

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

CINEMETRICS: 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

30 de julio del 2025

Introducción

[Introducción 3](#_Toc204622131)

[Justificación 4](#_Toc204622132)

[Antecedentes 5](#_Toc204622133)

[Definición de problema 6](#_Toc204622134)

[Avance de Estadísticos 7](#_Toc204622135)

[Grafica de visualización 8](#_Toc204622136)

## Introducción

Este proyecto tiene como objetivo desarrollar un modelo capaz de predecir el éxito de una película utilizando datos históricos sobre directores, actores, géneros y calificaciones. Para ello, se ha trabajado con el conjunto de datos **Cinemetrix**, que incluye información detallada como nombres, calificaciones y premios de los involucrados en cada producción.

En esta etapa del proyecto, se ha realizado un análisis exploratorio de los datos. Primero, se identificaron y clasificaron las variables del conjunto según su tipo: cualitativas o cuantitativas, y dentro de estas, si eran continuas o discretas. Esto permitió organizar mejor la información para los análisis posteriores.

También se calcularon estadísticas descriptivas, como promedios, medianas y desviaciones estándar, especialmente para variables numéricas como la calificación IMDb y las valoraciones asignadas a actores y directores. Estos resultados ofrecen una visión general del comportamiento de los datos.

Además, se elaboraron gráficos como diagramas de barras y de dispersión, los cuales ayudaron a identificar tendencias, por ejemplo, entre la clasificación por edad de las películas y sus calificaciones promedio. De igual forma, se observaron posibles correlaciones entre el género de las películas y su desempeño general.

Finalmente, se han comenzado pruebas con modelos predictivos básicos, lo que ha permitido explorar la viabilidad del modelo final y sentar las bases para las siguientes etapas del proyecto.

## Justificación

La elección de este tema surge del interés común del grupo por el cine y la posibilidad de aplicar conocimientos estadísticos a un contexto real y atractivo. Además, el análisis predictivo basado en datos cinematográficos permite ejercitar habilidades técnicas como la limpieza de datos, la interpretación de variables y el uso de modelos probabilísticos. Esta investigación no solo aporta valor académico, sino que también podrían servir como insumo para productoras, plataformas de streaming o incluso críticos, ayudándoles a identificar patrones de éxito o preferencias del público basados en evidencia cuantitativa.

## Antecedentes

El cine ha sido, desde sus orígenes, una de las expresiones artísticas más influyentes a nivel global, tanto por su capacidad de narración como por su impacto cultural y económico. A lo largo de las décadas, se han producido miles de películas que no solo entretienen, sino que también reflejan realidades sociales, estilos de dirección, evolución tecnológica y preferencias del público.

Con el avance de las tecnologías de la información y la disponibilidad de plataformas de datos como IMDb, se ha hecho cada vez más común el análisis cuantitativo y cualitativo del cine. Estos análisis permiten estudiar patrones de éxito, impacto de los actores y directores, evolución de géneros, y características que influyen en la recepción crítica y comercial de una película.

La industria cinematográfica ha sido objeto de numerosos estudios predictivos a través de bases de datos públicas como IMDb o TMDB. Estas bases han sido empleadas para analizar tendencias de consumo, determinar factores de éxito y hasta evaluar éxito de una nueva película en base a las altas calificaciones previas de estos datos. Proyectos previos han demostrado que variables como la popularidad del elenco, genero, mes de estreno y la expectativa hacia la siguiente obra de un determinado director o guionista tienen una fuerte relación con la recepción del público. En este contexto, el uso de modelos de machine learning permite llevar estos estudios a un nuevo nivel de precisión y aplicabilidad.

## Definición del problema

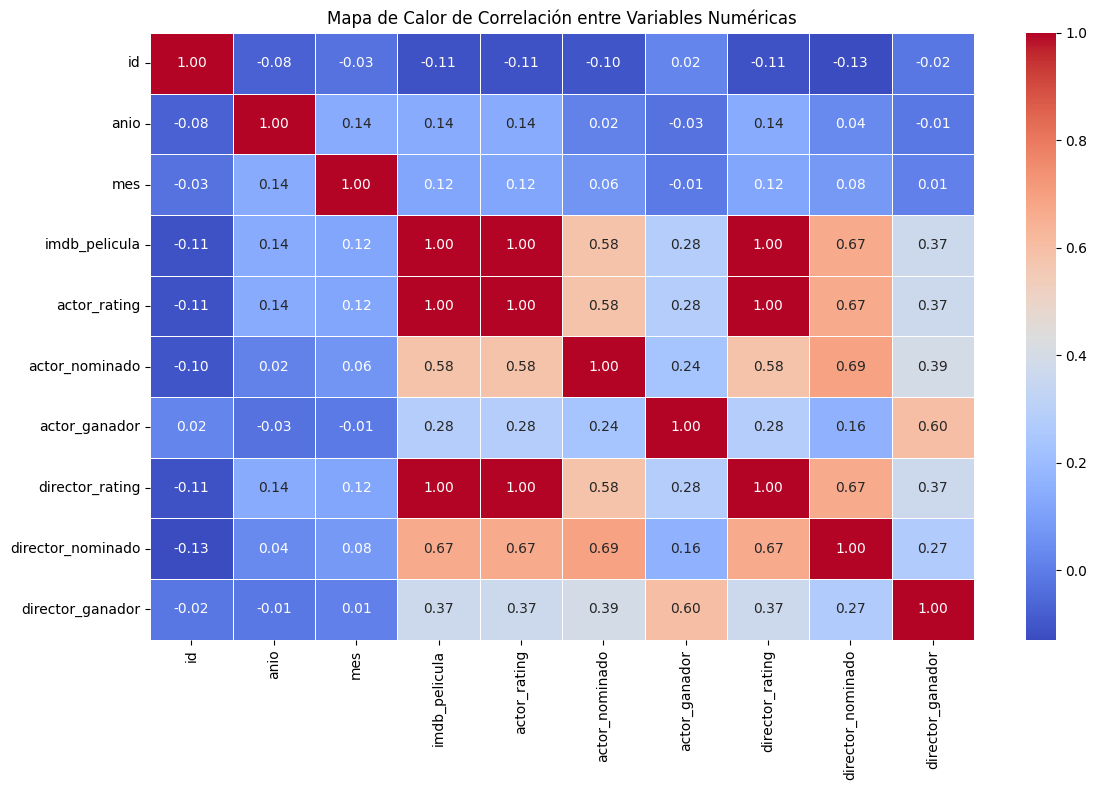
La industria cinematográfica, en constante evolución debido al auge del streaming y los cambios en los hábitos de consumo, enfrenta el desafío de predecir el éxito de sus películas, un factor clave para minimizar riesgos financieros y conectar con audiencias cada vez más diversas. Este proyecto busca desarrollar un modelo predictivo basado en el rating de IMDb, utilizando variables como el historial del director y elenco (premios, rating previo), género cinematográfico y clasificación por edad, con el objetivo de identificar patrones que determinen el éxito de una película y ofrecer una mejor comprensión aplicable tanto al ámbito académico como a la toma de decisiones en la industria del entretenimiento.

## Avance de Estadísticos

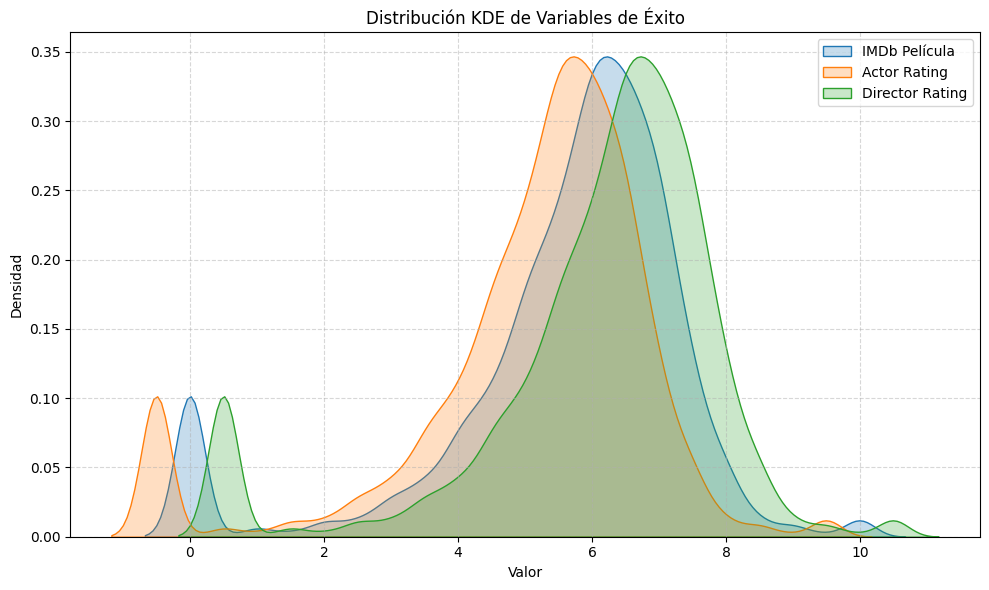
Hasta el momento, se han realizado los siguientes avances:

* Se identificaron y clasificaron las variables del conjunto de datos si eran cualitativas y cuantitativas, continuas y discretas.
* La realización de la base de datos para este trabajo, ya que ninguna de las Data Bases en línea ninguna se acomodaba a lo que nos planteamos.
* Se han calculado estadísticas descriptivas como promedios, medianas y desviaciones estándar para variables numéricas como la calificación IMDb y las valoraciones de actores y directores.
* Se elaboraron gráficas de barras y diagramas de dispersión que muestran tendencias entre la clasificación por edad y la calificación promedio.
* Se observaron correlaciones preliminares entre el género de la película y su desempeño general.
* También se han comenzado pruebas con modelos predictivos básicos para explorar la viabilidad del modelo final.

## Grafica de visualización



Este mapa de calor permite identificar rápidamente qué variables numéricas tienen relaciones significativas entre sí. En este caso, las calificaciones y nominaciones de actores/directores tienen una fuerte relación con la calificación IMDb, lo cual puede ser útil en análisis predictivos o de éxito cinematográfico.



El gráfico muestra que las tres variables clave de éxito (calificación IMDb, rating del actor y del director) tienen distribuciones similares, concentradas principalmente entre los valores 5 y 7. Esto sugiere una relación consistente entre estos factores en la percepción general del éxito de una película.