Jonh Stiven Arboleda Gallego A00359103

Isabella Ortiz Jaramillo A00358451

1. **IDENTIFICACIÓN DEL PROBLEMA**

Netflix cuenta con 200 millones de usuario activos aproximadamente, siendo una de las principales compañías de streaming por suscripción y también una de las pioneras en este campo. Como toda compañía Netflix cuenta con ciertas falencias en la precisión a la hora de sugerir películas según nuestro historial, así que se decidió realizar una mejor clasificación de las películas teniendo en cuenta más factores que el género de la película como normalmente lo realiza la plataforma, factores como: la fecha de lanzamiento, el director de la película, el país de origen y el reparto. Las sugerencias con factores diferentes y más específicos se hacen con el fin de que los usuarios reciban recomendaciones más acordes a sus gustos, a lo que frecuentan en sus películas y a lo que buscan ver en el momento.

La base de datos manejada fue elegida de kaggle.com, siendo de libre uso.

1. **RESEARCH**

**REQUERIMIENTOS FUNCIONALES**

1. Realizar tablas con la información de películas obtenidas del dataset de películas en la plataforma de Netflix.
2. Realizar 5 gráficos representando información importante para el estudio de datos.
3. Un gráfico ilustrando la cantidad de películas por años que existen en la plataforma
4. Un gráfico ilustrando la cantidad de películas por algunos países
5. Un gráfico que muestra algunos de los directores más conocidos en la plataforma
6. Un gráfico que ilustre las películas por el género.
7. Un gráfico ilustrando la duración de las películas por intervalos.
8. Filtrar la información de acuerdo al criterio escogido por el usuario.
   1. Primer criterio: Se filtran las columnas establecidas internamente con un valor de cadena, como lo son: Id de la película, título de la película, fecha de publicación, nombre del director, elenco de la película, país de origen.
   2. Segundo criterio: Se filtran las columnas establecidas internamente con un valor numérico, como lo son: Año de públicas en Netflix, duración en minutos de la película. Estos valores se filtran por medio de intervalos.
   3. Tercer criterio: Se filtran las columnas establecidas internamente con un valor categórico, como lo es: la calificación de la película.
9. Listar los registros y filtrarlos por valores particulares de una columna.
10. Sugerir las películas más adecuadas al usuario de acuerdo a su historial, basándonos en el entrenamiento de un árbol de decisión de implementación propia.
11. Usar el atributo género para el entrenamiento del árbol de decisión, así se hará la clasificación sobre los valores que puede obtener este atributo en nuestro dataset.

**REQUERIMIENTOS NO FUNCIONALES**

* 1. Implementación propia de un árbol de decisión para las sugerencias al usuario.[[1]](#footnote-1)
  2. Implementación de un árbol de decisión en base a librerías de C#

Para la solución de este problema específicamente es necesario conocer acerca de diferentes temas, algunas de estos temas pueden ser:

**Árboles de decisión**

Diagram

Description automatically generatedUn árbol de decisiones es una representación esquemática que facilita la toma de decisiones, especialmente cuando existen riesgos, costos, beneficios y múltiples opciones al representar visualmente las diferentes posibilidades que existen ante un escenario; además, de las posibles consecuencias que cada escenario podría traer. Su nombre se da debido al parecido que tiene el esquema con las ramas de un árbol; un árbol de decisiones puede utilizarse en cualquier aspecto de la vida cotidiana, desde decisiones difíciles en la familia, hasta aplicaciones complejas en los negocios y en la inteligencia artificial.[[2]](#footnote-2) [[3]](#footnote-3)

Además, estos arboles deben de cumplir unas reglas[[4]](#footnote-4) específicas para su implementación, las cuales son:

* 1. Al comienzo del juego se da un nodo inicial que no es apuntado por ninguna flecha, es el único del juego con esta característica.
  2. El resto de los nodos del juego son apuntados por una única flecha.
  3. De esto se deduce que hay un único camino para llegar del nodo inicial a cada uno de los nodos del juego. No hay varias formas de llegar a la misma solución final, las decisiones son excluyentes.

**Machine Learning[[5]](#footnote-5)**

Es una disciplina científica del ámbito de la Inteligencia Artificial que crea sistemas que aprenden automáticamente. Aprender en este contexto quiere decir identificar patrones complejos en millones de datos. **La máquina que realmente aprende es un algoritmo** que revisa los datos y es capaz de predecir comportamientos futuros. *Automáticamente*, también en este contexto, implica que estos sistemas se mejoran de forma autónoma con el tiempo, sin intervención humana. Veamos cómo funciona. Esta herramienta es de uso común en plataformas de recomendación, plataformas como Netflix, Disney+, Spotify, Google, Microsoft, Amazon, etc.

1. **SOLUCIONES CREATIVAS**

Para la solución de este punto llevamos a cabo el método de **lluvia de ideas** [[6]](#footnote-6)la cual consiste en seguir 4 pasos para un desarrollo óptimo, dichos pasos son:

* **Suspender el juicio**. Eliminar toda crítica. Cuando brotan las ideas no se permite ningún comentario crítico. Se anotan todas las ideas.
* **Pensar libremente**. Es muy importante la libertad de emisión. Las ideas locas están bien. Las ideas imposibles o inimaginables están bien. De hecho, en cada sesión tendría que haber alguna idea suficientemente disparatada que provocara risa a todo el grupo.
* **La cantidad es importante**. Hace falta concentrarse en generar un gran número de ideas que posteriormente se puedan revisar. Cuanto mayor sea el número de ideas, más fácil es escoger entre ellas.
* **El efecto multiplicador**. Se busca la combinación de ideas y sus mejoras. Además de contribuir con las propias ideas, los participantes pueden sugerir mejoras de las ideas de los demás o conseguir una idea mejor a partir de otras dos. ¿Qué tiene de bueno la idea que han dicho? ¿Qué se puede hacer para mejorarla o para hacerla menos salvaje?

Dicho lo anterior, procedemos a las ideas.

* + 1. **Búsqueda manual:** El usuario guarde las películas que más le llamen la atención para verlas en un futuro, cuando desee ver las películas ya guardadas.
    2. **Guardar las películas vistas:** Guardar las películas ya vistas por el usuario en una lista, al completar cierta cantidad de películas vistas comenzar a recomendar de acuerdo a su género más visto.
    3. **Recomendar a través de sus vistas:** Por medio de árboles de decisión tomar decisiones de recomendaciones en base las películas que busca, teniendo en cuenta el género, los directores, etc.
    4. **Utilizar listas de películas predeterminadas:** Tener listas con diversas películas por género, para recomendarlas de acuerdo al género más visto.

1. https://www.kaggle.com/shivamb/netflix-shows [↑](#footnote-ref-1)
2. <https://profesionistas.org.mx/que-son-y-como-hacer-arboles-de-decisiones/> [↑](#footnote-ref-2)
3. <https://www.gestiondeoperaciones.net/procesos/arbol-de-decision/> [↑](#footnote-ref-3)
4. https://es.wikipedia.org/wiki/Árbol\_de\_decisión#Reglas [↑](#footnote-ref-4)
5. https://cleverdata.io/que-es-machine-learning-big-data/ [↑](#footnote-ref-5)
6. https://es.wikipedia.org/wiki/Lluvia\_de\_ideas#:~:text=La%20lluvia%20de%20ideas%2C%20también,originales%20en%20un%20ambiente%20relajado. [↑](#footnote-ref-6)