

PROYECTO RECOMENDACIÓN DE ANIMES

INTEGRANTES:

EMIRO MORENO SOTO

JENNIFER DUQUE ORTIZ

DANIELA JIMÉNEZ GÓMEZ

CURSO:

INTRODUCCIÓN A LA INTELIGENCIA ARTIFICIAL

UNIVERSIDAD DE ANTIOQUIA

FACULTAD DE INGENIERÍA

23 DE ABRIL DEL 2023

- Dadas la necesidad de una empresa de streaming especializada en anime en saber las características de los más detectados y los mejores puntuados por los usuarios, se va a predecir aquellos factores que hacen relevante que el dicho anime tenga éxito.
- Usando el dataset de kaggle esta competición <https://www.kaggle.com/datasets/hernan4444/anime-recommendation-database-2020> Este conjunto de datos contiene información sobre 17.562 anime y la preferencia de 325.772 usuarios diferentes. En particular, este conjunto de datos contienen, calificaciones dadas por los usuarios a los animes que han visto por completo entre otros datos.
- Como métrica de Machine Learning se utilizará el método de regresión Gretl, el cual se podrá usar gracias a unas algunas librerías que importadas para esto, como hipótesis se emplearán 9 variables para saber que tan acogido será el anime y en base a esto elegir si lanzarlo o no. Como métrica de negocio se podría verificar el incremento en usuario, si durante un mes se ve un aumento del 8% se dejará el modelo y se le harán mejoras.
- Si los anime que han sido seleccionado con la ayuda de este modelo predictivo no tienen gran acogida después de un semestre se dejarán de usar.

Para este avance de proyecto sobre “Recomendaciones de animes” donde se va a predecir factores que hacen relevante que dicho anime tenga éxito, se toma como punto de inicio el conjunto de datos que contiene la información oportuna para implementar el adecuado desarrollo del mismo.

Esta información contiene la preferencia de diferentes usuarios que alimentan la base de datos a utilizar; los cuales muestran los listados por usuario, con su estado de descartado, planeado para ver, los que ven actualmente y aquellos que tienen en espera, de esta forma también contiene las características como género, estudio, estadísticas y calificaciones que dan los usuarios de 1 a 10 y si lo han visto completamente. Estas bases ayudan a separar las particularidades de cada anime como una información HTML, en estos archivos se encuentra además reseñas, sinopsis, información sobre el personal, sus estadísticas de anime entre otras características.

Las siguientes imágenes evidencian el paso a paso de lo realizado al momento:

1. Se importan las librerías necesarias para que se ejecuten las líneas correctamente.
2. Se da el permiso al programa para que acceda al archivo alojado en el drive.

3. Se carga el dataset seleccionado (contiene 17562 datos).
4. Se define las columnas que se desean mostrar, debido a que es un conjunto de datos amplio, por lo tanto, es necesario limitarlo
5. Se define que sólo se desea mostrar 35 columnas del data set.
6. Se verifican los encabezados antes mencionados.
7. Se realiza una limpieza de los datos para evitar errores en la ejecución.
8. Se filtran las variables que se requieren para el análisis.

```

PROYECTO DEL CURSO INTELIGENCIA ARTIFICIAL

Haz doble clic (o pulsa Intro) para editar

[ ] #Importamos las librerías que necesitamos
import math
import matplotlib.pyplot as plt
import numpy as np
import pandas as pd
import seaborn as sns
import sklearn

from sklearn.preprocessing import scale
from sklearn.model_selection import train_test_split
from sklearn.metrics import mean_squared_error
from sklearn.metrics import r2_score

```

```

from google.colab import drive
drive.mount('/content/drive')

#Damos permiso para usar archivos de nuestro Drive

Drive already mounted at /content/drive; to attempt to forcibly remount, call drive.mount("/content/drive", force_remount=True).

[ ] df=pd.read_csv('/content/drive/MyDrive/anime.csv')
df.shape
#Cargamos el data set a usar y vemos sus dimensiones

(17562, 35)

[ ] pd.set_option('display.max_columns',35)
#definimos las columnas para mostra por pantalla

[ ] df.head()
#vemos los encabezados antes definidos

```

	MAL_ID	Name	Score	Genres	English name	Japanese name	Type	Episodes	Aired	Premiered	Producers	Licensors	Studios	Source	Duration
0	1	Cowboy Bebop	8.78	Action, Adventure, Comedy, Drama, Sci-Fi, Space	Cowboy Bebop	カウボーイビバップ	TV	26	Apr 3, 1998 to Apr 24, 1999	Spring 1998	Bandai Visual	Funimation, Bandai Entertainment	Sunrise	Original	24 min. per ep.
1	5	Cowboy Bebop: Tengoku no Tobira	8.39	Action, Drama, Mystery, Sci-Fi, Space	Cowboy Bebop:The Movie	カウボーイビバップ 天国の扉	Movie	1	Sep 1, 2001	Unknown	Sunrise, Bandai Visual	Sony Pictures Entertainment	Bones	Original	1 hr. 55 min.
2	6	Trigun	8.24	Action, Sci-Fi, Adventure, Comedy, Drama, Shounen	Trigun	トライガン	TV	26	Apr 1, 1998 to Sep 30, 1998	Spring 1998	Victor Entertainment	Funimation, Geneon Entertainment USA	Madhouse	Manga	24 min. per ep.
3	7	Witch Hunter ROBIN	7.97	Action, Mystery, Police	Witch Hunter ROBIN	ウィッチハンターロビン	TV	26	Jul 2, 2002 to	Summer	TV Tokyo, Bandai Visual	Funimation, Bandai Entertainment	Sunrise	Original	25 min.

```
df = df.replace('Unknown', float(0))
#Hacemos una limpieza de datos reemplazndo string por numeor flotante cero para evitar error con el algoritmo

df.head(2)
```

	MAL_ID	Name	Score	Genres	English name	Japanese name	Type	Episodes	Aired	Premiered	Producers	Licensors	Studios	Source	Duration	Rating
0	1	Cowboy Bebop	8.78	Action, Adventure, Comedy, Drama, Sci-Fi, Space	Cowboy Bebop	カウボーイビバップ	TV	26	Apr 3, 1998 to Apr 24, 1999	Spring 1998	Bandai Visual	Funimation, Bandai Entertainment	Sunrise	Original	24 min. per ep.	R - 17+ (violence & profanity)
1	5	Cowboy Bebop: Tengoku no Tobira	8.39	Action, Drama, Mystery, Sci-Fi, Space	Cowboy Bebop: The Movie	カウボーイビバップ 天国の扉	Movie	1	Sep 1, 2001	0.0	Sunrise, Bandai Visual	Sony Pictures Entertainment	Bones	Original	1 hr. 55 min.	R - 17+ (violence & profanity)

```
[ ] df=df.drop(['MAL_ID', 'Name', 'Genres', 'English name', 'Japanese name', 'Type', 'Aired', 'Premiered', 'Producers', 'Licensors', 'Studios', 'Source', 'Duration', 'Rating', 'Ranked', 'Score-10', 'Score-9', 'Score-8', 'Score-7', 'Score-6', 'Score-5', 'Score-4', 'Score-3', 'Score-2', 'Score-1'],axis=1)

#filtramos solo las variables a usar

[ ] df.head()
```

	Score	Episodes	Popularity	Members	Favorites	Watching	Completed	On-Hold	Dropped	Plan to Watch
0	8.78	26	39	1251960	61971	105808	718161	71513	26678	329800
1	8.39	1	518	273145	1174	4143	208333	1935	770	57964
2	8.24	26	201	558913	12944	29113	343492	25465	13925	146918
3	7.27	26	1467	94683	587	4300	46165	5121	5378	33719
4	6.98	52	4369	13224	18	642	7314	766	1108	3394