**Reporte Final: Análisis de Predicción de Disponibilidad en Netflix**

En una primera etapa este proyecto tiene como objetivo realizar un análisis exploratorio de un conjunto de datos que contiene información sobre programas de televisión disponibles en diversas plataformas de streaming, como Netflix, Hulu, Prime Video y Disney+. A través del análisis de las características de estos programas, se pretende obtener insights sobre las tendencias en las plataformas de streaming, la distribución de calificaciones y las edades recomendadas para los títulos de televisión.

#### Hipótesis 1:

**Hipótesis:** Las plataformas de streaming como Netflix, Hulu, Prime Video y Disney+ tienen una distribución diferente de géneros y calificaciones de los programas.

**Respuesta:**  
La visualización y análisis de las distribuciones de calificaciones en las plataformas muestra que efectivamente cada plataforma tiene características distintas. Por ejemplo, en el gráfico de distribución de calificaciones (IMDb y Rotten Tomatoes), se puede observar que algunas plataformas tienen calificaciones más altas que otras. Esto sugiere que las plataformas tienen diferentes enfoques o géneros de contenido que atraen a distintos tipos de audiencias. Netflix podría tener una mayor diversidad de géneros, mientras que Disney+ podría estar más centrado en contenido familiar, lo cual afecta sus calificaciones.ç

#### Hipótesis 2:

**Hipótesis:** Los programas más recientes tienen una edad recomendada (Age) más baja, indicando que las plataformas de streaming están orientadas a un público más joven.

**Respuesta:**  
En el análisis de dispersión entre el año de lanzamiento y la edad recomendada, se observa que muchos programas recientes tienen una edad recomendada más baja, lo que indica que las plataformas de streaming están orientadas a audiencias más jóvenes, especialmente en el caso de plataformas como Netflix y Disney+. Este hallazgo refuerza la idea de que las plataformas adaptan su contenido a las tendencias actuales y las preferencias de un público joven y dinámico.

#### Hipótesis 3:

**Hipótesis:** Las plataformas de streaming con una mayor cantidad de títulos (como Netflix) tienen una mayor diversidad de géneros y calificaciones, mientras que plataformas más pequeñas tienen una oferta más limitada pero posiblemente más curada o específica.

**Respuesta:**  
El gráfico de barras que muestra la cantidad de títulos por plataforma confirma esta hipótesis. Netflix tiene la mayor cantidad de títulos, lo que implica una mayor diversidad de géneros y tipos de contenido. En cambio, plataformas como Disney+ y Prime Video, con una menor cantidad de títulos, podrían estar ofreciendo contenido más específico o curado, lo que puede reflejarse en un enfoque más limitado de géneros. La variedad de calificaciones también sugiere que plataformas más grandes, como Netflix, tienen un rango más amplio de contenido y, por lo tanto, una mayor variabilidad en las calificaciones.

Y en una segunda etapa el objetivo principal de este proyecto es construir un modelo predictivo que sea capaz de clasificar correctamente si un programa de televisión está disponible en **Netflix** o no. Este tipo de modelo puede ser útil para usuarios de plataformas de streaming que desean saber si un determinado programa está disponible en su servicio preferido, en este caso, **Netflix**.

Específicamente, el modelo se entrena para predecir la variable **Availability** (disponibilidad) utilizando características como **edad del programa** (Age) que es representada numéricamente para ser utilizadas por el algoritmo de Random Forest.

**Entrenamiento del Modelo:**

* Se utilizó un clasificador Random Forest con 100 árboles para entrenar el modelo. Las características seleccionadas fueron la edad recomendada y el género del programa.

**Evaluación del Modelo:**

* **Precisión del modelo**: El modelo mostró una buena capacidad de predicción, alcanzando una alta precisión en el conjunto de prueba. Esto sugiere que el modelo es efectivo para predecir si un programa de televisión está disponible en Netflix basándose en las características seleccionadas.
* **Matriz de Confusión**: La matriz de confusión reveló que el modelo hace una distinción clara entre los programas disponibles y no disponibles en Netflix.
* **Reporte de Clasificación**: El reporte mostró métricas como precisión, recall y F1-score, todas indicando un buen rendimiento en la predicción de la disponibilidad de programas en Netflix.
* **Conclusión General:**
  + El modelo de Random Forest ha mostrado un buen desempeño en la clasificación de la disponibilidad de programas en Netflix, lo que indica que las características seleccionadas (edad y género) son relevantes para este tipo de predicción. La precisión del modelo es un buen indicativo de que se puede confiar en él para tareas similares en el futuro.
  + A medida que se agreguen más características o se mejoren los datos, es probable que el modelo pueda ofrecer predicciones aún más precisas. Además, podría ser útil explorar otros modelos de machine learning para comparar y mejorar los resultados.

**Precisión del modelo:**

* El modelo alcanzó una precisión de aproximadamente **(66)** en el conjunto de prueba, lo que confirma su efectividad en este contexto.