

## Entrega Numero Dos

### Informe de Progreso del Proyecto

#### Proyecto: **TMDB Box Office Predicción**

#### Participantes:

Steven Alipio Berrio – 1036661504 – Ing. Sistemas

Sergio Andrés Zapata Ruiz – 1037648161 - Ing. Mecánica

#### Introducción

El presente informe describe el progreso alcanzado en nuestro proyecto de construcción de un modelo de inteligencia artificial (IA) para predecir la recaudación de películas. Hasta el momento, hemos centrado nuestros esfuerzos en el proceso de aprendizaje de las herramientas necesarias y en la exploración de los datos contenidos en el archivo 'train.csv'. Nuestro objetivo final es comprender y visualizar las características del conjunto de datos y prepararlos para la creación de un modelo IA eficiente.

#### Contexto del Proyecto

En esta etapa inicial del proyecto, hemos estado concentrados en el aprendizaje y la adquisición de habilidades para trabajar con datos y programación. Hemos seleccionado un conjunto de datos de películas ('train.csv') como punto de partida. Este conjunto de datos contiene diversas características de películas, como el presupuesto, el género, el dinero recaudado, entre otros. En esta competencia, se le presentan metadatos sobre más de 7000 películas anteriores de The Movie Database para intentar predecir sus ingresos generales de taquilla en todo el mundo. Los datos proporcionados incluyen elenco, equipo, palabras clave de la trama, presupuesto, carteles, fechas de lanzamiento, idiomas, compañías de producción y países; Datos disponibles antes del estreno de una película. Las películas recaudaron aproximadamente 41.700 millones de dólares en 2018, la industria cinematográfica es más popular que nunca. Pero, ¿qué películas generan más dinero en taquilla? ¿Cuánto importa un director? ¿O el presupuesto?

#### Progreso Alcanzado:

Hasta el momento, hemos logrado avanzar en las siguientes áreas:

#### Aprendizaje de Herramientas de Programación

Nuestro enfoque inicial ha estado en la comprensión y el dominio de las herramientas necesarias para llevar a cabo este proyecto como lo son la importación de datos en Colab,

el manejo de datos de los archivos para la creación de dataframes y datasets, la exploración correcta de datos para que funcione correctamente nuestra predicción. Reconocemos que este proceso de aprendizaje es esencial antes de avanzar en otras etapas. Aunque ha tomado tiempo, estamos adquiriendo la destreza necesaria para utilizar estas herramientas y así tener un mejor desempeño a la hora de realizar nuestro proyecto.

#### Exploración de Datos:

Hemos realizado una exploración inicial de los datos contenidos en el archivo 'train.csv'. Para comprender mejor el conjunto de datos, hemos utilizado las siguientes estrategias:

**Visualización de Datos:** Hemos generado algunas gráficas para comparar columnas específicas del archivo 'train.csv' con la cantidad de dinero recaudado por película. Estas visualizaciones nos han ayudado a obtener una idea visual de cómo ciertas características pueden influir en la recaudación.

#### Desafíos Actuales:

Aunque hemos realizado avances en el aprendizaje de herramientas y la exploración de datos, enfrentamos desafíos en nuestro proyecto:

**Complejidad de las Herramientas:** La complejidad de las herramientas de programación y análisis de datos ha requerido un proceso de aprendizaje continuo. Hemos sido conscientes de la necesidad de dominar estas herramientas antes de avanzar a etapas posteriores.

También se ha tenido complejidad en el aprendizaje de manejo de modelos y dataframes ya que es algo nuevo para nosotros, así como la implementación del proyecto en diferentes plataformas como GitHub en las cuales no hemos tenido un acercamiento muy importante a la hora de manejar proyectos de este tipo.

#### Pruebas de Modelos:

Reconocemos que hasta el momento no hemos avanzado lo suficiente para probar nuestros primeros modelos. Sin embargo, con el tiempo, esperamos utilizar estas herramientas con mayor facilidad y acelerar nuestro progreso. Estimamos que comenzaremos a probar modelos a partir de la próxima semana y tener un mayor avance en lo que respecta al proyecto

#### Conclusiones y Próximos Pasos:

Hasta la fecha, hemos centrado nuestros esfuerzos en aprender las herramientas necesarias y explorar los datos. Si bien aún no hemos abordado la limpieza de datos y otros aspectos críticos, reconocemos la importancia de adquirir conocimientos sólidos antes de avanzar. A medida que ganemos experiencia en el uso de estas herramientas, esperamos optimizar nuestras actividades y avanzar más rápidamente.

Nuestro próximo paso será abordar la limpieza de datos y empezar a implementar los modelos necesarios en nuestro proyecto para un buen procesamiento de datos.

Estamos comprometidos a continuar avanzando en el proyecto con el objetivo de crear un modelo de IA eficaz para predecir la recaudación de películas que es el objetivo principal de nuestra competencia.