



## PRUEBA ML ENGINEER

### Whale & Jaguar

**CLASIFICADOR DE NOTICIAS:** Descarga [aquí](#) el *dataset* de titulares de noticias (y sus descripciones cortas) publicadas entre 2012 y 2018 en el HuffPost y resuelve los siguientes retos. Puedes utilizar librerías y modelos pre-entrenados de acceso público (referenciar).

#### Retos:

1. ¿Se pueden catalogar las noticias con la descripción y los titulares? Compara tu clasificación con las categorías incluidas en el set de datos.
2. ¿Existen estilos de escritura asociados a cada categoría?
3. ¿Qué se puede decir de los autores a partir de los datos?
4. Ahora, utilizando técnicas de aprendizaje no supervisado, trata de identificar temas, "protagonistas" u otras entidades de las noticias.
5. Basándote en el texto de la descripción corta, caracteriza este *dataset*.
6. ¿Qué otra información útil se puede extraer de los datos?

#### Entregables:

1. Link de acceso al proyecto en GitHub (u otro repositorio)
2. Presentación (diapositivas) que comunique de la mejor manera posible y **para cualquier público** los resultados encontrados.