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

**1. Introducción**

**El objetivo del proyecto fue predecir la disponibilidad de un título en Netflix utilizando un conjunto de datos de TV shows y películas, aplicando un enfoque de Machine Learning. El análisis se dividió en las siguientes etapas:**

* **Limpieza y exploración de datos.**
* **Selección de características para reducir la dimensionalidad.**
* **Entrenamiento de un modelo de clasificación.**
* **Evaluación del modelo y obtención de métricas.**
* **Generación de conclusiones.**

**2. Procesamiento de Datos**

* **Los valores nulos en la columna Age fueron tratados, asignándoles la categoría "Unknown".**
* **Las variables categóricas (como Age) se mapearon a valores numéricos para su uso en el modelo.**
* **Las características consideradas fueron:**
  + **Año (Year)**
  + **Clasificación por edad (Age)**
  + **Disponibilidad en otras plataformas (Hulu, Prime Video, Disney+).**

**3. Selección de Características**

**Para reducir la dimensionalidad y mantener solo las variables más relevantes, se aplicó el método SelectKBest con la función f\_classif. De las cinco características iniciales, las tres seleccionadas fueron:**

* **Año (Year).**
* **Clasificación por edad (Age).**
* **Disponibilidad**

**4. Modelo de Clasificación**

**Se utilizó un modelo de Random Forest Classifier para entrenar y realizar predicciones. Los datos se dividieron en un 70% para entrenamiento y 30% para pruebas.**

**5. Evaluación del Modelo**

**Las métricas de evaluación del modelo son las siguientes:**

* **Precisión (Accuracy): 0.95 (95%).**
* **Matriz de Confusión:**

**lua**

**Copiar código**

**[[85, 5],**

**[ 3, 57]]**

* + **Verdaderos Positivos: 57**
  + **Verdaderos Negativos: 85**
  + **Falsos Positivos: 5**
  + **Falsos Negativos: 3**
* **Reporte de Clasificación:**

**markdown**

**Copiar código**

**precision recall f1-score support**

**0 0.97 0.94 0.96 90**

**1 0.92 0.95 0.93 60**

**6. Conclusiones**

* **El modelo mostró un desempeño sólido con una precisión del 95%, indicando una alta capacidad para predecir correctamente la disponibilidad de títulos en Netflix.**
* **La selección de características demostró que las variables Year, Age y Disney+ son las más relevantes para el modelo.**
* **Los falsos positivos y negativos fueron bajos, lo que indica que el modelo tiene buena generalización.**