# DOCUMENTACIÓN VARIABLES DEL DATASET

[DOCUMENTACIÓN VARIABLES DEL DATASET 1](#_Toc172201684)

[**1.** **Definición de variables independientes** 2](#_Toc172201685)

[**1.1.** **Variables del bloque USERS** 2](#_Toc172201686)

[**1.2.** **Variables del bloque PASSWORD** 4](#_Toc172201687)

[**1.3.** **Variables del bloque CHANNEL** 11](#_Toc172201688)

[**1.4.** **Variables del bloque FILE** 16](#_Toc172201689)

[**2.** **Definición de la variable dependiente** 17](#_Toc172201690)

## **Definición de variables independientes**

### **Variables del bloque USERS**

|  |  |  |  |
| --- | --- | --- | --- |
| **Variable name** | **Type** | **Description** | **Score Type** |
| ***VIP\_credentials*** | Categorical | VIP user / not VIP user | 0 / 10 |

Esta variable de tipo categórico define si un usuario es VIP o no lo es, tomando como VIP los usuarios registrados como tal para la detección del Bot.

**Asignación de valores:**

* Se asigna una puntuación de 0 las credenciales que sean de un usuario no VIP, entendiendo que estas presentan un menor riesgo.
* Se asigna una puntuación de 10 a las credenciales que sean de un usuario VIP, entendiendo que estas presentan un mayor riesgo.

|  |  |  |  |
| --- | --- | --- | --- |
| **Variable name** | **Type** | **Description** | **Score Type** |
| ***VIP\_group*** | Categorical | Groups: "CEO", "Directiva", "Ciber", "otros" | 0 - 10 |

Esta variable de tipo categórico define el grupo VIP al que corresponde el usuario, siguiendo un orden jerárquico partiendo del grupo VIP 1 como la categoría de más alto nivel, hasta tantos grupos VIP como se quieran establecer.

* ***VIP 1***
* ***VIP 2***
* ***VIP 3***
* ***VIPX…***
* ***Otros***

Para el desarrollo del Dataset en este caso se han tenido en cuenta 3 categorías de usuarios Vip: VIP 1 = “CEO”; VIP 2 = “Directiva” y VIP 3 = “Departamento de Ciberseguridad”.

**Asignación de valores:**

Se asignan los valores de menor a mayor grado de riesgo a los distintos grupos de VIP según se considera la importancia que representa cada uno de ellos.

* Se asigna una puntuación de 0 a las credenciales que no se correspondan con ninguna de las categorías VIP establecidas, es decir la categoría “Otros”.
* Se asigna una puntuación de 5 a las credenciales que se correspondan con la categoría de “Departamento Ciberseguridad”.
* Se asigna una puntuación de 8 a las credenciales que se correspondan con la categoría de “Directiva”.
* Se asigna una puntuación de 10 a las credenciales que se correspondan con la categoría de “CEO”.

|  |  |  |  |
| --- | --- | --- | --- |
| **Variable name** | **Type** | **Description** | **Score Type** |
| ***User\_status*** | Categorical | Active / Inactive user (check with the AD if the user activity) | 0 / 10 |

Esta variable de tipo categórico define si el usuario de la credencial detectada en el fichero está activo actualmente o está inactivo. Se comprueba la actividad del usuario con el directorio de actividad (AD), donde se indica si el usuario está activo o inactivo en casos de estar jubilado, retirado, etc.

**Asignación de valores:**

* Se asigna una puntuación de 0 a las credenciales que correspondan a un usuario inactivo, entendiendo que tienen menor riesgo por ser credenciales en desuso.
* Se asigna una puntuación de 10 a las credenciales que correspondan a un usuario activo, entendiendo que tienen mayor riesgo por ser credenciales en uso.

### **Variables del bloque PASSWORD**

|  |  |  |  |
| --- | --- | --- | --- |
| **Variable name** | **Type** | **Description** | **Score Type** |
| ***Password update*** | Categorical | Actual (contains "2024") / not actual (does not contain "2024") | 0 / 10 |

Esta variable de tipo categórico define la credencial como actual si la contraseña contiene las cadenas de caracteres “2024” y como no actual si no contiene ninguna de esas cadenas de caracteres.

**Asignación de valores:**

* Se asigna una puntuación de 0 a las credenciales que no se consideran actuales, entendiendo que estas suponen un menor riesgo por estar más obsoletas.
* Se asigna una puntuación de 10 a las credenciales que se consideran actuales, entendiendo que estas suponen un mayor riesgo por su alta posibilidad de estar en uso o de haberse creado recientemente.

|  |  |  |  |
| --- | --- | --- | --- |
| **Variable name** | **Type** | **Description** | **Score Type** |
| ***Password type*** | Categorical | Types: Password for change; Default Password;  Personal Password | 0 - 10 |

Esta variable de tipo categórico define si la credencial contiene una contraseña con el patrón de la contraseña asignada por defecto, de la contraseña utilizada para forzar un cambio de contraseña al usuario, o si, por el contrario, no contiene ninguno de estos dos patrones y por tanto se trata de una contraseña única personalizada.

Como norma general, la mayoría de las empresas tienen una política de contraseña que exige el cambio de la contraseña asignada por defecto en un corto periodo de tiempo, por motivos de seguridad. Por otro lado, las contraseñas para forzar un cambio tienen como objetivo el cabio inmediato de contraseña por parte del usuario.

**Asignación de valores:**

* Se asigna una puntuación de 0 a las credenciales que tienen la contraseña para el cambio, entendiendo que estas suponen un menor riesgo ya que se asume que el usuario debería haberla cambiado de inmediato y por lo tanto con alta probabilidad no están en uso.
* Se asigna una puntuación de 2 a las credenciales que tienen la contraseña por defecto, entendiendo que estas suponen un bajo riesgo por la alta probabilidad de estar obsoletas.
* Se asigna una puntuación de 10 a las credenciales que no tienen la contraseña para el cambio, entendiendo que estas suponen un mayor riesgo ya que son con más alta probabilidad contraseñas en uso, al tratarse de contraseñas personalizadas.

|  |  |  |  |
| --- | --- | --- | --- |
| **Variable name** | **Type** | **Description** | **Score Type** |
| ***Leaked password*** | Numerical | Number of times a password has been compromised according to HaveIBeenPwned's source | 0 - 10 |

Esta variable de tipo numérico define el número de veces que una credencial detectada en un fichero ha sido comprometida anteriormente. Se obtiene este número gracias a un Script que utiliza la fuente “HaveIBeeenPwned”.

**Asignación de valores:**

Se asignan las puntuaciones entre 0 y 10 a los siguientes intervalos definidos, entendiendo que el riesgo será menor cuantas menos veces haya sido comprometida la contraseña anteriormente, y que el riesgo será mayor cuantas más veces se haya comprometido la contraseña, entendiendo así que ha tenido un mayor alcance su exfiltración.

* 0: 0 veces
* 1: de 1 a 9 veces
* 2: De 10 a 99 veces
* 3: De 100 a 999 veces
* 4: De 1,000 a 9,999 veces
* 5: De 10,000 a 49,999 veces
* 6: De 50,000 a 99,999 veces
* 7: De 100,000 a 399,999 veces
* 8: De 400,000 a 799,999 veces
* 9: De 800,000 a 999,999 veces
* 10: Más de 1,000,000 veces

**Complejidad de contraseñas:**

Para determinar la complejidad de las contraseñas de las credenciales detectadas en los ficheros analizados, se utiliza un script de Python que usa la librería “*zxcvbn”* desarrollada por Dropbox para evaluar la complejidad que presenta una contraseña.

La herramienta utiliza técnicas avanzadas para esta evaluación, analizando patrones para entender la estructura de la contraseña, estimando el número mínimo de intentos necesarios para adivinar una contraseña y la evaluación del tiempo de crackeo.

|  |  |  |  |
| --- | --- | --- | --- |
| **Variable name** | **Type** | **Description** | **Score Type** |
| **Guesses discover** | Numerical | Number of attempts it takes to guess the password by a brute force attack | 0 - 10 |

Esta variable de tipo numérico define el número de intentos que un atacante necesitaría para adivinar la contraseña. Se utiliza el logaritmo base 10 (log10) para expresar el número de intentos estimados ya que así se convierten grandes números en una escala más pequeña y manejable para el humano, que facilita comprender la fortaleza de una contraseña. (Ejemplo: Un guesses\_log10 de 3 significa 10^3 o 1000 intentos).

Para esta estimación, la herramienta *zxcvbn* considera los ataques de diccionario, de fuerza bruta y combina la información de ambos ataques ajustando las estimaciones basadas en la longitud y la complejidad detectada en los patrones.

**Asignación de valores:**

Se asignan las puntuaciones entre 0 y 10 a los siguientes intervalos definidos, entendiendo que cuanto mayor sea el valor que devuelve guesses\_log10, más fuerte es la contraseña, y por tanto presenta un menor riesgo; y que cuanto menor sea el valor de guesses\_log10, más débil es la contraseña, y por tanto presenta un mayor riesgo al ser más vulnerable.

* 0: Guesses\_log10 >= 12 (> 1 billón intentos)
* 1: De 10 a 12 guesses
* 2: De 9 a 10 guesses
* 3: De 8 a 9 guesses
* 4: De 7 a 8 guesses
* 5: De 6 a 7 guesses
* 6: De 5 a 6 guesses
* 7: De 4 a 5 guesses
* 8: De 3 a 4 guesses
* 9: De 2 a 3 guesses
* 10: Guesses\_log10 < 2 (< 100 intentos)

|  |  |  |  |
| --- | --- | --- | --- |
| **Variable name** | **Type** | **Description** | **Score Type** |
| **Cracking time** | Numerical | Estimated time to guess the password | 0 - 10 |

Esta variable de tipo numérico define el tiempo estimado que tomaría adivinar la contraseña bajo el escenario de un ataque online sin limitaciones, asumiendo 10 intentos por segundo. Se utiliza una representación legible para humanos que transforma los intentos por segundo en segundos. La librería “*zxcvbn”* devuelve el valor en unidades de tiempo medidas tanto en segundos, minutos, horas, días, meses y años, como en un total de segundos. Se utilizan para el cálculo en el dataset los valores medidos en segundos para utilizar la misma unidad de tiempo para todas las contraseñas.

**Asignación de valores:**

Se asignan las puntuaciones entre 0 y 10 a los siguientes intervalos definidos, entendiendo que cuanto más tiempo tome adivinar la contraseña, menor riesgo representa esta por ser más fuerte, y que cuanto menos tiempo tome adivinarla, mayor riesgo representará por ser más débil. Se especifica para cada intervalo el valor en segundos correspondiente del extremo máximo.

* 0: Más de 10000 años
* 1: 1000 años a 10000 años (315360000000s)
* 2: 100 años a 1000 años (31536000000s)
* 3: 10 años a 100 años (3153600000s)
* 4: 1 año a 10 años (315360000s)
* 5: 1 mes a 1 año (31536000s)
* 6: 1 semana a 1 mes (2592000s)
* 7: 1 día a 1 semana (604800s)
* 8: 1 hora a 1 día (86400s)
* 9: 5 minutos a 1 hora (3600s)
* 10: Menos de 5 minutos (300s)

|  |  |  |  |
| --- | --- | --- | --- |
| **Variable name** | **Type** | **Description** | **Score Type** |
| **Password strength** | Numerical | Score in a scale from 1 to 4 that represents the complexity of a password | 0 - 10 |

Esta variable numérica define la puntuación que la biblioteca “*zxcvbn*” le asigna a la contraseña analizada para indicar la fuerza que presenta. Esta escala va desde 0 hasta 4, representando el 0 a las contraseñas más débiles y el 4 a las más fuertes.

**Asignación de valores:**

* Se asigna una puntuación de 0 a las credenciales que obtengan un score de 4, al considerarse contraseñas muy fuertes representan un menor riesgo.
* Se asigna una puntuación de 3 a las credenciales que obtengan un score de 3, al considerarse contraseñas fuertes.
* Se asigna una puntuación de 6 a las credenciales que obtengan un score de 2, al considerarse contraseñas de fortaleza media.
* Se asigna una puntuación de 8 a las credenciales que obtengan un score de 1, al considerarse contraseñas débiles.
* Se asigna una puntuación de 10 a las credenciales que obtengan un score de 0, al considerarse contraseñas muy débiles representan un mayor riesgo (puede ser adivinada casi inmediatamente).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Variable name** | **Type** | **Units** | **Description** | **Score Type** |
| ***Password entropy*** | Numerical | Bits | Number of bits that represents the uncertainty of a password | 0 - 10 |

Esta variable de tipo continuo define el valor de entropía que presenta la contraseña de la credencial detectada en el fichero, es decir, cuál es su incertidumbre o desorden, o cuán de difícil es de adivinar mediante un ataque de fuerza bruta. Para calcular la entropía se utiliza un script de Python que calcula tanto la entropía de Shannon, como

**Asignación de valores:**

Se asignan las puntuaciones entre 0 y 10 a los siguientes intervalos definidos, entendiendo que el riesgo será menor cuanto más baja sea la entropía y que el riesgo será mayor cuanto más alta sea la entropía.

* 0: 100 o más bits de entropía
* 1: De 90 a 99 bits de entropía
* 2: De 80 a 89 bits de entropía
* 3: De 70 a 79 bits de entropía
* 4: De 60 a 69 bits de entropía
* 5: De 50 a 59 bits de entropía
* 6: De 40 a 49 bits de entropía
* 7: De 30 a 39 bits de entropía
* 8: De 20 a 29 bits de entropía
* 9: De 10 a 19 bits de entropía
* 10: Menos de 10 bits de entropía

### **Variables del bloque CHANNEL**

|  |  |  |  |
| --- | --- | --- | --- |
| **Variable name** | **Type** | **Description** | **Score Type** |
| ***Chat type*** | Categorical | Group / channel | 0 / 10 |

Esta variable de tipo categórico define si el fichero analizado ha sido compartido a través de un grupo o un canal de Telegram. Se realiza esta diferenciación debido a que se considera que los canales tienen un mayor alcance ya que están diseñados para difundir mensajes a grandes audiencias por sus ilimitados suscriptores; mientras que los grupos, a pesar de que se crean para que los usuarios puedan colaborar, tienen un número limitado de suscriptores.

**Asignación de valores:**

* Se asigna una puntuación de 0 a los archivos que han sido compartidos a través de un grupo, entendiendo que tienen menor riesgo por tener un menor alcance.
* Se asigna una puntuación de 10 a los archivos que han sido compartidos a través de un canal, entendiendo que tienen mayor riesgo por tener un mayor alcance.

|  |  |  |  |
| --- | --- | --- | --- |
| **Variable name** | **Type** | **Description** | **Score Type** |
| ***Channel privacity*** | Categorical | Public / Private | 0 / 10 |

Esta variable de tipo categórico define si el fichero analizado ha sido compartido a través de un canal público o privado. Se considera que los mensajes enviados por canales privados pueden ser más sensibles ya que estos tienen una política de acceso más restrictiva que los canales públicos (se necesita ser añadido por el propietario u obtener enlace de invitación).

**Asignación de valores:**

* Se asigna una puntuación de 0 a los archivos que han sido compartidos a través de un canal público, entendiendo que tienen menor riesgo por su sencilla accesibilidad.
* Se asigna una puntuación de 10 a los archivos que han sido compartidos a través de un canal privado, entendiendo que tienen mayor riesgo por su complicada accesibilidad. La opción de monitorizar un canal privado queda subyugada a los requisitos legales vigentes, esto es, una autorización judicial.

|  |  |  |  |
| --- | --- | --- | --- |
| **Variable name** | **Type** | **Description** | **Score Type** |
| ***Subscribers*** | Numerical | Number of channel or group subscribers | 0 - 10 |

Esta variable de tipo numérico define el número de suscriptores que tiene un canal o grupo para medir el alcance del mismo. Dentro de esta variable se realiza una diferenciación a la hora de asignar valores y establecer intervalos condicionada por la variable de “C*hat\_type*”, es decir, se diferencia entre público o privado para ajustar la variable a las características de cada uno.

De esta manera, cuando se trata de un canal privado, los extremos establecidos para considerar el riesgo que presenta el archivo son más bajos, ya que el acceso a este tipo de canales es más restringido. En cambio, cuando se trata de un canal público, los extremos establecidos para determinar el riesgo son más altos, al tratarse de un tipo de canal accesible para cualquier usuario.

**Asignación de valores:**

Se asignan las puntuaciones entre 0 y 10 a los siguientes intervalos definidos, entendiendo de manera general para ambos casos que cuantos menos suscriptores tenga el canal menor riesgo representa, y cuantos más suscriptores tenga, mayor riesgo representa, sea público o privado.

A continuación, se muestran las asignaciones de puntos para cada tipo de canal:

* **Canales públicos**
* 0: Menos de 1.000 suscriptores
* 1: De 1.000 a 1.999 suscriptores
* 2: De 2.000 a 2.499 suscriptores
* 3: De 2.500 a 2.999 suscriptores
* 4: De 3.000 a 3.499 suscriptores
* 5: De 3.500 a 3.999 suscriptores
* 6: De 4.000 a 4.999 suscriptores
* 7: De 5.000 a 5.999 suscriptores
* 8: De 6.000 a 6.999 suscriptores
* 9: De 7.000 a 8.499 suscriptores
* 10: Más de 8.500 suscriptores
* **Canales privados**
* 0: Menos de 20 suscriptores
* 1: De 20 a 59 suscriptores
* 2: De 60 a 99 suscriptores
* 3: De 100 a 139 suscriptores
* 4: De 140 a 179 suscriptores
* 5: De 180 a 219 suscriptores
* 6: De 220 a 259 suscriptores
* 7: De 260 a 299 suscriptores
* 8: De 300 a 399 suscriptores
* 9: De 400 a 499 suscriptores
* 10: Más de 500 suscriptores

**Estadísticas de canales de Telegram**

Para incluir variables en referencia a las características sobre un canal en Telegram se utiliza la herramienta ***telemetr.io***, la cual ofrece estadísticas de los canales de Telegram. Gracias a esta herramienta se han podido incluir las siguientes variables de canal:

|  |  |  |  |
| --- | --- | --- | --- |
| **Variable name** | **Type** | **Description** | **Score Type** |
| ***Engagement rate*** | Numerical | The ratio of the average number of views per post to the number of subscribers | 0 - 10 |

Esta variable de tipo numérico define el índice de compromiso, es decir, hace referencia a la actividad que tiene el canal actualmente, a través de la media del número de visitas por post respecto al número de suscriptores que tiene el canal. Además, esta medida se actualiza diariamente en la plataforma.

**Asignación de valores:**

Se asignan las puntuaciones entre 0 y 10 a los siguientes intervalos definidos, entendiendo que cuanto menos porcentaje de índice de compromiso tenga en canal menor es su actividad y por tanto corresponde a un canal menos crítico.

* 0: Entre 0% y 3%
* 1: Del 3% al 5%.
* 2: Del 5% al 8%
* 3: Del 8% al 10%
* 4: Del 10% al 12%
* 5: Del 12% al 15%
* 6: Del 15 % al 17%
* 7: Del 17% al 20%
* 8: Del 20% a 25%
* 9: Del 25% al 30%
* 10: Más del 30%

|  |  |  |  |
| --- | --- | --- | --- |
| **Variable name** | **Type** | **Description** | **Score Type** |
| ***Mentions*** | Numerical | Total count of channel mentions over its existence | 0 - 10 |

Esta variable de tipo numérico define el número total de veces que un canal ha sido mencionado por otros canales en sus publicaciones desde que el canal fue creado. Indica cuánto de promocionado y cuánto alcance podría tener un canal en Telegram.

**Asignación de valores:**

Se asignan las puntuaciones entre 0 y 10 a los siguientes intervalos definidos, entendiendo que cuantas menos menciones tenga en canal menor es su popularidad y cuantas más menciones tenga mayor será ésta y por tanto más alcance tendrán sus mensajes, siendo más crítico el archivo proveniente de estos canales.

* 0: Entre 0 y 10 menciones
* 1: De 10 a 20 menciones
* 2: De 20 a 30 menciones
* 3: De 30 a 50 menciones
* 4: De 50 a 65 menciones
* 5: De 65 a 80 menciones
* 6: De 80 a 100 menciones
* 7: De 100 a 120 menciones
* 8: De 120 a 140 menciones
* 9: De 140 a 160 menciones
* 10: Más de 160 menciones

|  |  |  |  |
| --- | --- | --- | --- |
| **Variable name** | **Type** | **Description** | **Score Type** |
| ***Posts day*** | Numerical | Average number of posts per day for the last 30 days | 0 - 10 |

Esta variable de tipo numérico define el valor medio de posts que un canal publica por día, medido en base a las publicaciones de los últimos 30 días. Esta variable refleja la dinámica de publicaciones que presenta un canal, representando así su actividad recientemente.

**Asignación de valores:**

Se asignan las puntuaciones entre 0 y 10 a los siguientes intervalos definidos, entendiendo que cuanto menor sea la media de publicaciones por día menos actividad presenta el canal y por tanto es menos crítico, y que por el contrario cuando mayor sea este valor, más intensa será su actividad y por tanto más criticidad representará.

* 0: 0 publicaciones
* 1: 1 publicación
* 2: 2 publicaciones
* 3: 3 publicaciones
* 4: 4 publicaciones
* 5: 5 publicaciones
* 6: 6 publicaciones
* 7: 7 publicaciones
* 8: 8 publicaciones
* 9: 9 publicaciones
* 10: 10 o más de 10 publicaciones

|  |  |  |  |
| --- | --- | --- | --- |
| **Variable name** | **Type** | **Description** | **Score Type** |
| ***Reposts*** | Numerical | Average number of shares per post over the past 30 days | 0 - 10 |

Esta variable de tipo numérico define el valor medio de reenvíos por publicación que un presenta un canal, medido en base a las publicaciones de los últimos 30 días. Esta variable refleja cuánto de viral es el contenido de un canal y su calidad.

**Asignación de valores:**

Se asignan las puntuaciones entre 0 y 10 a los siguientes intervalos definidos, entendiendo que cuanto menor sea la media de reenvíos por publicación, menos viral es el contenido del canal y por tanto representa un menor riesgo, pero cuanto mayor sea más riesgo representa por ser más viral y de mayor calidad.

* 0: 0 reenvíos
* 1: 1 reenvío
* 2: 2 reenvíos
* 3: 3 reenvíos
* 4: 4 reenvíos
* 5: 5 reenvíos
* 6: 6 reenvíos
* 7: 7 reenvíos
* 8: 8 reenvíos
* 9: 9 reenvíos
* 10: 10 o más de 10 reenvíos

|  |  |  |  |
| --- | --- | --- | --- |
| **Variable name** | **Type** | **Description** | **Score Type** |
| ***Channel country*** | Categorical | “Spain” / other | 0 / 10 |

Esta variable de tipo categórico define si el canal del cual proviene el mensaje es un canal creado en España o si por el contrario ha sido creado en otro país distinto. Se define esta variable para darle valor a los archivos que provengan de canales españoles por el interés del contenido que se está monitorizando, pero se podría sustituir España por cualquier otro país.

**Asignación de valores:**

* Se asigna una puntuación de 0 a los archivos que han sido compartidos a través de un canal procedente de otro país distinto a España.
* Se asigna una puntuación de10 a los archivos que han sido compartidos a través de un canal procedente de España.

### **Variables del bloque FILE**

|  |  |  |  |
| --- | --- | --- | --- |
| **Variable name** | **Type** | **Description** | **Score Type** |
| ***Country file name*** | Categorical | File name is: "Spain" or similar, "Europe" or similar, or other | 0 - 10 |

Esta variable de tipo categórico define la categoría que se asigna a la credencial en función del nombre del fichero analizado. Se evalúa si este contiene la palabra “Spain” en cualquier idioma o alguna similar que haga referencia a España, si contiene “Europe” en cualquier idioma o alguna similar que haga referencia a Europe, o si por el contrario no contienen ninguna de estas referencias en el nombre de archivo. Se entiende que tiene más riesgo un archivo que contiene credenciales asociadas a España y Europa, ya que las credenciales que se monitorean tienen relación con este país / continente.

**Asignación de valores:**

* Se asigna una puntuación de 0 a los archivos que no contienen ni “Spain” ni “Europe” o similares en el nombre de archivo.
* Se asigna una puntuación de 5 a los archivos que contienen “Europe” en el nombre de archivo.
* Se asigna una puntuación de 10 a los archivos que contienen “Spain” en el nombre de archivo.

|  |  |  |  |
| --- | --- | --- | --- |
| **Variable name** | **Type** | **Description** | **Score Type** |
| ***Results file*** | Numerical | Amount of results in the analyzed file | 0 - 10 |

Esta variable de tipo numérico determina el número de resultados positivos que se han detectado en el fichero analizado, es decir, la cantidad de credenciales exfiltradas de la empresa que se detectan en un fichero.

**Asignación de valores:**

Se asignan las puntuaciones entre 0 y 10 a los siguientes intervalos definidos, entendiendo que cuantos menos resultados positivos contenga el fichero menor grado de riesgo representa, y cuantos más resultados contenga representa un mayor riesgo.

* 0: 1 resultado
* 1: 2 resultados
* 2: 3 resultados
* 3: 4 resultados
* 4: 5 resultados
* 5: 6 resultados
* 6: 7 resultados
* 7: 8 resultados
* 8: 9 resultados
* 9: 10 resultados
* 10: 12 o más resultados

## **Definición de la variable dependiente**

|  |  |  |
| --- | --- | --- |
| **Variable name** | **Type** | **Description** |
| ***TOTAL*** | Numerical | Total score assigned to the credential that determine the risk level |

Esta variable de tipo numérico se corresponde con la variable dependiente que determina el riesgo que presenta una credencial en base al resto de variables independientes del dataset.