



Universidad de  
**los Andes**  
Colombia

Universidad de los Andes  
Inteligencia de Negocios

**Proyecto – Etapa 2**  
**Automatización y uso de modelos de analítica**  
**turismo de los Alpes**

**Nicolás Díaz – 202021006**  
**Sebastián Casanova– 202115116**  
**Julio Peña– 201616539**

# Contenido

1. Proceso de Automatización .....	3
1.1 Cambios y revisión de la etapa 1 .....	3
1.2 Implementación de Pipeline .....	3
2. Desarrollo de la aplicación.....	4
2.1 Revisión de lo planteado de la etapa 1 .....	4
2.1 Descripción del Usuario/Rol .....	5
2.2 Implementación de la aplicación .....	5
2.3 Conexión de la aplicación y el proceso de negocio. ....	6
2.4 Importancia para el rol la existencia de la aplicación. ....	6
2.5 Feedback y contribución del grupo de estadísticas.....	6
3. Trabajo en equipo Roles .....	7
3.1 Roles de los integrantes.....	7
3.2 Tareas .....	7
3.3 Repartición de puntos .....	7
3.4 Posibles puntos a mejorar.....	8

# 1. Proceso de Automatización

## 1.1 Cambios y revisión de la etapa 1

El ingeniero correspondiente realizó la revisión de los modelos implementados en la etapa 1 para mejorar los resultados obtenidos en las métricas. En primer lugar, como resultado de la reunión con el grupo de estadística, se tomó la decisión de eliminar el proceso de lematización de los datos que se había implementado en la etapa 1. Esto se debió a que las palabras resultantes de este proceso muchas veces eran imposibles de entender, lo que hacía que los resultados arrojados por los modelos originalmente carecieran de valor, ya que era imposible extraer información valiosa. Por otro lado, se cambió completamente la implementación del modelo de regresión lineal, ya que este no contribuía en nada a nuestro problema de negocios. Como consecuencia, se utilizó un modelo de regresión logística con TF-IDF. Gracias al trabajo interdisciplinario y a los cambios mencionados anteriormente, conseguimos mejores métricas que en la etapa anterior.

## 1.2 Implementación de Pipeline

Para automatizar el proceso de preparación de datos, construcción del modelo y persistencia del modelo, se utilizó un pipeline. Esto se debe a que es una herramienta útil que permite automatizar y estandarizar el flujo de trabajo en el análisis de datos y la construcción de modelos. En este caso, para la implementación del pipeline, se utilizó el modelo de regresión logística, ya que este obtuvo los mejores resultados en las métricas.

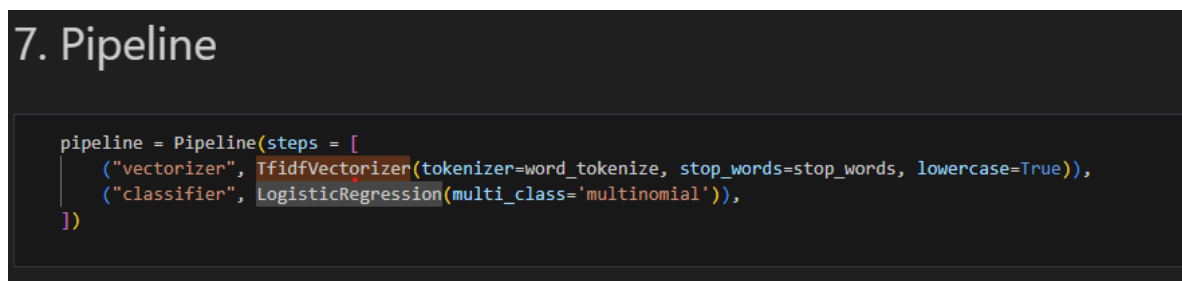


Ilustración 1. Captura de Visual Studio Code donde se muestra la implementación inicial del Pipeline.

## 1.3 Persistencia del modelo

Para persistir nuestro modelo ya entrenado, lo guardamos en un archivo con formato Joblib, el cual permite almacenar objetos de Python. Esto se realiza para que luego podamos utilizar el modelo conectándolo con el front para satisfacer las necesidades del usuario.

## 2. Desarrollo de la aplicación

### 2.1 Revisión de lo planteado de la etapa 1

Para la etapa 1 se realizó el siguiente mapa de actores relacionados con el producto:

Rol dentro de la empresa	Tipo de actor	Beneficio	Riesgo
Equipo de Análisis de Datos	Equipo Interno	Mayor comprensión de las características que hacen atractivos los destinos turísticos, lo que facilita la toma de decisiones estratégicas y la creación de campañas de promoción más efectivas.	Riesgo de que el modelo no proporcione insights relevantes o precisos, lo que podría resultar en la toma de decisiones erróneas y desperdicio de recursos.
Equipo de Marketing y Promoción	Equipo Interno	Acceso a información detallada sobre las preferencias y comportamientos de los turistas, lo que permite diseñar estrategias de marketing más segmentadas y personalizadas para promover los destinos turísticos de manera efectiva.	Posibilidad de malinterpretar los resultados del modelo y basar campañas de marketing en información incorrecta, lo que podría afectar la reputación de la organización y la atracción de turistas.
Equipo de Desarrollo de Productos y Servicios	Equipo Interno	Mejora en la oferta de productos y servicios turísticos, al contar con insights basados en datos sobre las preferencias y necesidades de los turistas, lo que aumenta la satisfacción del	Riesgo de depender demasiado del modelo y descuidar otros aspectos importantes de la gestión de productos y servicios, como la calidad y la innovación, lo que podría afectar la

		cliente y la competitividad de la organización.	experiencia del cliente.
Alta Dirección	Gestión	Mecanismo de toma de decisiones basado en información objetiva y precisa, lo que facilita la planificación estratégica y el cumplimiento de los objetivos organizacionales relacionados con el turismo.	Posibilidad de tomar decisiones erróneas si la información proporcionada por el modelo no es confiable o relevante, lo que podría afectar la reputación y el desempeño general de la organización.

Después de reunirse con el grupo de estadística y realizar la respectiva revisión, se concluyó que los roles planteados son correctos y no es necesario realizar cambios en la tabla de tipo de actores. Finalmente se concluyó que el directamente el usuario/rol que usará la aplicación serán los integrantes del equipo de análisis de datos de la empresa.

## 2.1 Descripción del Usuario/Rol

El desarrollador de Productos y Servicios tiene la responsabilidad de crear productos y servicios nuevos, así como de refinar y mejorar los servicios actuales. Nuestro problema de negocio indica que el objetivo del proyecto es potenciar el turismo local a través de la implementación de nuevas estrategias, las cuales se derivan del análisis de las reseñas proporcionadas por los usuarios. El desarrollador de productos y servicios está directamente involucrado en esta tarea, ya que es quien propone las ideas que se implementarán y que impactarán en el rendimiento de la industria hotelera.

## 2.2 Implementación de la aplicación

Para la implementación de la aplicación se utilizaron dos frameworks: FastAPI, con el que desarrollamos la parte del back-end, y ReactJS, con el que se diseñó la interfaz de usuario, es decir, el front-end.

### 2.2.1 Front-end

En el Front-end, el usuario comenzará por cargar un archivo con la o las reseñas de las cuales desea obtener la predicción de su calificación. Este archivo será enviado al Back-end para que este realice su tarea (que será explicada en el siguiente punto). Posteriormente, el Back-end retornará las métricas

obtenidas, además de un archivo .CSV que contiene las reseñas que el usuario envió inicialmente, junto con las clasificaciones predichas por el modelo montado en el pipeline. Entonces, el usuario tendrá la opción de descargar este archivo.

### **2.2.2 Back-end**

Gracias a FastAPI, el back-end recibe, mediante una solicitud tipo @POST, el archivo cargado por el usuario. A continuación, el back-end procede a procesar el archivo recibido utilizando el archivo .joblib que contiene nuestro pipeline con nuestro modelo. Este modelo se encarga de predecir las clasificaciones para las reseñas entregadas por el usuario. Las clasificaciones respectivas para cada reseña generadas por el modelo se agregan al archivo .CSV. Este archivo estará disponible para que el usuario lo descargue si realiza una solicitud tipo @GET.

## **2.3 Conexión de la aplicación y el proceso de negocio.**

La aplicación permite al desarrollador de Productos y Servicios obtener las calificaciones correspondientes a las reseñas proporcionadas por los usuarios. A partir de estas predicciones, el desarrollador tendrá más y mejor información para respaldar su proceso de creación de nuevos productos y servicios que impulsen la industria hotelera, lo que facilitará el logro de los objetivos establecidos por el proyecto.

## **2.4 Importancia para el rol la existencia de la aplicación.**

El desarrollador de Productos y Servicios tiene la responsabilidad de crear productos y servicios nuevos, así como de refinar y mejorar los servicios actuales. Nuestro problema de negocio indica que el objetivo del proyecto es potenciar el turismo local a través de la implementación de nuevas estrategias, las cuales se derivan del análisis de las reseñas proporcionadas por los usuarios. El desarrollador de productos y servicios está directamente involucrado en esta tarea, ya que es quien propone las ideas que se implementarán y que impactarán en el rendimiento de la industria hotelera.

## **2.5 Feedback y contribución del grupo de estadísticas.**

El grupo de estadística contribuyó al refinamiento de la información obtenida por los modelos. Recomendaron abordar de manera diferente el proceso de tratamiento de datos, especialmente el manejo de las raíces de las palabras, ya que estas por sí solas no proporcionaban suficiente información. También se recomendó usar otras métricas estadísticas adicionales que podrían ofrecer más claridad y confianza al usuario, permitiéndole tener más certeza sobre la fiabilidad de las predicciones del modelo. Además, se solicitó que el usuario pudiera conocer qué modelo se estaba utilizando para comprender mejor su funcionamiento.

### **3. Trabajo en equipo Roles**

#### **3.1 Roles de los integrantes.**

Líder de equipo: Julio Peña

Ingeniero de datos: Nicolas Diaz

Ingeniero de software: Sebastián Casanova.

#### **3.2 Tareas**

##### **3.2.1 Líder de equipo**

El líder del equipo se encargó de la delegación de las tareas del equipo, supervisó que se cumplieran correctamente y en los tiempos establecidos. También se encargó de supervisar que los productos entregados funcionaran correctamente, por lo que debía conocer e involucrarse en todas las áreas. Finalmente, el líder del equipo elaboró el documento donde están descritos todos los entregables para la parte dos del proyecto. Esto representó aproximadamente 13 horas de trabajo.

##### **3.2.2 Ingeniero de datos**

El ingeniero de datos revisó y corrigió los modelos generados para la etapa 1. Esto incluyó mejorar los procesos de preparación de datos, tokenización y la implementación de hiperparametros. Además, rehízo completamente el modelo de regresión lineal, sustituyéndolo por uno de regresión logística. Se encargó de mejorar los resultados en las métricas para que nuestros modelos fueran más confiables. Finalmente, el ingeniero de datos construyó el pipeline y generó el archivo .joblib para que el ingeniero responsable de la aplicación pudiera utilizarlo en el desarrollo final.

##### **3.2.3 Ingeniero de software responsable del diseño de la aplicación**

El ingeniero de software se responsabilizó del desarrollo total de la aplicación. En primer lugar, tomó la decisión final sobre los frameworks que se utilizarían en la implementación. En cuanto al front-end, el ingeniero de software diseñó todos los elementos involucrados en la interfaz del usuario utilizando React.js. Finalmente, desarrolló la lógica y las solicitudes del back-end, utilizando FastAPI, y se encargó de conectar ambos ends.

#### **3.3 Repartición de puntos**

Lider del equipo: 30

Ingeniero de datos: 35

Ingeniero de software: 35

### **3.4 Posibles puntos a mejorar.**

Para esta entrega, una distribución clara del trabajo contribuyó a que cada integrante realizara sus tareas, lo que resultó en una mayor calidad de los entregables y en una mejor gestión del tiempo. Sin embargo, faltó establecer plazos de entrega, especialmente porque el trabajo de algunos dependía directamente del de otros compañeros, lo que requería un mayor grado de coordinación y donde surgieron problemas. Por otro lado, la comunicación con el equipo multidisciplinario no fue efectiva; fue difícil establecer reuniones y compartir ideas, aunque al final resultó satisfactorio, consideramos que este es un aspecto en el que debemos trabajar para la siguiente entrega.