



Inteligencia de Negocios
Proyecto: Parte 2

Grupo:
Carlos Julio Pinto Rodríguez – 201616667
Juan Camilo González – 201911030
Grupo Estadística:
Letitia Sofia Russi Bello –

2023

Tabla de Contenido

Proceso de automatización del proceso de preparación de datos, construcción del modelo, persistencia del modelo y acceso por medio de API	2
Desarrollo de la aplicación y justificación.....	4
Interacción con el grupo de estadística	6
Trabajo en equipo	7

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

La fase inicial del despliegue de la solución analítica aborda la automatización integral del proceso, enfocándose en la preparación de datos, construcción del modelo y su persistencia. El ingeniero de datos implementa pipelines que orquestan de manera eficiente las tareas de limpieza y transformación de los textos en español, aplicando técnicas avanzadas de procesamiento de lenguaje natural que se ha realizado el proyecto, teniendo en cuenta los procesos de construcción de modelo realizado en la primera parte del proyecto. La eliminación de caracteres especiales, la lematización y la supresión de stopwords son algunas de las operaciones clave que aseguran la calidad de los datos de entrada al modelo. En esta fase, se puede observar en la imagen 1, donde se construye la automatización en el pipeline, cabe mencionar que el objetivo de crear un pipeline es automatizar todos los pasos realizados sobre los datos. Desde que salen de su fuente hasta que son ingresados al modelo de aprendizaje automático. Por ello, utilizamos la librería Scikit Learn ya que cuenta con una API para la creación de pipelines en la que pueden ser utilizados diferentes pasos para la transformación de los datos que serán aplicados secuencialmente.

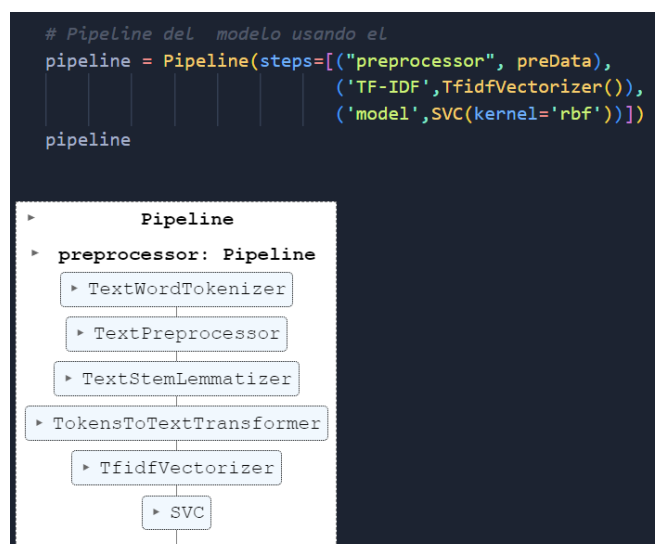


Imagen 1. Proceso de automatización involucrados en el pipeline

La construcción del modelo se lleva a cabo mediante técnicas de aprendizaje automático, utilizando algoritmos de clasificación de textos. Durante este proceso, el modelo previamente entrenado es cargado y configurado para interpretar y clasificar

textos relacionados con los Objetivos de Desarrollo Sostenible (ODS) de la Agenda 2030 de las Naciones Unidas. Esta etapa es crucial para asegurar que el modelo esté ajustado a las particularidades de los datos y sea capaz de realizar predicciones precisas. Por otro lado, la persistencia del modelo analítico es esencial para su disponibilidad continua y uso eficiente en el entorno de producción. Se exporta el modelo entrenado en un formato compatible con su reutilización, optimizando su despliegue. La persistencia asegura que el proceso de construcción del modelo no sea necesario en cada interacción, optimizando recursos y acelerando la respuesta del sistema. En este proceso, el pipeline construido donde están encapsulados los diferentes procesos requeridos para la preparación de datos, el entrenamiento del modelo y la predicción como proceso final del mismo, son exportados a un archivo binario o aun archivo ".joblib" utilizando la librería Joblib para posteriormente ser utilizado dentro de la API REST mediante peticiones HTTP.

A continuación, se implementa una API utilizando el framework FastAPI. Esta interfaz proporciona un punto de entrada para la interacción con el modelo clasificador. El ingeniero de datos configura la API para recibir solicitudes que contengan textos en español y devuelve predicciones correspondientes a los ODS. La API actúa como un puente eficaz entre el modelo analítico y los usuarios finales, facilitando la accesibilidad y la integración en diversas aplicaciones.

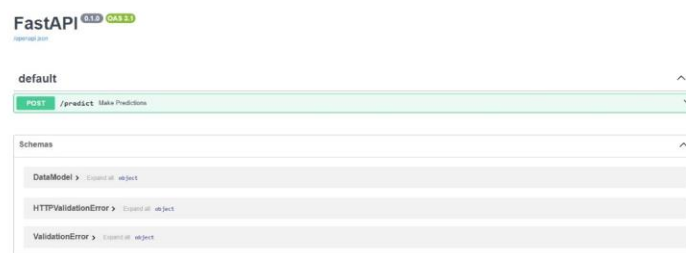


Imagen 2. Creación de API y verificación de ejecución del framework.

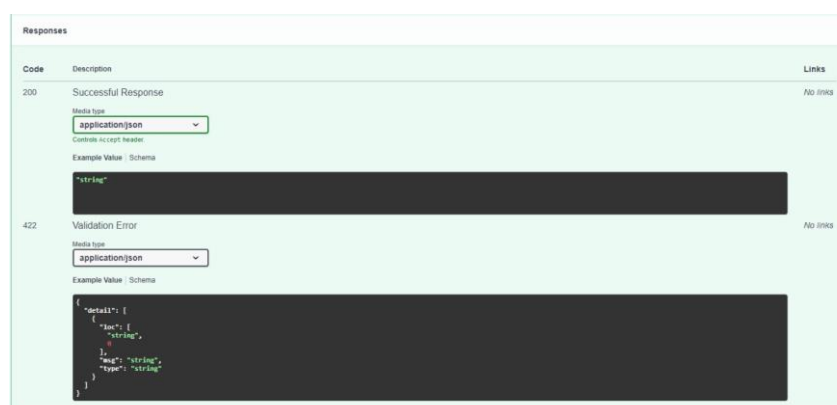


Imagen 3. Pruebas de ejecución de la creación de API y respuesta de modelo predictivo.

Dentro del desarrollo del framework API, FastAPI se destaca por su enfoque eficiente que puede brindar una infraestructura sólida para la interacción entre la aplicación web y el modelo predictivo. También, la capacidad de FastAPI para generar documentación interactiva automáticamente que es fundamental para el desarrollo de

la aplicación, este aspecto es beneficioso para la colaboración transdisciplinaria, ya que los expertos en estadísticas y otros usuarios pueden entender fácilmente la funcionalidad y los requisitos de la API, incluso sin un conocimiento profundo técnico. Creamos un API del framework basado en Python, dando instrucciones simples para la creación de rutas y la definición de parámetros, reduciendo la complejidad del código y acelerando el desarrollo. Por consiguiente, la clase “DataModel” permite la fácil especificación de los datos de entrada esperados, que puede facilitar así la validación y el manejo de errores en las solicitudes entrantes, destacando la utilización de la clase “BaseModel” de Pydantic.

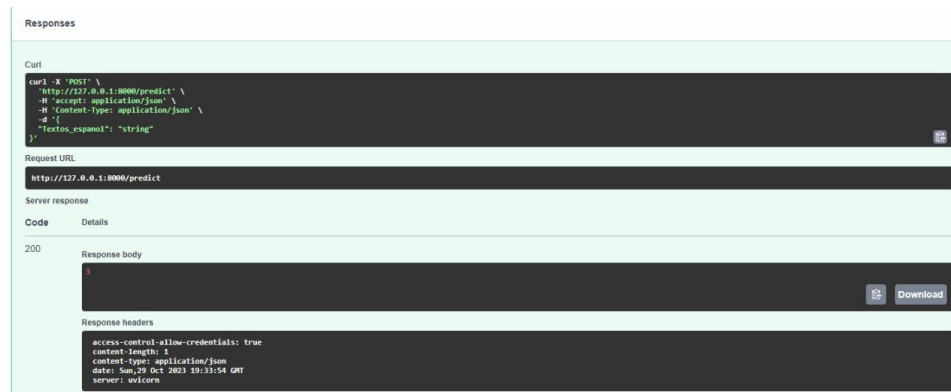


Imagen 4. Funcionamiento y respuesta de prueba a la API con el modelo entrenado de predicción.

Desarrollo de la aplicación y justificación

La aplicación desarrollada (Pagina web) se concibe o es utilizada como una herramienta clave para los actores estratégicos de la organización, en particular, para aquellos involucrados en la planificación y evaluación de políticas públicas relacionadas con los Objetivos de Desarrollo Sostenible (ODS) de la Agenda 2030 de las Naciones Unidas. Los usuarios principales son los analistas y planificadores del Fondo de Poblaciones de las Naciones Unidas (UNFPA), así como entidades públicas y colaboradores de la Universidad de los Andes. Por lo cual, la aplicación se puede percibir como un recurso valioso para estos actores, brindando un análisis automatizado de opiniones recopiladas a través de procesos de planeación participativa para el desarrollo a nivel territorial.

La conexión intrínseca entre la aplicación y el proceso de negocio se manifiesta o se ve reflejada en la capacidad de la herramienta para interpretar y clasificar textos según los ODS. Esta funcionalidad se integra perfectamente con las actividades de seguimiento y evaluación de políticas públicas que lleva a cabo el UNFPA, permitiendo una identificación más rápida y precisa de problemáticas locales, así como de soluciones. La aplicación se convierte en sí como un componente estratégico para la toma de decisiones basadas en datos, respaldando la misión de la organización de promover el desarrollo sostenible y útil para solucionar la problemática que es el objetivo del proyecto. (Relación aplicación y proceso de negocio)

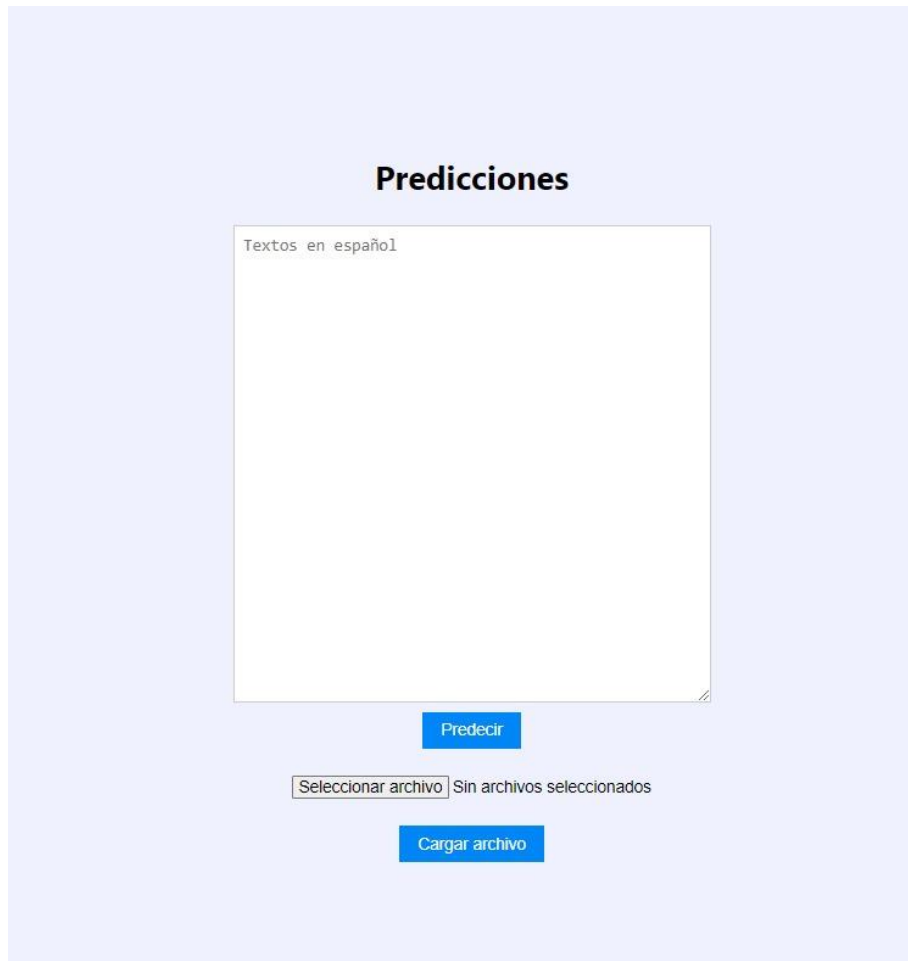
Entonces, el usuario final de la aplicación es muy importante para la aplicación ya que es guiado a través de una interfaz amigable que facilita la introducción de textos en

español relacionados con cuestiones territoriales. Cabe destacar que la importancia de esta aplicación (Pagina web) radica en su capacidad para traducir automáticamente estos textos en insights significativos, clasificándolos según los ODS. Los usuarios pueden introducir de manera sencilla textos en español relacionados con cuestiones territoriales, y la aplicación, mediante el modelo predictivo, clasifica estos textos según los ODS correspondientes. Esto proporciona a los usuarios una visión instantánea y estructurada de las preocupaciones locales, empoderándolos con información detallada y relevante para la toma de decisiones.

En el contexto de esta era globalizada y digitalizada en la que vivimos, el análisis de datos juega un papel crucial, pues ofrece una ventaja fundamental para las organizaciones públicas y privadas a la hora de planificar estrategias y desarrollarlas, aumentando las probabilidades de éxito y ayudando a tomar decisiones basadas en la realidad social en la que se busca generar un impacto. Además de permitir tomar decisiones actualizadas, el análisis de datos proporciona una visión más amplia al posibilitar hacer predicciones a futuro, de forma que las estrategias son planificadas y llevadas a cabo de forma más eficiente y productiva. Sin embargo, esta tarea se dificulta cuando se trata de grandes volúmenes de datos.

Para manejar un gran volumen de datos se requiere que sean previamente organizados y estructurados para ser procesados posteriormente, y a partir de esto poder sacar provecho de la información que nos proveen. Asimismo, es importante destacar que la flexibilidad a la hora de interpretar los datos es un factor esencial, pues es un proceso que varía en función de la necesidad del usuario.

La tecnología es una herramienta fundamental en la labor de análisis de datos y aún más cuando se trata de grandes volúmenes; esta aplicación representa una solución eficiente y confiable, diseñada para cumplir con los requerimientos establecidos en el análisis de datos cualitativos, filtrando y procesando la información de una forma segura y con gran precisión. En una era en la que los datos y la inmediatez son factores decisivos a la hora de tomar decisiones actualizadas y basadas en la realidad, esta aplicación tiene un rol central ya que contribuye en gran medida a aprovechar el verdadero potencial que tienen los datos.



The image shows a web application interface with a light blue background. At the top center, the title "Predicciones" is displayed in bold black text. Below the title is a large, empty white rectangular box with a thin border, containing the placeholder text "Textos en español" at the top left. Underneath this box is a blue button with the text "Predecir" in white. Below the button is a file selection interface consisting of a small rectangular box with the text "Seleccionar archivo" and the text "Sin archivos seleccionados" to its right. At the bottom of this interface is a blue button with the text "Cargar archivo" in white.

Imagen 4. Construcción de la aplicación web para el modelo predictivo.

En el desarrollo de la aplicación web proporciona un medio interactivo para que los usuarios exploren y comprendan el resultado del modelo. Al permitir la introducción de texto por parte del usuario y recibir predicciones instantáneas, la aplicación promueve la retroalimentación y mejora continua. Aunque en este punto no se ha recibido un apoyo directo del grupo de estadística, la colaboración anterior influyó en la definición de métricas y criterios de evaluación, asegurando la calidad del modelo y, por ende, la aplicación.

Interacción con el grupo de estadística

La definición de la aplicación y su enfoque centrado en el usuario se gestó mediante un proceso colaborativo. La transdisciplinariedad fue un aspecto clave, donde ingenieros y estadísticos cooperaron para transmitir e integrar sus conocimientos. Los ingenieros actuaron como desarrolladores de la aplicación con el objetivo de satisfacer las necesidades en materia de análisis de datos cualitativos, y los estadísticos desde el rol de usuario/analista de datos, testearon la aplicación y evidenciaron el proceso de desarrollo para determinar su funcionalidad, cumplimiento de las métricas y eficacia, brindando retroalimentación sobre el procesamiento de los

datos como base para el uso de métodos estadísticos posteriores, necesarios para interpretar la información procesada por la aplicación.

Trabajo en equipo

Roles para el proyecto:

- **Líder de proyecto (Juan Camilo Gonzalez):** Está a cargo de la gestión del proyecto. Define las fechas de reuniones, pre-entregables del grupo y verifica las asignaciones de tareas para que la carga sea equitativa. Se encarga de subir la entrega del grupo. Si no hay consenso sobre algunas decisiones, tiene la última palabra.
- **Líder de estadística (Letitia):** Está a cargo del grupo con el que está trabajando de manera transdisciplinar para mejorar el modelo y conseguir mejorar la experiencia usuario-cliente.
- **Ingeniero de datos (Carlos):** Es responsable de velar por la calidad del proceso de automatización relacionada con la construcción del modelo analítico.
- **Ingeniero de software responsable del diseño de la aplicación y resultados (Carlos):** Se encarga 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 (Juan Camilo Gonzalez):** Se encarga de gestionar el proceso de construcción de la aplicación

Se realizaron las siguientes reuniones de grupo:

- **Reunión de lanzamiento y planeación:** Se realizó esta reunión el miércoles 25 de octubre de 2023. La reunión, con duración de 40 minutos, consistió en definir los roles, la forma de trabajo del grupo y la interacción con el grupo de estadística. Se acuerdan las fechas de reuniones y pre-entregables del grupo, además de los métodos de comunicación para el desarrollo del proyecto.
- **Reunión con el grupo de estadística:** Se realizó esta reunión el jueves 26 de octubre de 2023. La reunión, con duración de 30 minutos, consistió en presentar el problema y solución al grupo de estadística y definir como apoyarían este proyecto.
- **Reuniones de seguimiento:** Se realizaron diferentes informes de seguimiento del proyecto mediante mensajes de texto. En estos mensajes se informaba de cada una de las tareas que realizaba cada uno de los integrantes del grupo. Sirven para mantener un registro constante del avance del proyecto y para asegurarse de que las tareas se completen de manera eficiente.
- **Reunión de finalización:** Se realizó esta reunión el domingo 29 de octubre de 2023. La reunión, con duración de 2 horas, consistió en una verificación del trabajo realizado por cada uno de los estudiantes en el proyecto, además se discutieron los posibles puntos de mejora para el siguiente proyecto. En la reunión, se consolidó todo el trabajo realizado. Una vez que se llegó a un consenso sobre la completitud y la satisfacción con los resultados, se realizó la entrega del proyecto.

