

# Evaluación 2

Natalie Julian

## ¿Cómo saber si una cuenta es *fake*?

Instagram es una red social lanzada al mundo el 10 de Octubre de 2010. La dinámica es simple: los usuarios comparten fotos y videos de forma instantánea, pudiendo interactuar con sus seguidores y amigos. Instagram tiene más de 500 millones de usuarios activos por mes. Sin embargo, una problemática no menor es que existen muchos usuarios que crean cuentas falsas, es decir, son usuarios *Fake*. Generalmente, este tipo de cuentas son creadas con fines maliciosos, como realizar spam, espiar o subir contenido inapropiado a Instagram, desagradable para los demás usuarios de la red social.

Usualmente se diseñan algoritmos de detección de perfiles falsos en base a sus características y comportamiento. El archivo **fake** contiene datos de cuentas de Instagram que el algoritmo ya ha clasificado como *fake* o real:

- **profilepic** Indica con un 1 si la cuenta tiene foto de perfil o con un 0 si no
  - **description** Indica la cantidad de caracteres usados en la descripción de la cuenta
  - **private** Indica si la cuenta es privada o pública
  - **posts** Indica la cantidad de publicaciones de la cuenta
  - **followers** Indica la cantidad de seguidores de la cuenta
  - **follows** Indica la cantidad de cuentas seguidas
  - **Fake** Indica si la cuenta correspondía a una cuenta falsa o no
- a) Cargue los datos del archivo **fake** en R. ¿Qué función utilizó para cargar los datos? ¿En qué paquete viene incorporada esta función? ¿Por qué es necesario cargar paquetes? Comente.
- b) ¿En qué formato se lee la variable *followers*? ¿Por qué se ha leído en este formato? Realice la modificación correspondiente de manera que el formato sea el adecuado. Verifique que se realizó correctamente la modificación.
- c) La variable *profilepic* es una variable categórica que indica si la cuenta tiene o no tiene foto de perfil. Sin embargo, está codificada de forma numérica (unos y ceros). ¿Por qué pudiera ser relevante detectar aquellas categorías codificadas numéricamente previo al análisis? ¿Qué errores podrían cometerse si no realizó este alcance? Comente. Recodifique esta variable en las categorías que corresponde. ¿Cuántos usuarios tenían foto de perfil? ¿Cuántos no?
- d) Se cree que en general, las cuentas *fake* no suelen tener foto de perfil. Muestre en una tabla cómo se distribuye el tener o no tener foto de perfil en las cuentas *fake* y *reales*. Comente. ¿Le parece que tener foto de perfil asegura que una cuenta sea real?

# Evaluación 2

Natalie Julian

- e) El cuociente de seguidores entre seguidos de una cuenta es:

$$\text{cuociente} = \frac{\text{followers}}{\text{follows}}$$

Este cuociente se utiliza como una variable relevante para clasificar si la cuenta es fake, menores valores del cuociente indicarían que una cuenta es *fake*. Calcule la variable *cuociente* y agréguela a los datos. ¿Qué rango de valores recorre dicho cuociente para las cuentas fake? ¿y para las cuentas reales? Compare. ¿Por qué tendría sentido realcionar menores valores del cuociente con cuentas fake? ¿Cómo se puede interpretar el cuociente en términos del contexto? Conjeture.

**BONUS)** Instale y cargue el paquete `dplyr`. Luego corra las siguientes líneas de código:

```
datos %>%  
  group_by(fake) %>%  
  summarise(Min=min(cuociente),  
            Max=max(cuociente))
```

¿Qué resultado observa? Explique con sus palabras qué cree que realiza cada línea de código para llegar al resultado obtenido.