

Proyecto 1 Entrega 2 Analítica de Textos – Turismo de los Alpes

Docente: Haydemar Nuñez

David Cuevas Alba - 202122284

Eduardo José Herrera Alba - 201912865

Samuel Alejandro Jiménez Ramírez - 202116652

Sección: 3

Universidad de los Andes

Departamento de Ingeniería de Sistemas y Computación

2024-10

1. Proceso de automatización y acceso mediante un API

Automatización:

El proceso de automatización está diseñado para optimizar la preparación y el análisis de datos a través de un sistema denominado *Pipeline*. Este sistema automatiza el tratamiento de los datos de entrada para adaptarlos a las necesidades del modelo analítico, facilitando una evaluación más eficaz. A continuación, se detallan los componentes y el flujo de trabajo que definimos en nuestro *Pipeline*:

- **Preprocesador:** Esta parte se encarga de tomar las reseñas (o comentarios), procesarlas y transformarlas en un formato adecuado para que el modelo las pueda procesar correctamente. Esta parte incluye la codificación vectorial de los datos, en la cual importamos el mismo vectorizador que usamos en la etapa anterior del proyecto (matriz de características TF-IDF).
- **Modelo:** El sistema carga un modelo de clasificación previamente entrenado, almacenado en un archivo .joblib. Este archivo contiene el modelo seleccionado en la primera etapa del proyecto. El modelo implementado es un clasificador de Regresión Logística, elegido tras evaluar diferentes modelos y múltiples combinaciones de hiperparámetros. Este modelo se consideró el más adecuado para la tarea según las métricas obtenidas que evaluaron su desempeño en los datos etiquetados proporcionados.

Acceso mediante un API:

Luego de haber creado el Pipeline, se procedió a integrarlo en una API utilizando FastAPI, un framework web para crear APIs de forma rápida en Python. Primero, se configuraron las distintas rutas disponibles para interactuar con el sistema.

Se definen varias rutas para diferentes propósitos:

- **Ruta para la página principal (HTML):** Esta ruta sirve una página HTML estática que probablemente contenga información sobre el servicio ofrecido o instrucciones de uso. La página se lee desde un archivo HTML ubicado en el directorio de plantillas (templates) y se devuelve como respuesta.
- **Ruta de bienvenida básica:** Una ruta simple que devuelve un mensaje de bienvenida en formato JSON cuando se accede a la raíz de la API ("/").

- **Ruta para leer un ítem específico:** Esta ruta permite acceder a información sobre un ítem específico identificado por su `item_id`. Puede aceptar un parámetro opcional `q` para realizar una consulta adicional.
- **Ruta para hacer predicciones a partir de un conjunto de datos (archivo):** Esta ruta recibe datos en formato `DataModel`, los convierte en un `DataFrame` de Pandas y los utiliza como entrada para el Pipeline previamente entrenado. Luego, se hacen predicciones sobre estos datos y se devuelven los resultados.
- **Ruta para hacer predicciones a partir de un único dato (individual):** Similar a la ruta anterior, pero diseñada para recibir un solo dato (en este caso, una reseña) en lugar de un conjunto de datos. El dato se procesa de manera similar y se hacen predicciones sobre él utilizando el Pipeline.

2. Desarrollo de la aplicación y Justificación

Descripción del usuario/rol de la organización:

El usuario principal de la aplicación desarrollada para el Ministerio de Comercio, Industria y Turismo de Colombia es probablemente un funcionario del gobierno dedicado a la gestión y promoción del sector turístico en el país. Este usuario podría tener un cargo como Analista de Turismo, Coordinador de Promoción Turística o similar. Su responsabilidad principal es desarrollar estrategias y políticas que impulsen el turismo en Colombia, para aumentar los ingresos generados por el sector y mejorar la calidad de vida de los involucrados en el turismo.

Conexión entre la aplicación y el proceso de negocio:

La aplicación se conecta estrechamente con el proceso de negocio del Ministerio de Comercio, Industria y Turismo relacionado con la promoción y el desarrollo del turismo en Colombia. Proporciona una herramienta tecnológica para analizar y procesar datos relacionados con el turismo, como reseñas de turistas, opiniones sobre servicios turísticos y tendencias de viaje. Estos datos son fundamentales para comprender las necesidades y preferencias de los turistas, identificar áreas de mejora en la calidad de los servicios turísticos y diseñar estrategias efectivas para atraer a más turistas y aumentar los ingresos del país por este sector.

Importancia de la aplicación para el rol:

La aplicación es muy importante para el papel del funcionario encargado de promover el turismo en Colombia. Le proporciona una herramienta poderosa y eficiente para analizar y comprender el mercado turístico, identificar oportunidades de mejora y diseñar estrategias basadas en datos sólidos y análisis detallados. La capacidad de realizar predicciones sobre el comportamiento de los turistas y evaluar la efectividad de las estrategias implementadas es fundamental para el éxito en la promoción del turismo y la generación de ingresos para el país. Además, la aplicación ayuda a mejorar la calidad de vida de las personas que dependen del sector turístico al impulsar el crecimiento y la sostenibilidad de la industria, lo que se traduce en más empleos, mejores oportunidades económicas y un mayor desarrollo para las comunidades locales. En resumen, esta aplicación es una herramienta estratégica y valiosa para el usuario/rol dentro del Ministerio de Comercio, Industria

y Turismo, ya que contribuye significativamente a alcanzar los objetivos de promoción turística y desarrollo económico del país.

Trabajo Transdisciplinar y Evidencia de Colaboración:

Durante el desarrollo de la aplicación, colaboramos estrechamente con un grupo de estadística que proporcionó una evaluación cualitativa y cuantitativa esencial para refinar nuestra herramienta. Este equipo destacó la importancia de una interfaz simple y sugerencias especializadas para mejorar continuamente la usabilidad y funcionalidad de la aplicación. Específicamente, recomendó el uso de gráficos de pastel claros y directos con títulos concisos como "Probabilidad para cada calificación", y sugirió ajustes visuales como el tamaño de la leyenda y el color para mejorar la legibilidad y la interpretación intuitiva de los datos.

Además, el grupo ayudó a definir con precisión el perfil del usuario final, proporcionando ejemplos y escenarios que nos permitieron ajustar la aplicación a las necesidades específicas del personal del Ministerio de Comercio, Industria y Turismo, así como de otros actores gubernamentales. La inclusión de una ventana de recomendaciones basada en los análisis de datos, como mejoras en la limpieza y políticas de servicio, fue otra aportación significativa que fortaleció la utilidad de la aplicación.

Este enfoque colaborativo y transdisciplinar no solo enriqueció el producto final, sino que también fomentó un entorno de aprendizaje y desarrollo profesional, crucial para la innovación y la mejora continua en el sector turístico. La aplicación final es un reflejo del compromiso conjunto para optimizar la promoción y el desarrollo turístico en Colombia, garantizando una herramienta estratégica y eficaz para los usuarios finales.

3. Trabajo en equipo

Trabajamos con los siguientes roles en el proyecto para mejorar el desempeño de este:

- Ingeniero de software responsable de desarrollar la aplicación final: Eduardo Alba – Algoritmo: Ridge Classifier – 33 Puntos. 11 horas dedicadas.
 - Planteamiento inicial de limpieza y entendimiento de datos.
 - Implementación de la API por medio de FastAPI.
 - Creación de archivos de requerimientos.
 - Redacción inicial del README.
 - Acople del pipeline con la aplicación.
- Ingeniero de software responsable del diseño de la aplicación y resultados: Samuel Jimenez – 34 Puntos. 10 horas trabajadas.
 - Planteamiento de objetivos de negocio
 - Diseño gráfico de la aplicación
 - Implementación del front-end por medio de Node.
 - Redacción final del README.
 - Acople del back y front.
- Ingeniero de datos: David Cuevas Alba – 33 Puntos. 12 horas dedicadas.
 - Selección de palabras descriptivas.
 - Construcción del preprocesador.
 - Exportación de modelos.
 - Descripción del pipeline.

Repartimos los 100 puntos equitativamente, ya que consideramos que trabajamos a la par y todos respondimos por la entrega final, por lo que cada uno obtuvo 33 puntos.

Se realizaron las siguientes reuniones:

1. Reunión de lanzamiento y planeación: Para definir roles y forma de trabajo del grupo.
2. Reunión de finalización: Para consolidar el trabajo final, verificar el trabajo del grupo y analizar los puntos a mejorar para la siguiente etapa del proyecto

4. Resultados

Durante la segunda etapa de nuestro proyecto, recibimos valiosas opiniones y recomendaciones de expertos en el campo del turismo y la tecnología. A continuación, se resumen los principales hallazgos y conclusiones:

1. **Avance y Explicación de la Aplicación:** Los expertos elogiaron el avance destacado de nuestra aplicación y la capacidad del equipo para explicar de manera clara y concisa su funcionamiento y utilidad. Se destacó especialmente la habilidad del equipo para comunicar conceptos técnicos de manera accesible, lo que refleja una excelente comprensión del dominio y una sólida habilidad en el manejo del lenguaje natural.
2. **Diseño y Presentación de Datos:** Se discutió la posibilidad de mejorar el diseño para mostrar las probabilidades de error de la predicción por estrella. Los expertos sugirieron representar estas probabilidades en gráficas de barras, aunque se recomienda explorar también el uso de gráficas circulares para una comprensión más intuitiva. Asimismo, se recomendaron títulos intuitivos y el uso efectivo de colores para mejorar la presentación de los datos.
3. **Usuarios y Mapa de Actores:** Se revisaron y ajustaron los usuarios y el mapa de actores de la aplicación. Se destacó cómo nuestra aplicación puede ser útil para miembros comerciales de la industria turística y hotelera, permitiéndoles evaluar el impacto de cambios específicos en la opinión pública. También se sugirió considerar la inclusión de otros agentes gubernamentales, como los legisladores, como posibles usuarios de la aplicación.
4. **Terminología y Conceptos:** Se elogió el uso del término "beneficiarios" en lugar de "usuarios" al referirse a los viajeros, destacando así que estos son beneficiarios de las mejoras implementadas por los establecimientos turísticos.

5. **Recomendaciones Generales:** Se discutió la posibilidad de incluir recomendaciones generales para mejorar el servicio cada vez que se predice una calificación específica para una reseña. Se reconoció esta idea como valiosa, aunque se señaló la importancia de que las recomendaciones sean interpretadas como generales y no necesariamente aplicables a todos los usuarios.

En resumen, los resultados de nuestro proyecto reflejan un progreso significativo y una comprensión profunda de las necesidades y desafíos del sector turístico. Las recomendaciones de los expertos nos han proporcionado una guía invaluable para mejorar y refinar nuestra aplicación, asegurando que sea una herramienta efectiva para promover el turismo y mejorar la calidad de los servicios en Colombia.

Referencias

GeeksforGeeks. (s/f). GeeksforGeeks. Recuperado el 7 de abril de 2024, de <https://www.geeksforgeeks.org/>

Scikit-learn. (s/f). Scikit-learn.org. Recuperado el 7 de abril de 2024, de <https://scikit-learn.org/stable/>

fastapi. (2024, April 19). PyPI. <https://pypi.org/project/fastapi/>

Index | Node.js v21.7.3 Documentation. (n.d.). <https://nodejs.org/docs/latest/api/>