



INTELIGENCIA DE NEGOCIOS

ISIS - 3301

PROYECTO 1 – ETAPA 2

POR:

DAVID ALEJANDRO FUQUEN

CÓDIGO: 202021113

VALERIA CARO RAMIREZ

CÓDIGO: 202111040

MARIANA RUIZ GIRALDO

CÓDIGO: 202011140

UNIVERSIDAD DE LOS ANDES

FACULTAD DE INGENIERÍA

SEPTIEMBRE DE 2024

Tabla de Contenido

1. Automatización del proceso de preparación de datos, construcción del modelo, persistencia del modelo y acceso por medio de API	2
2. Desarrollo de la aplicación	3
3. Resultados	4
4. Trabajo en equipo	5
5. Bibliografía	6

1. Automatización del proceso de preparación de datos, construcción del modelo, persistencia del modelo y acceso por medio de API

En esta etapa 2 del proyecto se completó el proceso iniciado en la etapa 1 para la automatización para la preparación de los datos, la persistencia del modelo y acceso al mismo por medio de una API REST utilizando el framework FastAPI. En particular, se creó una aplicación web en la cual se puede reentrenar el modelo y realizar predicciones sobre frases o archivos que introduzca el usuario.

Este flujo se gestiona a través de un pipeline que permite realizar la preparación de los datos y el entrenamiento del modelo, el cual nace a partir de lo realizado en la etapa 1 del proyecto. Tal como se mencionó en la etapa anterior, el modelo de clasificación elegida para el pipeline fue una Regresión Logística, que fue elegida por sus excelentes métricas sobre el conjunto de entrenamiento y testeo. Para el backend se utilizó FastAPI por su simplicidad y eficiencia en el desarrollo de APIs. En este, utilizando el pipeline realizado, se crearon 5 endpoints para cumplir con las dos funcionalidades especificadas en el enunciado:

- /uploadfile: permite subir un archivo en formato CSV o Excel para que el sistema procese los datos contenidos en dicho archivo. Este endpoint asegura que solo se acepten archivos con los formatos correctos y lee los datos, mostrando una muestra de cinco filas como verificación inicial.
- /retrain: está diseñado para permitir el reentrenamiento del modelo. El usuario puede subir un archivo de datos que incluya tanto las características como la variable objetivo. Estos datos reemplazan completamente los datos originales del modelo, con el objetivo de que el modelo se ajuste mejor a las características actuales de la población, en caso de que estas hayan cambiado considerablemente. En este endpoint, se asume que el usuario proveerá suficientes datos para un reentrenamiento efectivo. Al finalizar el proceso, se devuelven métricas clave como accuracy, precision, recall y f1_score, lo que permite evaluar la calidad del nuevo modelo entrenado.

- `/predict`: permite al usuario enviar una o más instancias de datos a través del cuerpo de la solicitud, las cuales serán procesadas por el modelo para devolver una predicción. Las predicciones se presentan en una lista, manteniendo el mismo orden de las instancias enviadas, y junto a la predicción se incluyen las probabilidades correspondientes para cada clase. Este endpoint asegura que los usuarios puedan realizar predicciones en tiempo real, ingresando datos manualmente o subiendo archivos.
- `/predict_from_excel`: permite subir archivos en formato CSV o Excel con los datos para predicción masiva. Este endpoint carga el modelo previamente entrenado, procesa el archivo y devuelve una lista con las predicciones y las probabilidades, manteniendo la estructura del archivo original.
- `/download-file` recibe los datos de predicciones y los devuelve en el formato solicitado, haciendo la aplicación más versátil y adaptable a distintos casos de uso.

Por otro lado, para el frontend se utilizó Next.js facilitando así la creación de una aplicación web dinámica y amigable para los usuarios.

2. Desarrollo de la aplicación

2.1. Descripción del usuario/rol de la organización que va a utilizar la aplicación

- **Analistas de Datos y Científicos de Datos:** Su función principal es procesar, analizar y derivar insights de grandes volúmenes de datos. Gracias a la aplicación, podrán automatizar el proceso de análisis de opiniones ciudadanas y así reducir la carga de trabajo, permitiéndoles enfocarse en tareas estratégicas y en el desarrollo de otros modelos y actividades menos manuales.
- **Políticos y Planificadores de Políticas Públicas:** Encargados de tomar decisiones basadas en datos para diseñar y aplicar políticas que respondan a las necesidades actuales de la población. La aplicación les proporciona información relevante y oportuna sobre las áreas de intervención en los ODS 3, 4 y 5, apoyando su toma de decisiones.
- **Expertos en Desarrollo Sostenible:** Involucrados en evaluar y priorizar áreas de intervención que ayuden a cumplir los Objetivos de Desarrollo Sostenible. La aplicación les permite identificar rápidamente las problemáticas prioritarias y las necesidades de los ciudadanos según las opiniones que han dado, facilitando su trabajo en la planificación de intervenciones al identificar a cuál ODS (3,4 o 5) pertenece cada uno de los textos introducidos.

2.2. Conexión entre la aplicación y el proceso de negocio que se va a aplicar

La aplicación está directamente conectada al proceso de recopilación y análisis de opiniones ciudadanas en relación con los ODS. Dentro del proceso de negocio, podemos ver cómo la aplicación impacta en tres aspectos en particular:

1. **Recopilación de Opiniones Ciudadanas:** La aplicación recibe datos textuales de diferentes fuentes, lo que incluye encuestas, foros de participación y redes sociales. Estos datos se integran automáticamente en la aplicación para ser procesados.
2. **Clasificación y Análisis Automatizado:** Una vez que se recopilan las opiniones, la aplicación utiliza técnicas de aprendizaje supervisado para clasificar y relacionar el contenido con los ODS 3, 4 y 5. El modelo analítico entrenado identifica patrones a cual de los ODS es más probable que pertenezca el texto.
3. **Generación de Reportes e Insights:** Finalmente, la aplicación genera reportes que destacan los temas prioritarios y las áreas de intervención necesarias según cada uno de los textos, proporcionando insights a los políticos y planificadores. Estos insights son vitales para el diseño y ajuste de políticas públicas.

2.3. Importancia para el rol de la existencia de la aplicación

- **Eficiencia y Reducción de Carga de Trabajo para los Analistas de Datos y Científicos de Datos:** La automatización del proceso de análisis de datos permite reducir el tiempo y los recursos empleados en la interpretación manual de grandes volúmenes de opiniones.
- **Toma de Decisiones Basada en Datos para los Políticos y Planificadores de Políticas Públicas:** Para los políticos y planificadores, la aplicación proporciona datos precisos y actualizados sobre las necesidades ciudadanas en los ODS, permitiéndoles diseñar políticas efectivas que respondan a la realidad del momento.
- **Facilitación del Trabajo de los Expertos en Desarrollo Sostenible:** Estos expertos pueden identificar áreas prioritarias de intervención de forma más rápida y con mayor exactitud.

3. Resultados

3.1. Pruebas de facilidad de uso de la aplicación

Nuestra página ha sido diseñada pensando en la accesibilidad para garantizar una experiencia inclusiva para todos los usuarios. Uno de los aspectos destacados por nuestros usuarios fue la paleta de colores, la cual ha sido seleccionada cuidadosamente para ser agradable a la vista y facilitar la navegación. En términos de usabilidad, nos hemos asegurado de que el sitio sea completamente responsive, adaptándose a distintos dispositivos sin comprometer la funcionalidad. Además, nos hemos esforzado en ofrecer diversas maneras para que los usuarios puedan subir

datos y así facilitar el proceso de predicción. La página también cuenta con un sistema que notifica al usuario cuando los datos se están cargando, y, en caso de error al subir un archivo, hemos habilitado la opción para corregirlo fácilmente sin necesidad de reiniciar el proceso.

3.2. Link a la aplicación web
<https://proyectobi.pukini.dev/>

4. Trabajo en equipo

4.1. Roles dentro del equipo

- **Líder de proyecto:** Mariana Ruiz Giraldo, se encargó de pactar las reuniones semanales, de definir cada una de las tareas y revisar el trabajo de cada uno de los integrantes con el fin de mantener coherencia en el proyecto.
- **Ingeniera de datos:** Valeria Caro, encargada de velar por la calidad del proceso de automatización relacionado con la construcción del modelo analítico.
- **Ingeniero de software responsable del diseño de la aplicación y resultados:** David Fuquen, encargado de liderar el diseño de la aplicación y de la generación del video con los resultados obtenidos.
- **Ingeniero de software responsable de desarrollar la aplicación final:** Valeria Caro, encargada de gestionar el proceso de construcción de la aplicación.

4.2. Reuniones realizadas

- Reunión de lanzamiento y planeación: 28/09/2024. En esta reunión se definieron roles y se definieron tareas a realizar por cada integrante.
- Reunión de seguimiento 1: 01/10/2024. En esta reunión se mostró un primer avance del proyecto. Se definió la estructura de la aplicación y se realizó el pipeline para el modelo realizado en la etapa 1.
- Reunión de seguimiento 2: 05/10/2024. En esta reunión se mostró el segundo avance del proyecto. Se mostró el backend casi finalizado y un primer borrador del frontend.
- Reunión de seguimiento 3: 08/10/2024. En esta reunión se mostró el tercer avance del proyecto. Se mostró el backend finalizado y funcional junto al frontend pronto a terminarse. Asimismo, se realizó la presentación a mostrar en le video.
- Reunión para grabación del video: 11/10/2024. En esta reunión se mostró la versión final de la aplicación. Tras una exhaustiva revisión del código, se realizó la grabación del video.
- Reunión de finalización: 12/10/2024. En esta reunión se consolidó el trabajo final, se terminó la realización de este documento, se verificó el trabajo del grupo y se analizaron los puntos a mejorar. Hecho esto, se realizó el envío del trabajo.

4.3. Distribución de tiempos y puntos

- Mariana Ruiz: 14 horas de trabajo, utilizadas en la realización del endpoint de reentrenamiento del modelo y modelaje del frontend. De 100 puntos repartidos, 33.33 serían dados a este integrante.
- Valeria Caro: 14 horas de trabajo, utilizadas en la realización del pipeline para adaptar el modelo de la etapa anterior y la realización del endpoint de predicción de archivos. De 100 puntos repartidos, 33.33 serían dados a este integrante.
- David Fuquen: 14 horas de trabajo utilizadas en la realización del endpoint de predicción de frases insertadas por el usuario y la realización del video. De 100 puntos repartidos, 33.33 serían dados a este integrante.

5. Bibliografía

United Nations. (2023). *Objetivos de desarrollo sostenible*. Obtenido de <https://www.un.org/sustainabledevelopment/es/objetivos-de-desarrollo-sostenible/#>