

UNIVERSIDAD TEGNOLÓGICA DE PANAMÁ

FACULTAD DE INGENIERIA EN SISTEMAS COMPUTACIONALES

LICENCIATURA EN INGENIERIA EN SISTEMA Y COMPUTACIÓN

PROBABILIDAD APLICADA A TECNOLOGÍA DE INFORMACIÓN Y COMUNICACIÓN

CINEMETRIX: MODELADO PREDICTIVO DE CALIFICACIÓNES CINEMATOGRÁFICAS CON MACHINE LEARNING

Estudiantes:

Aldahir Aguilar 8-1029-1115

Andrés Flores 8-1025-1254

Diego García 8-1034-95

Luis Torné 8-1032-1644

Rashell Vidal 8-1028-643

Profesor:

Juan Marcos Castillo, PhD

Salón: 1IL124

22 de junio del 2025

Índice

[**Introducción** 3](#_Toc204624794)

[**Descripción de la Base de Datos** 3](#_Toc204624795)

[**Motivos de Selección** 5](#_Toc204624796)

[**Línea de Tiempo** 6](#_Toc204624797)

[**🔹 Semana 1 – Selección del tema y planteamiento del problema** 6](#_Toc204624798)

[**🔹 Semana 2 – Búsqueda y elección de la base de datos** 6](#_Toc204624799)

[**🔹 Semana 3 – Análisis estadístico y visualización de datos** 6](#_Toc204624800)

[**🔹 Semana 4 – Elaboración del informe final y presentación** 6](#_Toc204624801)

# **Introducción**

En el cine, saber si una película será exitosa antes de estrenarla es una tarea difícil. Sin embargo, hoy en día tenemos acceso a una gran cantidad de datos sobre películas anteriores., como sus géneros, directores, actores, calificaciones y premios recibidos. Esta información puede ser utilizada para crear modelos que ayuden a predecir el posible éxito de una nueva producción.

Este proyecto tiene como objetivo construir un modelo que, basándose en el historial de calificaciones y reconocimientos de los directores y actores principales, así como en el género de la película, permita estimar si una nueva producción tiene altas probabilidades de éxito.

Para ello, se utilizará la base de datos **Cinemetrix**, que contiene información detallada sobre películas, incluyendo nombres de directores, actores, sus calificaciones previas y premios obtenidos.

Con este modelo se busca brindar una herramienta útil para productores e inversionistas, que les ayude a tomar mejores decisiones antes de desarrollar o financiar una nueva película.

# **Descripción de la Base de Datos**

La base de datos analizada contiene información relacionada con películas, incluyendo datos generales, artísticos y técnicos. Desde una perspectiva estadística, se identifican distintos tipos de variables. Entre las variables cualitativas nominales se encuentran los campos como *título*, *género*, *estudio*, *actor principal* y *director*, ya que representan categorías sin un orden inherente. Las variables como *clasificación por edad* se consideran cualitativas ordinales, debido a que implican un nivel jerárquico o secuencial (por ejemplo, clasificaciones como PG, R, etc.).

Por otro lado, variables como *año*, *mes* e *id* son cuantitativas discretas, ya que representan conteos o valores enteros no fraccionarios. En cambio, campos como *calificación IMDb de la película*, *calificación del actor* y *calificación del director* se clasifican como cuantitativas continuas, al estar representadas por números reales con decimales. Finalmente, los campos booleanos que indican si un actor o director fue nominado o ganador de un premio son considerados cualitativos nominales dicotómicos, ya que solo tienen dos posibles valores: verdadero o falso.

# **Motivos de Selección**

En el contexto de esta investigación, se ha optado por emplear una base de datos de calificaciones de películas como sujeto de análisis para la implementación de modelos predictivos fundamentados en algoritmos de aprendizaje automático con un enfoque probabilístico. Esta decisión se basa en diversas razones tanto técnicas como prácticas que refuerzan la relevancia y el potencial de la investigación.

En primer lugar, las bases de datos asociadas con películas (como IMDb, TMDB o Rotten Tomatoes) ofrecen una amplia accesibilidad al público y cuentan con una considerable variedad de características. Estos comprenden datos tales como el género, el año de lanzamiento, la duración, el elenco, los directores, el presupuesto y la fama, lo que facilita la creación de modelos predictivos multivariables con numerosas combinaciones posibles. Este tipo de información permite investigar relaciones complicadas entre diferentes factores y calcular las probabilidades ligadas a determinados resultados, como alcanzar una calificación elevada.

Desde una perspectiva metodológica, esta cuestión es adecuada para implementar modelos de aprendizaje supervisado, en los cuales la variable de interés (la calificación de la película) se encuentra ya establecida. Esto facilita el entrenamiento de algoritmos como la regresión logística, los árboles de decisión, los bosques aleatorios o las redes neuronales, que son capaces de estimar probabilidades con precisión. Asimismo, la adopción de un enfoque probabilístico posibilita la interpretación de los resultados en relación con el riesgo, la confianza y la probabilidad condicional, lo que favorece un análisis más exhaustivo de los patrones ocultos en los datos.

En el cine constituye un campo de interés diverso y cautivador, lo que contribuye a la comprensión y difusión de los resultados de la investigación entre públicos no especializados. Este aspecto también ayuda a mantener la motivación y la continuidad del proyecto, ya que se trata de un tema que está relacionado con el interés general.

# **Línea de Tiempo**

## **🔹 Semana 1 – Selección del tema y planteamiento del problema**

* Durante esta primera semana, como grupo nos reunimos para discutir posibles ideas y enfoques. Después de intercambiar opiniones, logramos definir el tema central del proyecto, el cual se basa en el análisis del éxito cinematográfico y elegimos un nombre “CineMetrics”. En este punto también identificamos el problema que queríamos abordar, así como el propósito del análisis, acordando que trabajaríamos con herramientas estadísticas y visualización de datos.

## **🔹 Semana 2 – Búsqueda y elección de la base de datos**

* En la segunda semana, nos enfocaremos en buscar una base de datos que se ajustara al tema seleccionado. Evaluaremos distintas opciones y fuentes hasta encontrar una solución más conveniente que cumpla con los requisitos del contenido, estructura y relevancia. Una vez elegida, la prepararemos para un análisis, haciendo una revisión general para entender su composición y campos principales.

## **🔹 Semana 3 – Análisis estadístico y visualización de datos**

* Durante esta semana nos dividiremos las tareas para avanzar con mayor eficiencia. Algunos comenzarán a trabajar en el planteamiento de los problemas estadísticos, mientras otros se encargarán de construir gráficas y representaciones visuales. También realizaremos algunos ajustes y limpieza de datos para asegurar resultados precisos. Este será un periodo clave donde empezaremos a obtener conclusiones preliminares.

## **🔹 Semana 4 – Elaboración del informe final y presentación**

* En la última semana, unificaremos todo el trabajo realizado en un reporte estructurado. Organizaremos los resultados obtenidos, incluiremos las gráficas con su respectiva interpretación y redactaremos nuestras conclusiones. Paralelamente, diseñaremos una presentación para exponer el proyecto, repartiéndonos los temas a presentar y practicando en equipo para asegurar una exposición clara y coherente.