use of SSL. Install a web application firewall that has known-vuln-detect	HTTP Request Splitting (also known as HTTP Request Smuggling) is an attack pattern where an attacker attempts to insert additional HTTP requests in the body of the original (encapsulating) HTTP request in such a way that the browser interprets it as one request but the web server interprets it as two. There are several ways to perform HTTP request splitting attacks. One way is to include Length headers in the request to exploit the fact that the browser parsing the request may each use a different method. Another

## EJERCICIO 5:

Para este ejercicio hemos decidido utilizar otra API, en este caso hemos creado un formulario en el que introduces un dominio y a través de la API urlscan se obtiene información sobre él, después repitiendo la mecánica del ejercicio 4 lo mostramos en nuestra plantilla HTML con un formato más legible.

### Información sobre un dominio

Introduce un dominio para obtener toda la información sobre él

Dominio:  Enviar

Volver

Esta es la información sobre el dominio introducido

TASK	STATS	PAGE	_ID	GROUP																																																				
<table><tr><td>visibility</td><td>public</td></tr><tr><td>method</td><td>manual</td></tr><tr><td>domain</td><td>www.aulavirtual.urjc.es</td></tr><tr><td>apexDomain</td><td>urjc.es</td></tr><tr><td>time</td><td>2020-07-13T10:22:54.450Z</td></tr><tr><td>uuid</td><td>e4fc483-d2fa-424c-b28f-a4c3e6b19f0b</td></tr><tr><td>url</td><td>https://www.aulavirtual.urjc.es/moddle/mod/quiz/</td></tr></table>	visibility	public	method	manual	domain	www.aulavirtual.urjc.es	apexDomain	urjc.es	time	2020-07-13T10:22:54.450Z	uuid	e4fc483-d2fa-424c-b28f-a4c3e6b19f0b	url	https://www.aulavirtual.urjc.es/moddle/mod/quiz/	<table><tr><td>uniqueIPs</td><td>5</td></tr><tr><td>uniqueCountries</td><td>2</td></tr><tr><td>dataLength</td><td>4193527</td></tr><tr><td>encodedDataLength</td><td>1294782</td></tr><tr><td>requests</td><td>23</td></tr></table>	uniqueIPs	5	uniqueCountries	2	dataLength	4193527	encodedDataLength	1294782	requests	23	<table><tr><td>country</td><td>ES</td></tr><tr><td>server</td><td>Apache/2.4.41 (Unix) OpenSSL/1.0.2k-fips PHP/7.2.23</td></tr><tr><td>ip</td><td>212.128.240.26</td></tr><tr><td>mimeType</td><td>text/html</td></tr><tr><td>url</td><td>https://www.aulavirtual.urjc.es/moddle/mod/quiz/</td></tr><tr><td>ttlValidDays</td><td>730</td></tr><tr><td>ttlAgeDays</td><td>4</td></tr><tr><td>ttlValidFrom</td><td>2020-07-09T00:00:00.000Z</td></tr><tr><td>domain</td><td>www.aulavirtual.urjc.es</td></tr><tr><td>apexDomain</td><td>urjc.es</td></tr><tr><td>asname</td><td>REDIRIS RedIRIS Autonomous System, ES</td></tr><tr><td>asn</td><td>AS766</td></tr><tr><td>ttlIssuer</td><td>GEANT OV RSA CA 4</td></tr><tr><td>status</td><td>404</td></tr></table>	country	ES	server	Apache/2.4.41 (Unix) OpenSSL/1.0.2k-fips PHP/7.2.23	ip	212.128.240.26	mimeType	text/html	url	https://www.aulavirtual.urjc.es/moddle/mod/quiz/	ttlValidDays	730	ttlAgeDays	4	ttlValidFrom	2020-07-09T00:00:00.000Z	domain	www.aulavirtual.urjc.es	apexDomain	urjc.es	asname	REDIRIS RedIRIS Autonomous System, ES	asn	AS766	ttlIssuer	GEANT OV RSA CA 4	status	404	e4fc483-d2fa-424c-b28f-a4c3e6b19f0b	<ul style="list-style-type: none"><li>1594635774450</li><li>e4fc483-d2fa-424c-b28f-a4c3e6b19f0b</li></ul> <a href="https://artican.c...d2fa-424c-4...">https://artican.c...d2fa-424c-4...</a>
visibility	public																																																							
method	manual																																																							
domain	www.aulavirtual.urjc.es																																																							
apexDomain	urjc.es																																																							
time	2020-07-13T10:22:54.450Z																																																							
uuid	e4fc483-d2fa-424c-b28f-a4c3e6b19f0b																																																							
url	https://www.aulavirtual.urjc.es/moddle/mod/quiz/																																																							
uniqueIPs	5																																																							
uniqueCountries	2																																																							
dataLength	4193527																																																							
encodedDataLength	1294782																																																							
requests	23																																																							
country	ES																																																							
server	Apache/2.4.41 (Unix) OpenSSL/1.0.2k-fips PHP/7.2.23																																																							
ip	212.128.240.26																																																							
mimeType	text/html																																																							
url	https://www.aulavirtual.urjc.es/moddle/mod/quiz/																																																							
ttlValidDays	730																																																							
ttlAgeDays	4																																																							
ttlValidFrom	2020-07-09T00:00:00.000Z																																																							
domain	www.aulavirtual.urjc.es																																																							
apexDomain	urjc.es																																																							
asname	REDIRIS RedIRIS Autonomous System, ES																																																							
asn	AS766																																																							
ttlIssuer	GEANT OV RSA CA 4																																																							
status	404																																																							
		<table><tr><td>country</td><td>ES</td></tr><tr><td>server</td><td>Apache/2.4.41 (Unix) OpenSSL/1.0.2k-fips PHP/7.2.23</td></tr></table>	country	ES	server	Apache/2.4.41 (Unix) OpenSSL/1.0.2k-fips PHP/7.2.23																																																		
country	ES																																																							
server	Apache/2.4.41 (Unix) OpenSSL/1.0.2k-fips PHP/7.2.23																																																							

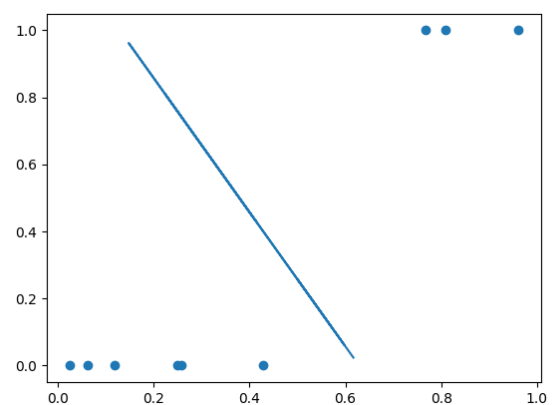
## EJERCICIO 6:

Para este apartado hemos utilizado 3 modelos. Antes de utilizar cada uno hemos cargado los datos del JSON a nuestro código en Python, hemos eliminado los nombres ya que daban problemas y hemos separado los datos en entrenamiento y test. En nuestro caso hemos decidido que el 70% serían para entrenar y el 30% restante para el test.

El primer modelo es el de regresión lineal, lo que hace este modelo es representar en el eje X la probabilidad de pinchar en spam y en el Y si es crítico o no en función de esto. El resultado de la regresión es una recta que separa claramente 2 regiones, si está por debajo será no vulnerable y si está por encima lo será. Aplicado a los datos de test el accuracy es de 66,66% y en el modelo real los resultados obtenidos son los siguientes:

```
El modelo de regresion lineal ha predicho:
3 vulnerables y 6 no vulnerables
El accuracy es de: 66.66666666666666 %

El modelo de regresion lineal sobre los datos reales ha predicho:
7 vulnerables y 23 no vulnerables
-----
```

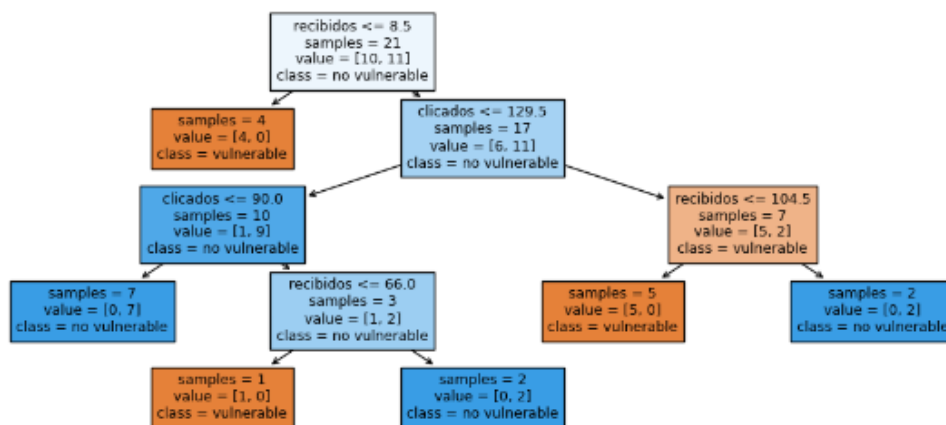


El segundo método es el del árbol de decisión, construimos el modelo y le pasamos los mismos datos que al anterior obteniendo los siguientes resultados y la representación del árbol creado:

```
El modelo de arbol de decisión ha predicho:
5 vulnerables y 4 no vulnerables
El accuracy es de: 88.8888888888889 %

Profundidad del árbol: 4
Número de nodos terminales: 6

El modelo de arbol de decisión sobre los datos reales ha predicho:
10 vulnerables y 20 no vulnerables
```



Por último, el Random Forest que es similar al árbol ya que lo único que hace es crear varios árboles más sencillos con conjuntos de datos diferentes, luego para clasificar hace un consenso entre todos y deciden cual es el mejor resultado. Se obtiene lo siguiente y los gráficos de los árboles se crean en una carpeta llamada randomForestTrees.

```
El modelo de random forest ha predicho:
4 vulnerables y 5 no vulnerables
El accuracy es de: 77.7777777777779 %

El modelo de random forest sobre los datos reales ha predicho:
12 vulnerables y 18 no vulnerables
```

Vemos que el árbol de decisión es el mejor modelo con un accuracy del 88% seguido del random forest y por último el de regresión. Sorprende que el random forest no sea el mejor pero probablemente se deba a la escasez de datos para el análisis ya que los árboles salen muy sencillos y es razonable que con un conjunto de datos mayor los resultados se ajustasen más a la realidad y el random obtuviese los mejores resultados.