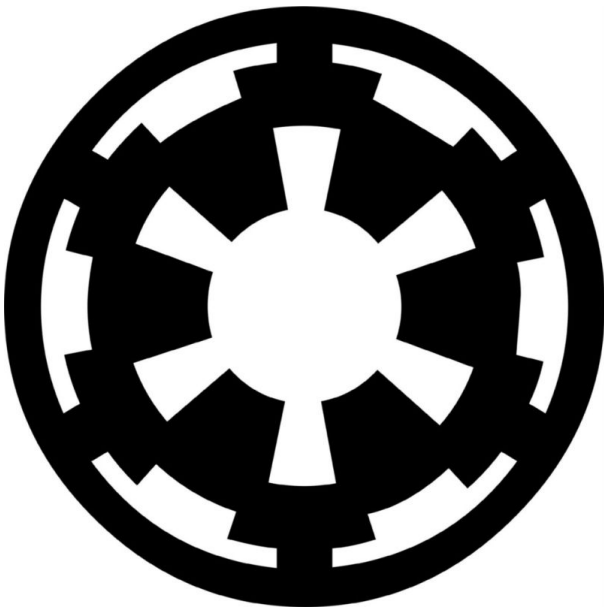


Seedtag Codetest 4: Data Scientist



===== comienzo de la transmisión =====

Agente Kallus,

si está leyendo este mensaje es que ha alcanzado el piso franco que El Imperio tiene instalado cerca de la base Rebelde en la ciudad de Norg Bral. Bien hecho, aunque su trabajo aquí no ha hecho más que empezar.

Como sabe, el planeta Mandalore es de una gran importancia estratégica para nosotros y tenemos que estar preparados ante cualquier movimiento de los comandos Rebeldes que se encuentran estacionados allí. Para anticiparnos a ellos, hemos estado desarrollando en los últimos meses un sistema de interceptación de comunicaciones que estamos cerca de poner en marcha y que se encuentra instalado en su localización. El último módulo que queda por desarrollar es el subsistema de clasificación de transmisiones. Aquí es donde entra usted.

Detalles de la misión

Nuestro departamento de inteligencia ha clasificado todas las comunicaciones interceptadas durante las últimas semanas en 7 categorías distintas, atendiendo a las divisiones Rebeldes a las que pertenecen. Su misión es conseguir que el sistema de clasificación sea capaz de clasificar una nueva transmisión de manera automática para que el departamento de operaciones pueda evaluar las amenazas más rápido.

El resultado de esta misión deben ser 2 ejecutables: * `train`: Recibirá como único argumento el nombre de una carpeta de entrada. Esta carpeta contendrá una subcarpeta por cada categoría,

dentro de las cuales estarán los ejemplos. A partir de esa carpeta, `train` entrenará un modelo y generará un archivo "model" con el modelo entrenado. Ejemplo de uso:

```
$ ./train dataset
```

- `classify`: Recibirá un número variable de argumentos, el primero con el nombre del archivo del modelo que se cargará y todos los posteriores representando rutas a archivos a clasificar. Deberá imprimir por pantalla una línea por cada archivo recibido con la forma: `<nombre_del_archivo><1 espacio><categoría>` Ejemplo de uso:

```
$ ./classify model archivo1 archivo2 archivo3
archivo1 categoríal
archivo2 categoría2
archivo3 categoríal
```

Finalmente, se deberá adjuntar un archivo llamado 'ANALYSIS' (del tipo de archivo que se desee), que incluya los pasos seguidos hasta llegar a la solución final, además de un estudio del funcionamiento del clasificador obtenido, desglosando los detalles y las métricas que considere más interesantes.

Como comprenderá, es fundamental que el sistema generalice bien a comunicaciones no vistas anteriormente.

Instrucciones de entrega

Deberá generar un fichero zip con el nombre `<username-en-github>_codetest4_seedtag.zip` que contenga los dos ejecutables, el informe de rendimiento y un fichero README indicando las instrucciones para poder ejecutar el código y enviarlo adjunto en un email a alianza@seedtag.com. En caso de que su sistema de correo no le permita enviar el fichero, puede renombrar la extensión a `.a`.

Suerte en su misión y que la fuerza le acompañe.

===== fin de la transmisión =====