

## **ROBIN ROVER - PRUEBA TÉCNICA DATA SCIENCE**

El siguiente ejercicio tiene como objetivo una categorización de productos en base a un set de productos entrenados.

El set de productos entrenados contiene la información de la categoría asociada (tanto su código como su descripción) a cada descripción de producto, así como su marca y un id de la tienda donde se ha adquirido.

El ejercicio consta de varias partes.

### **Parte 1 – Preparación de la información.**

Dispones de un csv (productAttributes.csv) que contiene información de productos previamente categorizados y revisados por el experto. Estos serán tus datos de entrenamiento.

Deberás leer esta información y llevártela a un dataframe en python. Emplea las librerías / soluciones que estimes convenientes, total libertad.

Una vez obtenido este dataframe, verás que existe un campo llamado marketplace, este marketplace contiene un \_id de marketplace. Para poder obtener toda la información relacionada con el marketplace, podrás acceder a una API que hemos diseñado para este propósito. Puedes encontrar la documentación de su uso en:

<https://stores.robinrover.io/swagger/>.

Como en el caso anterior tienes total libertad para emplear la librería de cliente http que prefieras.

Una vez obtenida esta información dispondrás de un conjunto de datos que deberás almacenar en mongo. Para ello puedes emplear una base de datos local, pero es importante emplear el cliente de mongo para python.

La colección deberá llamarse testPythonDB, y tendrá el siguiente esquema:

- \_id: id unívoco del objeto
- name: Descripción del producto
- categoryName: Descripción de la categoría
- categoryCode: código de la categoría
- brandName: Marca del producto
- marketplaceId: id de la tienda
- marketplaceName: Nombre de la tienda

Tienes total libertad para tipar las propiedades de la colección como estimes más conveniente / lógico, así como definir arrays/objetos si entiendes que puede ser útil de cara a posterior explotación de la información. Igualmente puedes definir propiedades adicionales si crees que pueden ser útiles de cara al paso posterior.

## Parte 2 - Categorización

Una vez la información almacenada en tu base de datos local, deberás desarrollar un modelo que permita establecer una clasificación determinista dada una descripción de producto. Es decir, en base a la información existente en la base de datos de entrenamiento, establecer el algoritmo que entiendes más completo de cara a poder clasificar automáticamente un producto nuevo.

Entendemos que leerás la colección creada en el paso anterior (con lo que emplearás el mismo cliente mongo que empleaste en el paso 1) y operarás con ella para poder obtener el resultado necesario, pero puedes utilizar cualquier algoritmo y cualquier librería que se te ocurra. Tienes total libertad para diseñar el modelo. No hay una solución correcta absoluta, con lo que cualquier alternativa tendrá un determinado grado de confianza. Siéntete libre de investigar, experimentar.

No es preciso desarrollar ningún front vía Bottle o cualquier otra aplicación, aunque si te sientes más cómodo sin problema! Si prefieres simplemente usar un comando por SSH también es perfecto. Lo relevante es que la aplicación de permitir introducir una descripción y el algoritmo nos devuelva el top 5 de categorías más probables.

Opcionalmente también sería deseable aplicar un algoritmo para la determinación de la marca.

Se valorará muy positivamente que apliques algún tipo de arquitectura de organización y gestión de componentes o de código aplicado. Otro aspecto que es opcional pero valoraríamos muy positivamente es realizar testing y preparar el despliegue en local haciendo uso de las tecnologías docker/docker-compose.

El proyecto debe de estar en GitHub. A la hora de entregarlo tienes que darnos acceso al repositorio.

Estamos deseando ver tu solución. Da lo mejor de ti en esta prueba!