



SCI: Inteligencia artificial – Proyecto 1

Análisis de Películas y Series de TV en IMDB

Introducción

El objetivo de este proyecto es realizar un análisis exploratorio de datos y modelado matemático para entender la relación entre diferentes características de películas y series de TV, como la calificación en IMDB, la recaudación bruta y el número de votos. Los estudiantes aplicarán conceptos de álgebra lineal, cálculo, y análisis de datos utilizando Numpy, pandas, y matplotlib.

Dataset

Utilizaremos el **IMDB Dataset of Top 1000 Movies and TV Shows** de [Kaggle](https://www.kaggle.com/datasets/lk1989/imdb-top-1000-movies-and-tv-shows), que contiene información sobre películas y series, incluyendo:

- Título
- Año de lanzamiento
- Certificación
- Duración
- Género
- Calificación en IMDB
- Resumen
- Puntuación en Metacritic
- Director
- Actores principales
- Número de votos
- Recaudación bruta
- Conformación de Equipos

Conformación de equipos

El proyecto podrá desarrollarse de manera personal o en parejas. Los equipos de trabajo deberán ser confirmados antes del **11 de julio**

Entregables

1. **Presentación (4 puntos)**
 - Explicar el análisis de datos exploratorio, limpieza de datos y los resultados obtenidos.



- Explicar la implementación del modelo matemático
 - Duración ideal de 10 minutos (máximo 15 minutos).
2. **Código (6 puntos)**
- Incluir el código utilizado para el análisis y el modelado.
 - El código debe ser legible y reproducible, entregado en un único notebook o archivo comprimido en formato zip.
3. **Informe (10 puntos)**
- Describir el proceso de limpieza y análisis de datos.
 - Incluir el proceso de modelo analíticamente y `code snippets` que representan el mismo proceso

Actividades Sugeridas

1. **Carga y limpieza de datos:**
 - Utilizar pandas para cargar el dataset y realizar la limpieza básica de datos (manejo de valores nulos, tipos de datos).
2. **Análisis exploratorio de datos (EDA):**
 - Calcular estadísticas descriptivas básicas.
 - Generar gráficos con matplotlib para visualizar la relación entre las ventas de helados y la temperatura promedio.
3. **Modelado matemático:**
 - Aplicar álgebra lineal para encontrar la relación lineal entre las (regresión lineal simple).
 - Implementar el modelo de regresión lineal desde cero usando Numpy.
 - Calcular la derivada de la función de costo y realizar un descenso de gradiente simple para ajustar los parámetros del modelo.

El análisis de datos exploratorio de los estudiantes deberá responder a al menos 4 de las siguientes preguntas.

1. **Análisis de la recaudación bruta de una película vs. directores.**
 - ¿Qué directores tienden a generar mayores ingresos en taquilla?
2. **Análisis de la recaudación bruta de una película vs. diferentes actores.**
 - ¿Qué actores están asociados con películas que generan mayores ingresos?



3. **Análisis del número de votos de una película vs. directores.**
 - ¿Qué directores tienden a recibir más votos en IMDB?
4. **Análisis del número de votos de una película vs. diferentes actores.**
 - ¿Qué actores están asociados con películas que reciben más votos?
5. **¿Qué géneros de películas son los que los actores prefieren?**
 - Analizar la preferencia de actores por diferentes géneros cinematográficos.
6. **¿Qué combinación de actores (Star1, Star2, Star3 y Star4) está obteniendo buenas calificaciones en IMDB la mayor parte del tiempo?**
 - Identificar combinaciones de actores que tienden a recibir buenas calificaciones en IMDB.
7. **¿Qué combinación de actores (Star1, Star2, Star3 y Star4) está obteniendo buenos ingresos en taquilla?**
 - Identificar combinaciones de actores que tienden a generar mayores ingresos en taquilla.

Fechas Importantes

- **Inicio del proyecto:** 8 de julio
- **Entrega del código y el informe:** 22 de julio
- **Presentación final:** 22 de julio