Proyecto Machine Learning

**Primeros Pasos**

Lo primero que necesitamos es buscar los datos que vamos a utilizar, para este proyecto necesitamos una base de datos de myanimelist.com el cual tiene como datos los títulos el score el ranking entre otros y otra base de datos de puras imágenes para la otra parte del proyecto.

**Limpieza**

Ahora vamos a realizar una limpieza primero vamos a eliminar las columnas que no necesitamos, y chequeamos que no tengamos nulos, o ‘Unknown’ que no nos sirven para realizar comparativas de correlación ni los podemos ubicar en los gráficos, adicionalmente la columna aired que es la que contiene la fecha de lanzamiento está en formato str lo cual utilizando una función creada por mi nos genera una columna con solo los años.

**Predicción de Score**

Este modelo de Machine Learning, busca predecir el Score que tendrá un anime que esta por salir, metiéndonos en el papel de Data Scientist de un estudio de animación nos piden realizar este modelo para saber si tendría sentido sacar otra temporada.

Para este modelo primero comenzaré con el dataset de MAL sin los usuarios ya que en el mismo están los puntajes dados por ellos sin los datos personales que no nos importarían mucho

Lo primero es importar las librerías seguido por hacer una limpieza como se explica anteriormente, luego realizamos una visualización para ver como están distribuidos los datos.

Ahora vamos con la parte de ML lo primero seria hacer un Split de los datos para poder tener unos datos para comparar al realizar las predicciones.

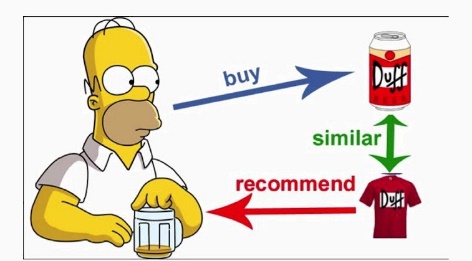
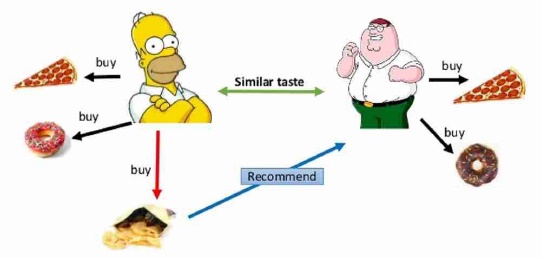
En este caso en particular vamos a utilizar un modulo llamado Pycaret el cual nos ayuda haciendo una comprobación con varios modelos de regresión.

Ahora sabiendo que el mejor modelo es un random forest procedemos a realizar las predicciones y determinar las métricas.

**Sistemas de Recomendación**

Como en todo modelo de ML lo primero que debemos realizar es una limpieza, una vez terminada esta parte se procede a hacer el modelo.

Al investigar me di cuenta que el sistema de recomendación consta de dos partes una que es recomendación por producto y la segunda parte es de recomendación por gustos similares

En este caso pude realizar el primero con la función cosine\_similarity y la segunda parte es la que utiliza los modelos de ML pero no me ha dado tiempo para terminarlo ya que prefiero hacer otro proyecto.

**Generador de Personajes con GAN (Redes Generativas Antagónicas)**

Para este proyecto no vamos a necesitar hacer como tal una limpieza sino un reshape de las imágenes para que todas las imágenes tengan el mismo tamaño y así poder trabar mejor.

Una vez que tengamos las imágenes del mismo tamaño, tendríamos que convertirlas en array o tensores ya que es la única forma que podamos tratarlas.

Como vamos a utilizar una red GAN, lo primero es entender como funciona, toda red GAN consta de un generador y un discriminador, lo que podríamos ver como un falsificador y un policía, de los cuales el falsificador debe tratar de realizar una imagen que el policía no pueda determinar si es verdadero o falso, por el resto es una red neuronal normal y corriente.

En este caso utilizamos un modelo sequential tanto para el generador y el discriminador de 4 capas conv2d con leakyrelu.

Utilizamos un optimizador Adam con un los del tipo 'binary\_crossentropy', también necesitamos hacer una normalización de los datos para que estén entre [-1 y 1] que don los valores que permite el leakyrelu, también vamos a crear un par de funciones para ir guardando tanto los modelos como una muestra cada 10 iteraciones, como parámetros necesarios tenemos el latent\_dim que es prácticamente la cantidad de compresión a la que vamos a someter las imágenes, también la cantidad de epochs y el numero de batch, que son los mismo que las redes neuronales iguales.

**Conclusiones**

1.- Para el primer modelo podemos concluir que es un buen modelo ya que podemos predecir cual seria el score que tendrá una serie antes de su estreno lo cual podría ayudar a la empresa a decidir cuanto podría invertir en la publicidad

2.- Para el segundo modelo al no terminarlo no podemos ser muy objetivos, pero viendo la primera parte nos indica que vamos por buen camino, con lo cual podemos concluir que sería un muy buen proyecto y debemos dedicar más esfuerzo.

3.- El tercer modelo es con el cual podemos presumir ya que si lo utilizamos correctamente, podemos utilizarlo para tener una idea de cómo podría ser el protagonista del próximo éxito.

…………………………………………………………………………………………….

Pude realizar varios proyectos ya que utilice mi EDA como comienzo lo cual me ahorro mucho tiempo.

………………………………………………………………………………………………….

Al igual que en mi EDA voy a realizar un requirements.txt, el cual contine los módulos con la versión exacta que utilice para que no allá problemas al tratar de correr los programas, se coloca en un requirements.txt porque al hacer un

\*\*pip install -r requirements.txt\*\*

Y que se nos haga más fácil

…………………………………………………………………………………………………..