

## Tabla de contenido

Tabla de contenido .....	1
1. Proceso de automatización .....	2
1.1 Propósito de la automatización.....	2
1.2 Preparación de los datos .....	2
1.3 Construcción del modelo.....	2
1.4 Persistencia del modelo .....	2
1.5 Desarrollo del API .....	3
2. Desarrollo de la aplicación y justificación .....	3
2.1 Aplicación .....	4
2.2 Conexión entre la aplicación y el negocio .....	4
2.3 Relevancia de la aplicación .....	5
2.4 Mejoras al hablar con el experto.....	5
3. Trabajo en equipo.....	6
3.1 Roles dentro del equipo .....	6
3.2 Reflexión.....	7

# 1. Proceso de automatización

## 1.1 Propósito de la automatización

Para poder generar una aplicación que agregue valor a la organización es necesario que la organización pueda ingresar la información de la forma más natural posible, para poder lograr esto es necesario que el proceso de preparación de datos, carga del modelo y predicción por parte del modelo este de forma generalizada y automatizada. De este modo, preparamos los datos, cargamos un modelo ya entrenado y presentamos los resultados del modelo de forma natural.

## 1.2 Preparación de los datos

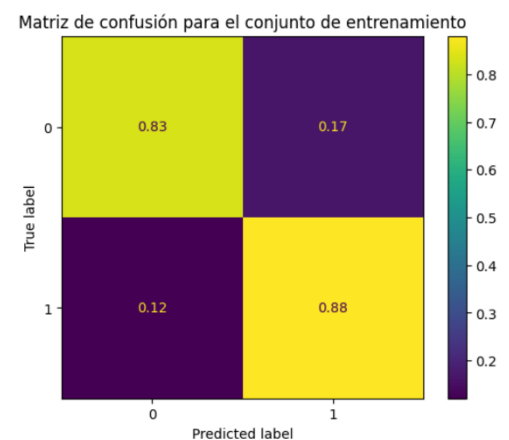
Para poder preparar los datos, hacemos varias transformaciones de los datos que ingresa la empresa o el usuario de nuestra aplicación. Las transformaciones que realizamos son:

- Quitar todos los caracteres que no fueran ASCII de las reseñas cargadas por el usuario.
- Convertir todas las palabras a minúsculas para unificar palabras que contengan minúsculas y mayúsculas bajo la misma semántica.
- Eliminar la puntuación de las palabras, de igual forma que lo anterior para unificar a la semántica de palabras que tienen puntuación.
- Cambiar los números a texto para mantener la unicidad.
- Remover las "stop words" que hacen referencia a todas aquellas palabras que tienen una correlación muy baja o nula con la variable de salida que es el sentimiento. Esto se hace eliminando conectores y artículos de los textos.
- Vectorizar las palabras ingresadas y preprocesadas con las transformaciones anteriores bajo el método *count vectorizer* que aplica *bag of words* como método de vectorización.

## 1.3 Construcción del modelo

El modelo fue construido con anterioridad en la anterior etapa del proyecto, de este modo construimos un pipeline que exportamos mediante un *joblib*. El modelo fue realizado con el mejor modelo que logramos generar, que fue el de **regresión logística**. Debido a esto, este modelo fue introducido al pipeline que exportamos junto con sus mejores hiperparámetros. El modelo fue escogido, al igual que sus hiperparámetros en la anterior etapa del proyecto. Ya que en esta etapa habíamos encontrado el mejor modelo y sus mejores hiperparámetros, decidimos entrenar el modelo que utilizaremos en el API, con *todos* los datos que teníamos, para así aumentar al máximo posible la precisión del modelo para datos nuevos.

F1: 0.8628628628628628



## 1.4 Persistencia del modelo

El modelo fue guardado como un *joblib*, que fue entrenado con absolutamente todos los datos, de tal forma que el *joblib* tiene el mejor resultado posible. También definimos una clase de utilidades con la clase de preprocesamiento, de tal forma que el al cargar el modelo, este pudiera utilizar todo el

procesamiento que habíamos realizado anteriormente. El modelo es cargado una vez al correr el backend en Django, de tal forma que se optimiza el tiempo al sólo tener que ser cargado una vez para su correcto funcionamiento. Una vez ya el modelo está cargado, podemos ejecutar el método *predict* del pipeline para los nuevos casos entrantes.

## 1.5 Desarrollo del API

Para desarrollar el modelo utilizamos Django, usando en general 2 endpoints principales. El primero recibe una JSON que contiene una lista de reseñas que van a ser procesadas, estas reseñas son las que el usuario ingresa manualmente a través de la interfaz de usuario. Mientras que el segundo endpoint recibe un texto plano, que en realidad es el csv que contiene una lista de reseñas al ser cargadas desde un archivo csv. Esta decisión de diseño fue tomada para facilitar la extracción de los textos en Django directamente y no en el front. En Django se organiza esta información en un DataFrame, el cual contiene todas las reseñas de forma organizada, hace las transformaciones necesarias y devuelve el análisis de sentimiento con respecto a cada una de las reseñas dadas.

El front fue creado sobre un template de Django en el que se permite el uso de los endpoints anteriormente mencionados. Dentro de los campos de texto se puede ingresar directamente una o más reseñas añadiendo más campos de texto. El front permite:

1. Añadir más campos de texto para escribir reseñas directamente en ellos, y poder analizarlas.
2. Importar un csv desde la memoria local de tu computador. Si el csv tiene un tamaño superior a 1MB, las reseñas no se verán cargadas en la interfaz (debido a que serían demasiadas para ser mostradas) y se asume que en este caso lo importante sería simplemente el csv que retornamos con la clasificación de cada una de las reseñas que fueron enviadas.

El front también hace uso de una paleta de colores intuitiva, de forma que las reseñas positivas se mostrarán verdes, mientras que las reseñas negativas se mostrarán con un color rojo al ser negativas.

## Análisis de sentimiento

### Reseñas de películas

¡Bienvenido!

Para utilizar la aplicación puedes hacerlo de dos modos:

#### 1. Ingresando reseñas en las áreas de texto

- 1) Presiona el botón "+" para agregar una nueva área de texto
- 2) Presiona el botón "Analizar" para analizar las reseñas ingresadas
- 3) Presiona el botón "Borrar" para borrar las reseñas ingresadas

#### 2. Importando un archivo CSV con las reseñas a analizar:

- 1) Selecciona el separador del archivo CSV
- 2) Selecciona la columna del archivo CSV que contiene las reseñas
- 3) Presiona el botón "Importar CSV" para seleccionar el archivo CSV

**NOTA** Si el archivo CSV tiene un peso mayor a 1MB, no se mostrarán las reseñas en las áreas de texto pero sí se descargará un CSV con los resultados

Descargar CSV

+

Borrar

Analizar

Importar CSV

Separador de columnas: ,

Columna de reseñas: review\_es

Ingrese los textos a analizar:

## 2. Desarrollo de la aplicación y justificación

## 2.1 Aplicación

La aplicación es una página web que permite usar todo el potencial del modelo y pipeline desarrollado a través de una interfaz gráfica intuitiva y amigable con el usuario. La Aplicación está compuesta de diferentes elementos, por un lado, se pueden crear diferentes campos de texto para ingresar una reseña, por otro lado, se puede escribir en estos campos de texto las diferentes reseñas de personas sobre una película correspondiente y finalmente permite que el usuario ingrese reseñas desde un CSV y que sea capaz de descargar un csv con los resultados después del análisis del modelo con respecto a las reseñas que bien sea fueron cargadas o introducidas a mano en la aplicación.

Descargar CSV

Borrar

Importar CSV

+

Analizar

Separador de columnas: 

.

Columna de reseñas: 

review\_es

Ingrese los textos a analizar:

La película 'Parásitos' de Bong Joon-ho es una obra maestra del cine contemporáneo. La trama es ingeniosa y está repleta de giros inesperados que mantienen al espectador en vilo hasta el final. Además, la actuación del reparto es increíblemente buena, especialmente la interpretación de Song Kang-ho como el padre de la familia Kim. La película ofrece una mirada crítica e inteligente sobre las divisiones de clase en la sociedad y deja al espectador reflexionando sobre ella mucho después de que los créditos hayan terminado. En resumen, 'Parásitos' es una película que no te puedes perder.

La la land' es una película que enamora a primera vista. El dúo de Ryan Gosling y Emma Stone nos cautiva con su química en pantalla y con su talento musical. La cinematografía es impresionante, y la banda sonora es tan pegajosa que te la encontrarás tarareando incluso después de salir del cine. La trama es emotiva y nos hace reflexionar sobre nuestras propias decisiones en la vida. En general, 'La la land' es una película que tiene todo lo que se podría pedir en un musical moderno: estilo, corazón y una buena dosis de nostalgia por el Hollywood clásico

Sucker Punch' es una película pretenciosa y confusa que no logra ofrecer una experiencia satisfactoria. La trama es difícil de seguir y las actuaciones son mediocres en el mejor de los casos. A pesar de tener un reparto estelar, la película parece más interesada en mostrar secuencias de acción sin sentido que en contar una historia coherente. Además, el tratamiento que hace la película de temas serios como la opresión y el abuso es superficial y decepcionante. En resumen, 'Sucker Punch' es una película que intenta ser más de lo que es, pero que no logra cumplir con las expectativas

The Last Airbender' es una película lamentable que hace un flaco favor a la serie animada en la que se basa. La trama está mal adaptada, el diálogo es aburrido y las actuaciones son pésimas. Además, los efectos especiales son mediocres y no logran capturar la esencia de la magia y la fantasía que hicieron que la serie fuera tan popular. En resumen, esta película es una decepción absoluta que no merece ser vista

El gran hotel Budapest' es una película divertida, ingeniosa y visualmente impresionante. La trama está bien construida y las actuaciones son sobresalientes, especialmente la de Ralph Fiennes, que ofrece una actuación cómica memorable. La cinematografía es hermosa y la música complementa perfectamente el ambiente de la película. Además, la película es capaz de equilibrar hábilmente la comedia y la tragedia, y el resultado es una experiencia cinematográfica que te dejará riendo y reflexionando al mismo tiempo. En resumen, 'El gran hotel Budapest' es una película que vale la pena ver

Transformers: El último caballero' es una película ruidosa, caótica y sin sentido. La trama es confusa y las actuaciones son sobreactuadas. Además, los efectos especiales, aunque impresionantes, no logran compensar la falta de cohesión en la historia. Es una película que parece estar diseñada para los fans acérrimos de la franquicia, pero incluso ellos podrían encontrarla difícil de ver. En resumen, esta película es una muestra de cómo no hacer cine de acción y ciencia ficción

## 2.2 Conexión entre la aplicación y el negocio

La aplicación está fuertemente relacionada con el negocio, esto se debe a que es una herramienta de utilidad directamente este. Para este caso se asume que el negocio es una empresa que se dedica a vender películas por suscripción como Netflix, Amazon Prime o Disney Plus. Esto le beneficia al negocio de forma directa. Esto se da por varias razones, a continuación, se presentan los beneficios que le aporta al negocio:

- El negocio podrá recoger información cuando sus clientes reseñen una película nueva que hayan subido a la plataforma, para así saber si fue una buena decisión meter la película a la plataforma y en caso dado de que haya sido buena idea dejarla, si no, quitarla para mejorar el catálogo general.
- Esto también permite que la empresa pueda descargar las reseñas generales de toda la plataforma en un csv, para que sea analizado por la aplicación y esto les permita tener un histórico del porcentaje de películas que tienen reseñas positivas en la plataforma.
- La plataforma también permite que una empresa pueda comunicarse con empresas que reseñan películas, como por ejemplo Rotten Tomatoes para así saber si una película nueva es buena o no y con esta información revisar si valdría la pena incluirla en la plataforma.

- Como empresa proveedora de películas que a la vez genera películas saber si las reseñas sobre las películas que desarrolla la empresa son buenas o malas es importante para la empresa, ya que esto les permite filtrar las reseñas malas para ver qué factores se deben mejorar para desarrollar mejores películas en el futuro.

## 2.3 Relevancia de la aplicación

La aplicación es relevante, debido a que permite una interfaz más cómoda para la empresa, sin tener que correr ni una línea de código la empresa va a poder tener acceso a las funcionalidades de la aplicación. Además, le da versatilidad a la empresa, al tener diferentes formas de ingresar la información en la aplicación, a la vez que diferentes formas de extraer la información del sentiment análisis entregado. Esto se hace a través de csv, texto y colores que ilustran la clasificación del sentimiento de la reseña.

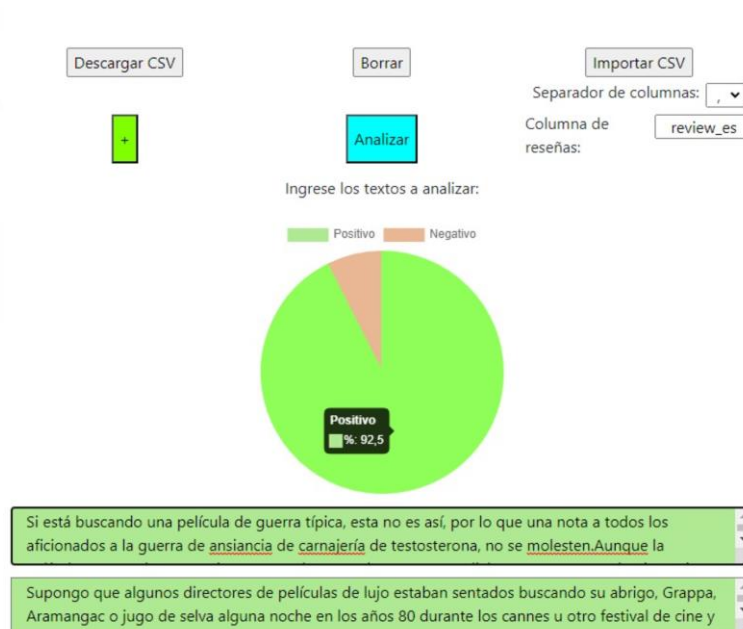
## 2.4 Mejoras al hablar con el experto

Al hablar con John, experto en estadística que fue asignado a nuestro equipo, nos brindó diferentes perspectivas con respecto a la validez de los resultados presentados por nuestra aplicación, así como diferentes métodos para mejorar los resultados tanto a nivel estadístico como a nivel de negocio.

En general, para poder justificar y validar que los resultados obtenidos por nuestra aplicación son válidos, es importante tener en cuenta:

1. **El tamaño de la muestra:** Un tamaño de muestra más grande generalmente proporciona una mayor precisión. Nuestro modelo fue entrenado con alrededor de 5000 reseñas diferentes, lo cual le brinda una mayor precisión.
2. **El método de muestreo:** Es ideal que los datos proporcionados provengan de fuentes aleatorias de modo que se eviten sesgos en las respuestas, lo cual pueda disminuir la precisión de la aplicación en un análisis real. En este sentido, las reseñas con las que fue entrenado nuestro modelo son aquellos datos que nos fueron suministrados, sin embargo, haciendo una revisión general de estas reseñas podemos observar que:
  1. No provienen de la misma película.
  2. Tienen una distribución bastante equilibrada de cerca del 50% positivas y 50% negativas.
  3. Se utilizan diferentes terminologías y semánticas dentro de las reseñas para indicar su opinión.
3. **La medida del sentimiento:** Debido a que se trata de una categorización binaria (1-positivo o 0-negativo) es correcto utilizar un análisis estadístico de frecuencia, ya que este representa en términos generales y sintetizados, qué porcentaje de las respuestas son positivas y negativas, así como un conteo directo, a la vez que es un estadístico que se puede obtener de manera directa de la clasificación que realiza nuestro modelo.
4. **El análisis estadístico:** El análisis de frecuencia permitiría identificar la cantidad de veces que se seleccionó cada categoría en la muestra, proporcionando información útil sobre la opinión de la audiencia hacia la película o cualquier otro objeto de estudio. En este caso, la medida del sentimiento puede ser útil para diferenciar entre los comentarios positivos y negativos, permitiendo un análisis más detallado del sentimiento de la audiencia.

Finalmente, otro de los aspectos principales dialogados con John, fue la importancia de la aplicación para las empresas objetivo que habíamos considerado. En este sentido, un punto clave dialogado fue la importancia de presentar un resumen directo de los datos en la aplicación, es decir, una forma fácil y rápida en la que la empresa objetivo pueda comprender el significado de los resultados obtenidos, sin necesidad de saber procesar datos en formato CSV u otros. Para esto se desarrolló un diagrama de torta que muestra de manera directa qué porcentaje de las reseñas son positivas y qué porcentaje negativas, con lo cuál se puede visualizar de manera directa si la mayoría de las reseñas son positivas o no.



En general poder hablar con John nos ayudó a tener una visión un poco más general y enfocada a un público más general, esto nos permite llegar a más personas con nuestra aplicación y facilitar su uso a los usuarios.

### 3. Trabajo en equipo

#### 3.1 Roles dentro del equipo

Brian Manuel Rivera:

Rol: El Ingeniero de software responsable del diseño de la aplicación y resultados: Brian se encargó en su mayor parte de la creación del proyecto y proyección de los resultados finales, y de velar por el aporte de información relevante para la organización en conjunto con la aplicación de las decisiones tomadas a raíz de las conversaciones con el experto.

Tony Santiago Montes

Rol: Ingeniero de software e Ingeniero de datos responsable de desarrollar la aplicación final: Tony se encargó de implementar un Pipeline que contuviera el mejor modelo en conjunto con su integración en el framework Django, además desarrolló el front de la aplicación en donde fue posible establecer una excelente interacción cliente-modelo.

Luccas Rojas

Rol: Líder de proyecto: Luccas Rojas fue el encargado de dirigir el proyecto, establecer los días y horas de reunión entre todos los integrantes, en organizar los entregables del grupo y asignar las tareas que cada uno de los integrantes iba a realizar para que fuera lo más equitativo posible. Además, realizó como tarea principal gran parte de la organización del documento y la información más relevante recopilada en esta.

### **3.2 Reflexión**

En general como grupo trabajamos muy bien, nos logramos repartir las diferentes secciones del proyecto de manera que todos los integrantes aportaran en la misma medida. Llevamos una buena comunicación en general y logramos desarrollar de forma adecuada el proyecto.

Tony Santiago Montes 33.3 - Brian Manuel Rivera 33.3 - Luccas Rojas 33.3

Aspectos por mejorar:

- Reunirnos más seguido
- Definir los roles y el trabajo por hacer desde el principio
- Mejorar la comunicación entre nosotros para coordinar las tareas y así no llegar a repetir las tareas que otros ya realizaron
- Mejorar la cooperación del equipo para que todos nos ayudemos en las tareas y podamos aprender de lo que el otro hizo.
- Tenemos que mejorar también la motivación que tenemos con respecto a las actividades, ya que llegamos a hacer el trabajo sólo por entregarlo.
- Debemos aprender a manejar mejor el tiempo para que nos rinda más y tengamos más tiempo para corregir los errores.