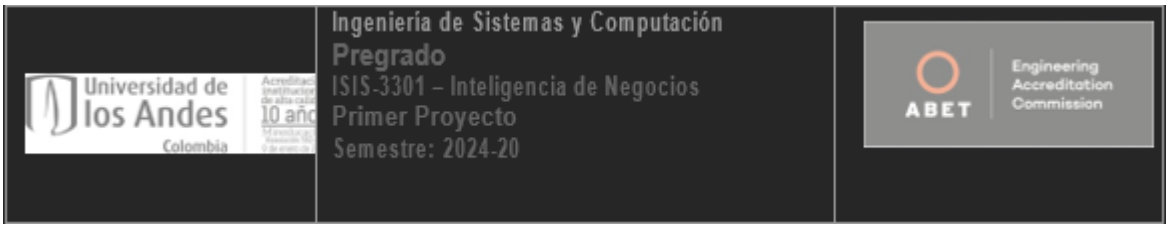


Proyecto 1 – Etapa 2

Integrantes:

- Alejandro González Salazar – 201921465
- Daniel Gómez Rey - 202122586
- María Alejandra Lizarazo – 202021385

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

El proceso de automatización incluye la creación del pipeline y el uso del modelo seleccionado en la etapa anterior. Para ello, utilizamos la clase textProcessor.py que implementa el método necesario para la limpieza de los datos. Este método ya está integrado en nuestro pipeline, y realiza tareas como:

- Eliminar caracteres no deseados (letras, símbolos).
- Eliminar tildes y caracteres diacríticos.
- Normalizar palabras y reducirlas a sus raíces.

El modelo utilizado es Naive Bayes, seleccionado por su rendimiento en la etapa anterior. Posteriormente, el pipeline se guarda utilizando joblib para su reutilización en la API. (Todo el código se encuentra en el notebook llamado Proyecto1Parte2.ipynb)

Los endpoints proporcionados en el código funcionan correctamente. Dentro de la aplicación se muestran tanto los resultados del modelo como las diferentes probabilidades de todas las predicciones. Adicionalmente, el segundo endpoint puede recibir un archivo XLSX o un CSV, el cual realiza un entrenamiento que combina los nuevos datos con los antiguos, volviendo a entrenar el modelo y retornando las nuevas métricas de calidad. Esta opción fue elegida porque permite al modelo conservar el conocimiento adquirido de los datos anteriores, mientras se adapta a posibles nuevos patrones. Esto ayuda a mejorar la generalización del modelo, evitando así olvidar características importantes.

Sección 2. Desarrollo de la aplicación y justificación

La aplicación será utilizada por el Fondo de Poblaciones de las Naciones Unidas (UNFPA) en Colombia específicamente por analistas y gestores de proyectos que trabajan en el ámbito del desarrollo social y la gestión de programas relacionados con la equidad de género, la salud reproductiva y el bienestar de las poblaciones vulnerables. La aplicación está diseñada para apoyar el proceso de monitoreo y evaluación de proyectos del UNFPA. Al proporcionar capacidades de clasificación y análisis de textos, permite a los analistas identificar rápidamente temas emergentes en las opiniones y comentarios de la población. Esto facilita la toma de decisiones informadas y la adaptación de estrategias y programas a las realidades de la comunidad colombiana.

Los profesionales necesitan herramientas que les permitan analizar y clasificar opiniones y comentarios de la comunidad, para comprender mejor las necesidades y preocupaciones de las personas. La existencia de esta aplicación es crucial, ya que les proporciona una herramienta eficiente y accesible para extraer insights valiosos de datos

El desarrollo de la API está enfocado en proporcionar predicciones basadas en el modelo entrenado. La API, ubicada en api.py, tiene la siguiente estructura:

Estructura de la API

- POST("/predict"): Recibe los datos en formato JSON para predecir la clase correspondiente (calificación de la reseña) y devuelve la predicción junto con la probabilidad. La entrada es un texto en lenguaje natural, que se procesa utilizando el pipeline previamente mencionado.
- POST("/retrain"): Este endpoint permite el reentrenamiento del modelo utilizando un nuevo conjunto de datos. La entrada incluye las características y la variable objetivo. Devuelve métricas de desempeño como Precision, Recall y F1-score.

Justificación

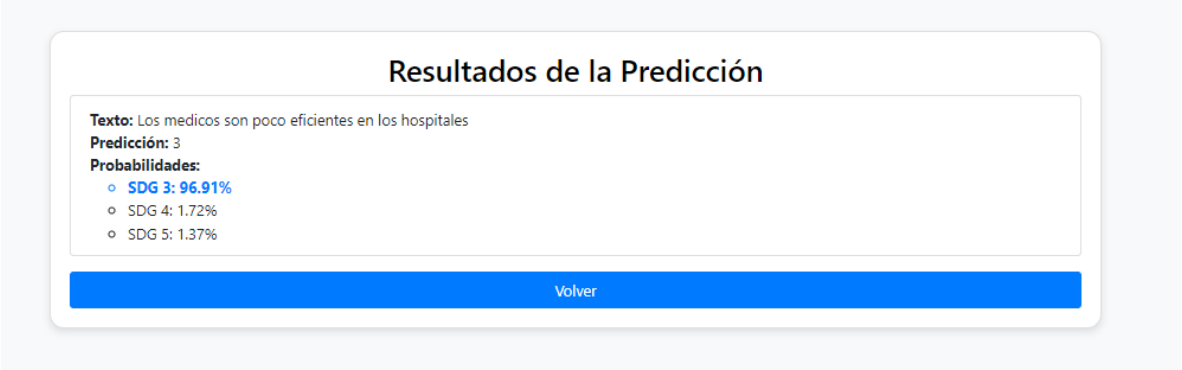
La aplicación fue diseñada pensando en los objetivos del Fondo de Poblaciones de las Naciones Unidas (UNFPA) y su necesidad de analizar y clasificar opiniones de los ciudadanos con relación a los Objetivos de Desarrollo Sostenible (ODS). La API permite automatizar este proceso mediante un modelo que predice la relevancia de las opiniones en relación con los ODS 3, 4 y 5.

Sección 3. Resultados

Los resultados obtenidos en la aplicación están explicados en el video publicado en el padlet, a continuación, se mostrarán las imágenes de las distintas funcionalidades de la aplicación web:

Página principal:

Clasificación de 1 o más textos:

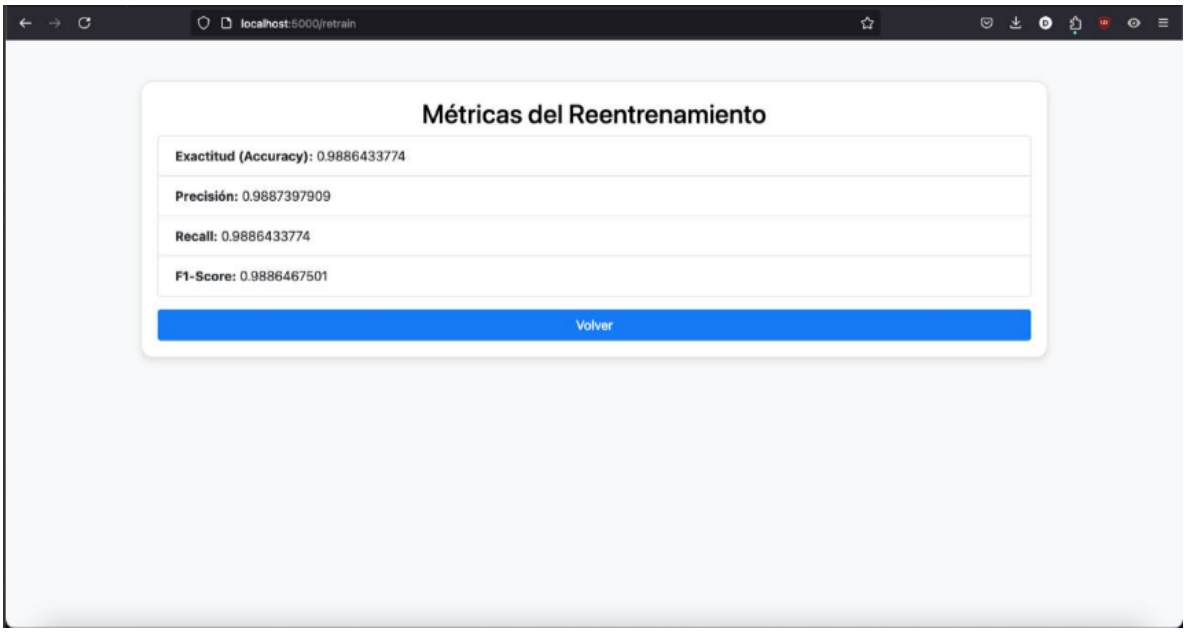


Clasificación de un archivo:

Predicciones de Opiniones desde el Archivo

Texto	Predicción	Probabilidades
Han examinado la contribuciÃ³n de las universidades y las instituciones de educaciÃ³n terciaria al desarrollo del capital humano y las competencias, la transferencia de tecnologÃa y la innovaciÃ³n empresarial, el desarrollo social, cultural y medioambiental, y la creaciÃ³n de capacidad regional. El proceso de revisiÃ³n ha facilitado la creaciÃ³n de asociaciones en ciudades y regiones al reunir a instituciones de educaciÃ³n terciaria y organismos pÃºblicos y privados para identificar objetivos estratÃ©gicos y trabajar juntos para alcanzarlos. Complementa las revisiones que se han llevado a cabo en la región transfronteriza México-Estados Unidos, de gran importancia estratégica y económica, como las de Nuevo León, la región Paso del Norte y, más recientemente, la del sur de Arizona (2011), junto al estado de Sonora.	4	• SDG 3: 5.03% • **SDG 4: 65.51%** • SDG 5: 29.46%
En la última década, y en particular desde 2010, el número de altas per cápita ha tendido a disminuir, y ahora está en línea con la media de la OCDE mencionada anteriormente. Por ejemplo, los habitantes de zonas rurales representan el 43% de la población, pero sólo el 32% de las altas hospitalarias. A falta de información complementaria, no es posible determinar si esta diferencia indica o no disparidades en el acceso de las poblaciones rurales y, en caso afirmativo, si las causas son geográficas, culturales o infraestructurales. No parece que se recojan datos sobre el acceso a la atención desde el punto de vista del paciente, lo que ayudaría a comprender la naturaleza de los obstáculos al acceso (y las soluciones). En general, no es posible determinar con los datos disponibles si los ingresos o las características socioeconómicas de los pacientes influyen a la hora de llegar a las puertas del hospital o de ser ingresado. El acceso desigual probablemente se deba también a la incapacidad de los centros de las distintas regiones para prestar los mismos servicios.	3	• **SDG 3: 84.83%** • SDG 4: 4.94% • SDG 5: 10.24%
¿En qué países los estudiantes de alto rendimiento se sienten atraídos por la docencia? Las encuestas de docentes en servicio a menudo muestran que los maestros actuales están muy motivados por los beneficios intrínsecos de la enseñanza: trabajar con niños y ayudarlos a desarrollarse y hacer una contribución a la sociedad, mientras que los estudios que encuestan a grandes grupos de graduados sobre sus opciones de carrera muestran que la los salarios relativos de las ocupaciones de los graduados juegan un papel en sus elecciones: si los salarios de los docentes hubieran sido más altos, más "docentes potenciales" habrían considerado seriamente una carrera en la enseñanza. A nivel de país, los resultados indican que tanto los salarios de los docentes como la estatus social de la profesión docente se asocian positivamente con las expectativas de los estudiantes de trabajar como docentes.	4	• SDG 3: 0.15% • **SDG 4: 98.34%** • SDG 5: 1.51%
A raíz de su preocupación por el hecho de que los médicos de todo el sistema sanitario japonés no fueran capaces de identificar los signos de sufrimiento psicológico que podrían ayudar a reducir los suicidios, la Asociación Médica Japonesa informa de que empezó a compartir información y orientaciones sobre la depresión con todos los médicos, primero en 2004 y luego de nuevo en 2009. En la actualidad, sólo se recogen indicadores sobre la tasa de reclusión y contención, y de ingresos involuntarios, por parte de los proveedores y a nivel prefectural. La información recopilada sistemáticamente sobre el sistema de atención de salud mental se limita a indicadores estructurales: instalaciones, número de personal, número de camas.	3	• **SDG 3: 98.35%** • SDG 4: 0.89% • SDG 5: 0.76%

Reentrenamiento del modelo:



Sección 4. Trabajo en equipo

Todos los integrantes del equipo participaron activamente en todas las fases del desarrollo de esta etapa del proyecto, tomando roles distintos, pero igualmente importantes para cumplir con los objetivos establecidos. En total, cada integrante dedicó entre 6 y 10 horas de trabajo para desarrollar las tareas asignadas. El equipo decidió que los puntos de la calificación se repartieran equitativamente entre los tres miembros del equipo, ya que todos aportaron de manera similar a las tareas principales.

Nombre	Tareas	Descripción
Alejandro Gonzalez	Ingeniero de software responsable del diseño de la aplicación y responsable de desarrollar la aplicación final	Fue el principal encargado de la realización del frontend del API, además contribuyó en el backend para realizar los respectivos ajustes para verificar el correcto funcionamiento
Maria Alejandra Lizarazo	Lider del proyecto e ingeniera de datos	Fue la encargada de establecer los pasos a seguir para la realización de esta etapa del proyecto, adicionalmente, realizó el Pipeline, construcción del modelo analítico y supervisión de la realización de la aplicación final
Daniel Gomez Rey	Ingeniero de software responsable del diseño de la aplicación y responsable de desarrollar la aplicación final	Fue el principal encargado de la realización del backend del API, además contribuyó en el frontend para mostrar algunos resultados

Criterios de evaluación

Siguiendo los criterios de evaluación establecidos en la matriz de evaluación del curso, todos los integrantes completaron las tareas indicadas en el enunciado, presentando documentación adecuada y detallada del proceso. Cada uno participó activamente en los roles descritos, contribuyendo de manera significativa a las tareas técnicas y de documentación, asegurando que los estándares de calidad del proyecto fueran cumplidos.