



Inteligencia de
Negocios

**Automatización de analítica de
Textos - Parte 2**

Objetivos de Desarrollo Sostenible

Antonin Bouillaud
Ernesto Duarte
Lina Gómez

Fecha:
Octubre 29, 2023



Contenidos

01

**Recapitulación -
Entendimiento de
Negocio**

- ODS Elegidos
- Objetivo Principal

02

**Proceso de
automatización**

- Preparación de datos
- Modelo y API

03

**Desarrollo de la
aplicación**

- Transformaciones
- Algoritmos
Seleccionados

04

Resultados

- Visualización en la
aplicación

Parte 1:

Entendimiento del Negocio

Objetivos del
Desarrollo
Sostenible

Fecha:
Octubre 29, 2023



Objetivo Principal

Implementar un modelo de **clasificación de textos** utilizando técnicas de procesamiento de lenguaje natural y **aprendizaje supervisado** para entrenar y evaluar el modelo con un conjunto de datos de textos etiquetados con los ODS 3, 4 y 5.



Parte 2:

Proceso de Automatización

Sustainable
Development Goals

Date:
Octubre 29, 2023

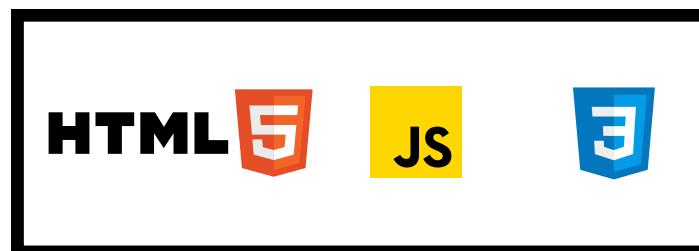
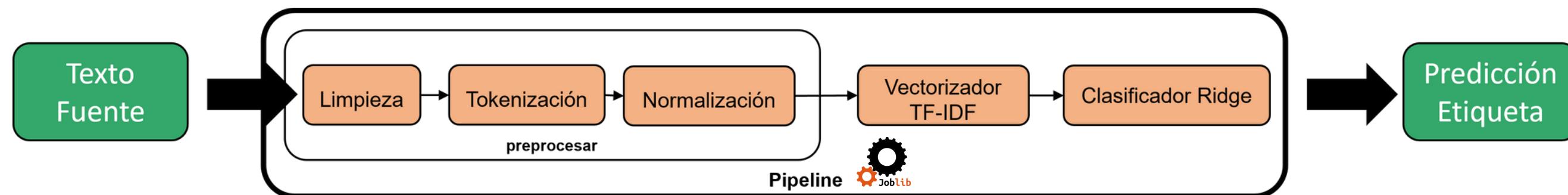




Inteligencia de
Negocios

Pipeline

Uso de **FastAPI** como framework para crear una API web que recibe solicitudes de comentarios de usuarios y devuelve predicciones de clasificación.



Front



Backend



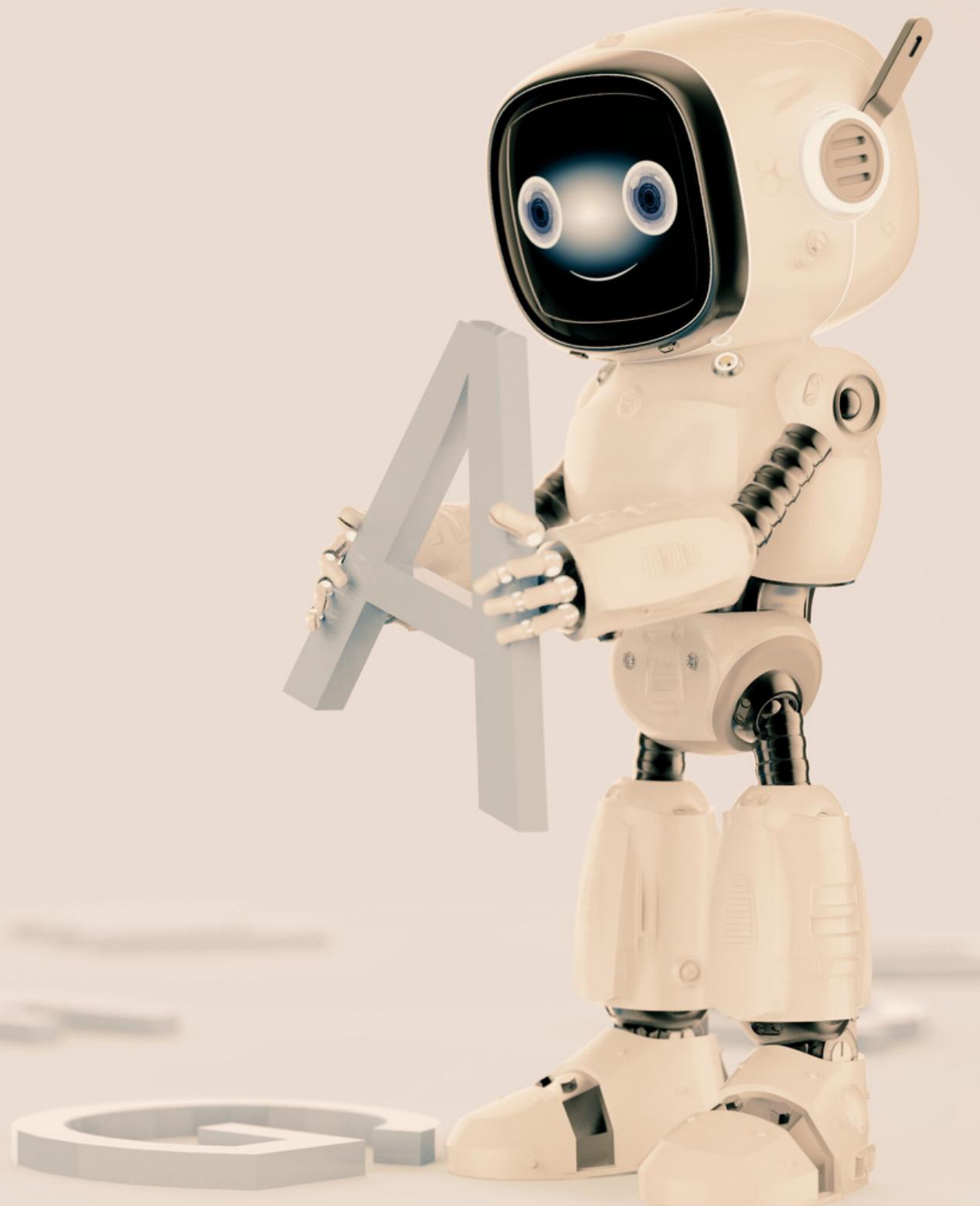
Persistencia

Parte 3:

Desarrollo de Aplicación

Objetivos de
desarrollo sostenible

iv





Función 1

Objetivo: Clasificar **un único texto** ingresado por un usuario.

Usuario: Habitante Local

Proceso: Identifican cuáles son los principales problemas del país para que luego se formulen buenas políticas

Función 2

Objetivo: Clasificar **un archivo excel (.xlsx)** con varios textos ingresado por un usuario.

Usuario: Trabajador Organización, CEO

Proceso: Identificar y contribuir a definir las prioridades estratégicas y operativas globales del UNFPA.



Conclusiones Y Recomendaciones



Modelo Seleccionado

Se utilizó **clasificador Ridge** con **TF-IDF** ya que es una buena opción para resolver este problema para UNFPA.

Interpretabilidad

El clasificador de Ridge le permite al negocio puede verificar cuáles **features** está utilizando más el modelo con la **funcionalidad de (WordCloud)**.

Monitoreo Y Re-entrenamiento

Es esencial tener en cuenta que el modelo se construye utilizando datos pasados y una pequeña base de datos. [Se necesitan + datos.](#)

Tiempo

El **tiempo de respuesta (latencia)** para el usuario es **ALTO**. Este es un factor importante a mejorar.



Inteligencia de
Negocios

¡ Gracias !

Para cualquier pregunta o
aclaraciones sobre nuestro
informe quedamos pendientes.





La persona encargada de probar la aplicación fue Juan José Suárez, quien es el experto en estadística asignado al proyecto.