Detección de influencers en RRSS

# Introducción

Una de las funcionalidades incorporadas al proyecto Social Brand es la búsqueda de influencers en redes. Con ello se pretende encontrar usuarios con una influencia suficiente en redes sociales como Twitter e Instagram. El objetivo es localizar posibles contratos publicitarios con la marca que use la aplicación.

Para ello, se utilizará la aplicación de ingesta de datos ya desarrollada. Gracias a dicha aplicación se buscarán perfiles de usuarios relacionados con las *keywords* que establezca la marca que use la aplicación.

Se obtendrá un conjunto de perfiles de relacionados donde buscaremos los influencers.

Obviamente no todos estos usuarios serán válidos, por lo que hay que hacer una criba para mantener únicamente los influencers potenciales.

Para ello, se han establecido una serie de criterios usados en una métrica que indica el grado de influencia del usuario.

# Requisitos

Para el correcto funcionamiento de este script son necesarios los siguientes archivos:

· *ingest\_extraction\_v2.py* – Permite la extracción y filtrado de los dataframes a partir del json de entrada.

· Librería *requests* – Para poder expandir las url.

· **LO QUE SEA QUE USE LUIS PREGUNTALE!!!!!**

# Criterios

**Followers y bots**

El primer criterio establecido es el número de seguidores que tenga la cuenta. Obviamente cuanta más gente siga al usuario, más lejos llegará la publicidad que este haga. Sin embargo, hay que comprobar que dichos followers sean gente real. Esto se debe a que es una práctica común comprar bots para aumentar el número de seguidores.

Por tanto, la aplicación es capaz de filtrar, devolviendo un porcentaje respecto al número de seguidores, aquellas cuentas cuyo porcentaje de bots supera el valor de 40 %. Si esto ocurre se considerará el perfil como *no válido* y se desechará.

**Otras redes sociales**

A la hora de buscar un influencer, se busca también que tenga influencia en otras plataformas. Entre estas plataformas, se valora especialmente aquellos usuarios que además tengan un canal de youtube, pues es una plataforma con mucha interacción, donde muchas veces siguen más personas al influencer que en el propio Twitter

Otras plataformas como blogspot o Tumblr también son bien valoradas.

**Verificación**

Otro campo a tener en cuenta es si la cuenta está verificada. Normalmente se verifican cuentas que tienen gran presencia social o una credibilidad elevada. También aquellas con riesgo de suplantación.

Todo esto hace que una cuenta verificada tenga más probabilidades de ser catalogada como influencer que una que no

**Engagement**

El engagement, o participación, es una medida muy útil a la hora de calcular la influencia en las redes de un usuario. Dicho valor se obtiene al hacer la media del número de favoritos o RTs que tienen los posts del 4 al 10 y dividir el resultado entre el número de followers. Si además multiplicamos lo obtenido por 100, se saca el porcentaje de participación de los seguidores.

Se calcula en los posts del 4 al 10 ya que los primeros tres posts, al ser más nuevos, siguen evolucionando, y los posts a partir del décimo no son tan valiosos ya que nos interesa más el engagement actual que el pasado.

**Grupos de influencia**

Por último, aunque de una complejidad mayor, existen los llamados grupos de influencia. Estos son grupos de usuarios con acuerdos entre ellos, a veces económicos, para comentarse y darse feedback. Dicho feedback se realiza de manera muy rápida, lo que hace que los algoritmos de Twitter le den más visibilidad. Sin embargo, son comentarios vacíos y que no aportan nada. Además de que el influencer carece de un público real.

# Métrica

Una vez tenemos todos estos datos, se hace una métrica aplicando distintos pesos a los valores.

En primer lugar, se ha dado un peso de **0.20** a aquellas cuentas que tengan **0% de bots**, variando dicho peso de forma inversamente proporcional respecto al porcentaje de bots.

Si el usuario posee un **canal de youtube**, esto le aportará un peso de **0.20**. (Esto se puede extender a varias redes, dando un mayor peso cuanto mayor sea el número de redes en las que se encuentre el influencer)

En cuanto a si la **cuenta está verificada**, se ha decidido darle un peso de **0.25**, ya que es un criterio que da bastante certeza de que el usuario es alguien importante. Además, al ser un valor cualitativo, pues será cierto o no, provoca que se note más la diferencia entre aquellos usuarios no verificados y los que sí lo están.

Por último, se ha dado un peso de **0.35** al **engagement**, pues es un criterio cuantitativo que aporta bastante información acerca de si llegan o no los mensajes del supuesto a influencer a sus seguidores que, al final, es el objetivo de esta funcionalidad.

# Código

A continuación se pretende explicar el código que permite sacar la lista con los datos de los posibles influencers. Se dividirá la explicación según los criterios.

**Funciones generales**

En primer lugar tenemos una serie de funciones generales que se utilizan para encapsular la extracción y el filtrado de los dataframes, pues esto se hace en otro fichero, *ingest\_extraction\_v2*.

Además, tenemos la función general, *get\_ranking()*, que recibe un dataframe con datos y, tras aplicar las métricas necesarias, devuelve una lista con los usuarios aceptados como posibles influencers y otros datos útiles de los mismos, como la puntuación obtenida, el nombre, etc…

Por último tenemos el método *get\_score()*, que recibe un id de usuario y el dataframe con la información y saca la puntuación del usuario tras aplicarle las métricas previamente mencionadas. Es en esta función donde se aplican los pesos de los criterios. Si el porcentaje de bots supera el 40% es en este método donde se desechará el perfil, devolviendo un score de -1.

**Porcentaje de bots**

Para hallar el porcentaje de bots se utilizan los métodos *get\_num\_followers(), get\_num\_bots()* y *get\_bot\_percentage()*

El primero recibe un perfil y devuelve el número de seguidores que tiene este.

**PREGUNTAR LUIS!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!**

El segundo recibe un dataframe con el perfil a analizar y devuelve el número de bots que se estima hay entre los seguidores del perfil.

Por último, el método *get\_bot\_percentage()* recibe un dataframe con los datos del perfil y llama a los dos anteriores, calculando y devolviendo posteriormente el porcentaje de bots.

**Verificado**

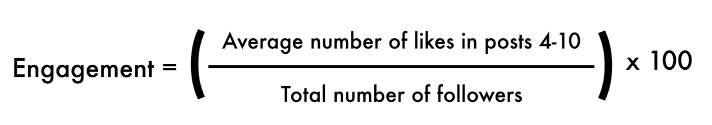
Para saber si un perfil está verificado o no se usa el método *is\_verified(),* que simplemente recibe un perfil y devuelve 100 si su campo Is\_Verified está a True y 0 si está a False.

**Otras redes sociales**

Utiliza el método *has\_other\_social\_network()*, que recibe un perfil y, mirando en la biografía, en la url del perfil y en los tweets, busca otras redes sociales en las que esté el usuario. El mayor peso que puede obtener es 0.20 si tiene canal de youtube y blog, por lo que el método devolverá 1 si posee ambos, 0.75 si solo posee canal de youtube, y 0.25 si posee cuenta de Tumblr o Blogspot.

El método se ayuda de la función *expand\_url()*, que obtiene la url completa para poder ser analizada.

**Engagement**

Como se ha comentado anteriormente, el engagement se obtiene de la fórmula: 

Por tanto, necesitamos algunos datos para calcularlo.

Para el número de followers que tiene el perfil se vuelve a usar el método utilizado para calcular el porcentaje de bots, mientras que para la media de likes primero tenemos que obtener los posts deseados. Estos se obtienen en el método *get\_posts()*, que recibe un dataframe con los tweets de un usuario, así como el post de inicio y el post final y devuelve una lista con los posts que se encuentren entre los valores especificados. Para nuestro caso, se llamará al método con los valores 4,10

A la hora de calcular la media de likes, se llama al método *get\_average\_likes()*, que recibe la lista de posts procedente de *get\_posts()* y calcula y devuelve la media de sus likes.

Por último, el método *get\_engagement()* combina los anteriores obteniendo los valores necesarios para usar la fórmula anterior. Devolverá pues el porcentaje de participación de los seguidores.

# Referencias

<https://hackernoon.com/how-to-reach-out-to-instagram-influencers-to-maximize-response-rates-16f429008f62>

<https://hackernoon.com/how-to-measure-what-an-instagram-influencer-is-worth-cb488f605ffe>

<https://www.agorapulse.com/es/blog/encontrar-influencers-facilmente>