

Proyecto: Turismo de los Alpes

Etapa 2

| | |
|--|---|
| Etapa 2 | 1 |
| 1. Equipo de trabajo y tareas asignadas | 2 |
| 2. Proceso de automatización | 2 |
| 2.1. Tecnologías utilizadas | 2 |
| 2.2. Preparación de Datos y Construcción del Modelo: | 2 |
| 3. Desarrollo de la Aplicación | 5 |
| 3.1. Descripción del usuario..... | 5 |
| 3.2. Funcionalidades de la aplicación | 5 |
| 3.3. Importancia para el usuario | 5 |
| 4. Comunicación con el grupo de estadística | 6 |
| 4.1 validación del enfoque del proyecto..... | 6 |
| 4.2 Prueba de la aplicación..... | 6 |

1. Equipo de trabajo y tareas asignadas

El equipo de trabajo está conformado por 3 integrantes:

- Nicolas Camargo: Ingeniero de Software (Desarrollo de Aplicación). Desarrolla la aplicación final. 10 horas de trabajo.
- Pablo Pedreros: Líder de proyecto e ingeniero de Datos. Se encarga de gestionar el proyecto y las tareas del equipo, prepara los datos y construye el modelo analítico. 10 horas de trabajo.
- Jairo Garavito Correa: Ingeniero de Software (Diseño de Aplicación y Resultados). Diseña la aplicación y genera el video con los resultados. Documentar todo su trabajo. 8 horas de trabajo.

El principal reto del proyecto fue diseñar el API con el pipeline del modelo analítico desarrollado en la primera etapa. Otro reto fue diseñar la aplicación para poder predecir las calificaciones.

Según lo trabajado por el grupo, los 100 puntos serían repartidos equitativamente. Para próximos proyectos la única mejora sería hacer contacto más estrecho y temprano con el equipo de expertos para enfocar el proyecto correctamente desde el inicio.

Así mismo, organizamos reuniones en dos momentos, para iniciar el backend de la aplicación y para la finalización, donde se realizó la prueba de la aplicación en su totalidad, comprobando la funcionalidad y presentando el resultado de esta al equipo de estadística.

2. Proceso de automatización

En base a los resultados obtenidos en la primera etapa del proyecto, se realizó el siguiente proyecto:

2.1. Tecnologías utilizadas

- FastAPI: Utilizado para construir la API que servirá el modelo analítico a la aplicación web desde el backend, y conectado a través del pipeline de preparación de datos.
- React: Utilizado para construir el frontend de la aplicación web, lo que permitirá una experiencia de usuario fluida e interactiva.
- Jupyter Notebook: En este proyecto, se utilizará Jupyter Notebook para desarrollar y entrenar el modelo analítico, y luego exportarlo como un pipeline que se puede cargar y utilizar en la API.

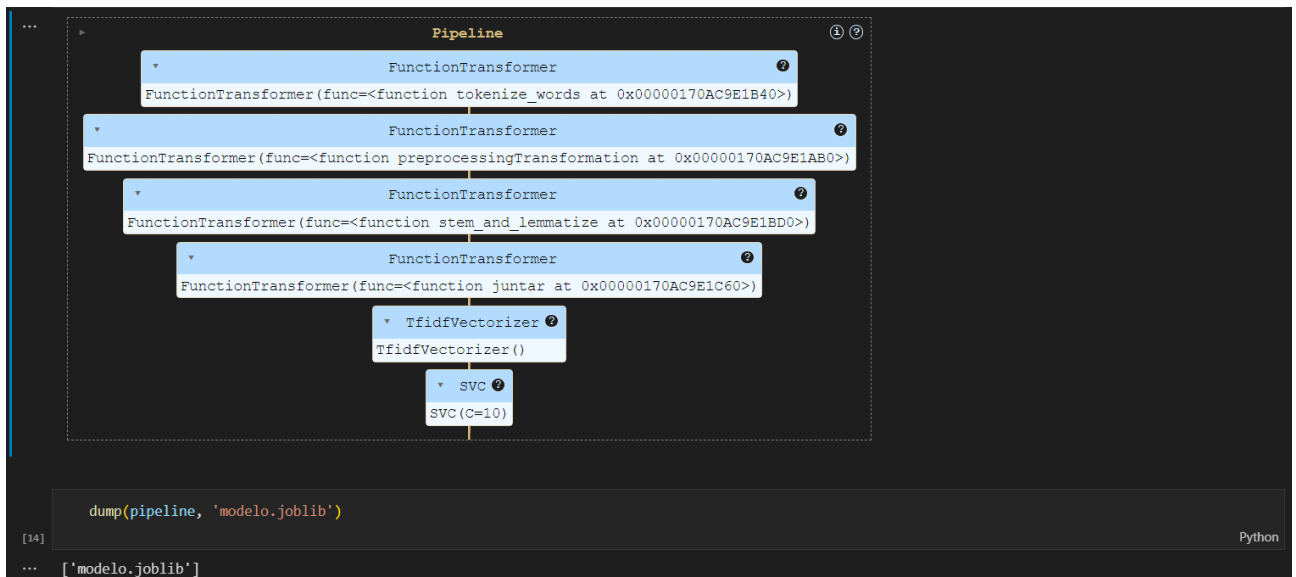
2.2. Preparación de Datos y Construcción del Modelo:

El ingeniero de datos será responsable de preparar los datos para el análisis y construir un modelo de análisis de sentimientos. Esto implica limpiar los datos, realizar análisis exploratorio y seleccionar o desarrollar un modelo de aprendizaje

automático para predecir el sentimiento de las reseñas. Este proceso cuenta con los siguientes pasos:

1. **Preparación de datos:** Se limpiarán los datos de reseñas de sitios turísticos junto con sus calificaciones asociadas. Es importante asegurarse de que los datos estén completos, sean coherentes y estén libres de ruido o datos irrelevantes.
2. **Selección y representación de características:** Se seleccionarán las palabras o características más relevantes de las reseñas que puedan influir en la calificación de los sitios turísticos. Estas palabras se utilizarán para representar las reseñas como variables en el modelo analítico. Se van a utilizar técnicas como la tokenización, eliminación de palabras irrelevantes (stopwords) y lematización para procesar el texto y extraer las características más importantes.
3. **Desarrollo del modelo analítico:** Se desarrollará un modelo de procesamiento de textos. Para esto se van a explorar diferentes algoritmos, como Multinomial Naive Bayes, árboles de decisión o Support Vector Machines, para construir un modelo que sea capaz de predecir la calificación de las reseñas basándose en las palabras seleccionadas como características.
4. **Validación del modelo:** Se evaluará el rendimiento del modelo utilizando técnicas de validación cruzada y métricas como precisión, sensibilidad (recall), precisión, y F1-score. Es crucial asegurarse de que el modelo sea capaz de generalizar correctamente a nuevos datos y que tenga un rendimiento satisfactorio en la clasificación de reseñas.

Para integrar el modelo en la página web, se tuvo que exportar el modelo del jupyter notebook por medio de la librería joblib. De esta forma, se creó el pipeline que se ve en la siguiente figura. La primera transformación tokeniza las palabras, la segunda hace todo el proceso de cambiar las palabras a minúsculas, quitar puntuación y caracteres no ASCII y quitar palabras no relevantes para el análisis (stopwords). Todas las transformaciones tuvieron que definirse en un archivo diferente dentro del proyecto para poder ser usadas tanto en el notebook, como en el back de la aplicación. Finalmente, el modelo se exporta a un archivo con la función dump.



Ya con este pipeline, se implementaron los dos endpoints que realizan predicciones en el back usando la función load de la librería joblib para poder cargar el modelo y hacer que prediga los resultados que serán guardados en ambas funciones como se ve en el siguiente código para luego ser mostrados al usuario final.

```
37 @app.post("/predictCSV")
38 async def make_predictions(file: UploadFile = File(...)):
39     contents = await file.read()
40     bytes_io = BytesIO(contents)
41     df = pd.read_csv(bytes_io)
42     model = load("assets/modelo.joblib")
43     df["Class"] = model.predict(df["Review"])
44     return df.to_dict()
45
46
47 class PredictionRequest(BaseModel):
48     text: str
49
50 class PredictionResponse(BaseModel):
51     predictions: dict
52
53 @app.post("/predictText", response_model=PredictionResponse)
54 async def make_predictionsText(request: PredictionRequest):
55     df = pd.DataFrame({'Review': [request.text]})
56     model = load("assets/modelo.joblib")
57     prediction = model.predict(df["Review"])[0]
58     prediction = int(prediction)
59
60     return {"predictions": {"Review": request.text, "Class": prediction}}
```

3. Desarrollo de la Aplicación

El ingeniero de software será responsable de desarrollar una aplicación web que permita a los usuarios interactuar con el modelo analítico. Esta aplicación debe permitir a los usuarios ingresar texto y recibir una predicción de sentimiento, junto con un puntaje asociado.

3.1. Descripción del usuario

El usuario de la aplicación es principalmente un profesional o experto en turismo que trabaja en el Ministerio de Comercio, Industria y Turismo de Colombia, en la Asociación Hotelera y Turística de Colombia (COTELCO) o en cadenas hoteleras reconocidas como Hilton, Hoteles Estelar, Holiday Inn, entre otros. Este usuario tiene un interés directo en comprender las características de los sitios turísticos y en mejorar su atractivo para los turistas locales e internacionales. Este usuario presenta el siguiente perfil:

- **Profesionales del Turismo:** Este usuario puede ser un analista de datos, un gerente de marketing, un planificador de destinos turísticos, o cualquier otro profesional involucrado en la gestión y promoción del turismo en Colombia.
- **Investigadores del Sector Turístico:** Aquellos que realizan investigaciones y análisis de mercado para comprender las tendencias del turismo y formular estrategias para mejorar la industria turística en Colombia.
- **Gestores de Sitios Turísticos:** Gerentes de hoteles, parques naturales, museos, y otros destinos turísticos interesados en recibir retroalimentación sobre la experiencia de los visitantes para mejorar sus servicios y atractivos.

3.2. Funcionalidades de la aplicación

La aplicación cuenta con las siguientes funcionalidades:

- **Ingreso de Reseñas por Texto o Archivo CSV:** Los usuarios pueden ingresar reseñas turísticas directamente en la aplicación a través de texto o cargar un archivo CSV con múltiples reseñas.
- **Predicción de Calificaciones:** La aplicación utiliza un modelo analítico para predecir las calificaciones que los turistas darían a los sitios turísticos basándose en las reseñas ingresadas. Esto permite a los usuarios obtener una comprensión rápida del sentimiento general de los visitantes hacia un destino turístico específico.
- **Visualización de Resultados:** Los usuarios pueden ver los resultados de las predicciones, incluyendo la calificación predicha en un formato de tabla y la probabilidad de acierto en dicha calificación.

3.3. Importancia para el usuario

El usuario de la página puede contar con los siguientes beneficios:

- **Toma de Decisiones Informadas:** Permite a los profesionales del turismo tomar decisiones informadas sobre cómo mejorar la experiencia de los turistas y promover destinos turísticos específicos.
- **Identificación de Oportunidades de Mejora:** Ayuda a identificar áreas de mejora en sitios turísticos que puedan estar recibiendo calificaciones bajas, lo que permite a los gestores tomar medidas correctivas.
- **Planificación Estratégica:** Facilita la planificación estratégica en la promoción y desarrollo de destinos turísticos, ayudando a priorizar recursos y esfuerzos en áreas que tengan el mayor impacto en la satisfacción del visitante.

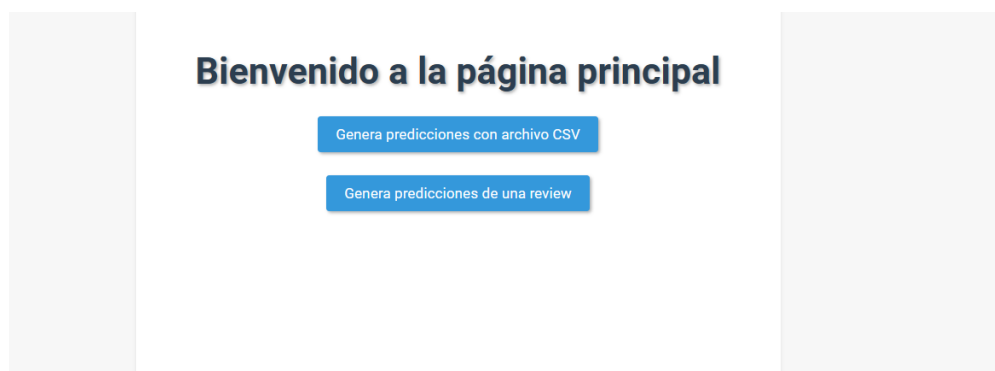
4. Comunicación con el grupo de estadística

4.1 Validación del enfoque del proyecto

Para comenzar esta etapa del proyecto se recibieron comentarios sobre el primer enfoque que se le dio al proyecto, en lo que se destaca la confirmación de los intereses del proyecto como el aumento de turistas y el desarrollo económico. Por otro lado, se encontraron falencias en la preparación del modelo, donde se menciona que algoritmo lo único que está haciendo es identificar cuáles fueron las palabras que más se repitieron e incidieron, además de que es probable que las métricas de datos que tienen tampoco sean idóneas.

4.2 Prueba de la aplicación

En este punto se presentó la aplicación en su fase final, para tener una retroalimentación. En primer lugar, la imagen siguiente presenta la página principal que encuentra el usuario con dos opciones, la de generar las predicciones por medio de un archivo CSV que debe subir o la de generar las predicciones ingresando el texto manualmente.



Posterior a esto, si ingresamos la opción 1 de cargar el archivo por CSV, ingresaremos a una página donde nos pide seleccionar un archivo de nuestro computador (.csv) y luego al darle al botón de generar predicciones arrojará una tabla con los resultados.

Carga de Archivo CSV y Predicciones

Seleccionar Archivo

Realizar Predicciones

Carga de Archivo CSV y Predicciones

particion_prueba_estudiantes - Copy.csv

Realizar Predicciones

Resultados de las Predicciones:

El modelo hace una predicción con una precisión de 60%

| Review No. | Review | Calificacion |
|------------|---|--------------|
| 0 | La primera noche nos encontramos en la habitación con un nido (5) de cucarachas muertas y la alfombra muy sucia...que por cierto nunca limpiaron hasta que reclamamos. La cena de fin de año fue un bufete que acabó en un tremendo desorden por el excesivo número de mesas vendidas, no se podía caminar para ir a servirte la comida, para luego encontrarte con bandejas vacías, mi esposa y yo y otra pareja 'NO CENAMOS' porque nos encontramos con puras bandejas vacías!!...pagamos \$2800 pesos por por pareja por nada!!! El servicio del restaurante pesimo, pesimo,pesimo....se tardaron muchísimo tiempo en atendernos, nos trajeron el desayuno en partes, nunca nos dieron el café que pedimos desde que llegamos....un pésimo servicio El frigobar vacío y después de reclamar lo surtieron pero_x0085_ | 1 |
| 1 | A una calle de catedral con platillos tradicionales, tipo Gourmet, de buen sabor y calidad. Si bien ofrecen una carta con los platillos más representativos de cada temporada como mole de caderas, poblano, chinicules, chiles en nogada, escamoles y cemitas; son principalmente para degustación, no...Más | 4 |

Por otro lado, al ingresar a la opción 2 de generar predicciones por texto plano, la página nos mostrara un recuadro vacío, en el que podemos ingresar la review que deseemos y al darle al botón de realizar predicciones, se genera el resultado de la predicción y su correspondiente probabilidad de ser correcto.

En general, este fue el despliegue que se presentó a las estudiantes de estadística, quienes nos dieron su retroalimentación respecto a lo esperado que establecieron para la etapa 1 y reafirmaron que en general tiene buen aspecto en términos de funcionalidad para la propuesta de negocio. Sin embargo, también recalcaron como algunos aspectos gráficos y de presentación podrían mejorarse, esto haciendo referencia al uso del usuario en el front. Finalmente, establecieron que, si bien no se realizó un cambio del modelado como tal, cumple con los requisitos que se nos plantearon para la entrega de este proyecto y la correcta predicción de calificaciones en base a las reseñas.

Predicciones con Texto Plano

Ingrese el texto para hacer predicciones:

El hotel fue medianamente decente, la comida estuvo bastante buena pero, habia mucho ruido, lo cual fue incomodo

Realizar Predicciones

Resultado de la Predicción:

El modelo hace una predicción con una precisión de 6 0%

Review: El hotel fue medianamente decente, la comida estuvo bastante buena pero, habia mucho ruido, lo cual fue incomodo

Clase: 3

4. Mapa de actores

En la identificación de los actores se construyó la siguiente tabla:

| Rol dentro de la empresa | Tipo de actor | Beneficio | Riesgo |
|---|--------------------|---|---|
| Turista | Beneficiado | Acceso a reseñas detalladas para decisiones de viaje más informadas. | Dependencia excesiva de opiniones individuales que podrían no coincidir con las expectativas personales. |
| Sitios turísticos | Cliente | Identificación de áreas de mejora y aumento de la popularidad gracias al análisis de las reseñas. | Riesgo de críticas negativas públicas que podrían afectar la reputación del sitio si no se gestionan adecuadamente. |
| Ministerio de Comercio, Industria y Turismo de Colombia | Financiador | Mejora del turismo y desarrollo económico a través de la promoción efectiva de destinos turísticos. | Riesgo de inversión no justificada si los resultados del análisis no se traducen en mejoras tangibles en el sector turismo. |
| Asociación Hotelera y Turística de Colombia – COTELCO | Proveedor de datos | Adaptación de servicios y promociones para satisfacer mejor las necesidades de los clientes. | Riesgo de violación de privacidad de datos si no se manejan adecuadamente. |

