
 <p>Universidad de los Andes Colombia</p>	<p>Ingeniería de Sistemas y Computación Pregrado ISIS-3301 – Inteligencia de Negocios Primer Proyecto Semestre: 2023-20</p>	
--	---	---

Proyecto 1 – Entrega 2

Integrantes:

Juan Sebastián Sánchez Delgado - js.sanchezd1

Miguel Cárdenas Cárdenas - ma.cardenasc1

Santiago Paerez Gonzales – s.paeres10

Institución:

Universidad De Los Andes

Curso:

Inteligencia de negocios – ISIS3301 (Sección 1)

Profesores:

Fabian Peña

María del Pilar Villamil

Haydemar Nuñez

1. Proceso de automatización del proceso de preparación de datos, construcción del modelo, persistencia del modelo y acceso por medio de API (20%)

Para el proceso de automatización destinado a la preparación de datos, construcción del modelo, persistencia y acceso por medio de una API, se utilizó el Pipeline creado en la primera etapa, el cual tiene como algoritmo el RandomForest y utiliza la estrategia TF-IDF para la cuantificar la relevancia de las palabras en el documento dentro de una colección de documentos, como se muestra a continuación:

```
# Crear el pipeline
pipeline_tfidf = Pipeline([
    ('tfidf', tfidf), # Etapa de TF-IDF
    ('classifier', tfidf_model) # Etapa de modelo de clasificación
])
```

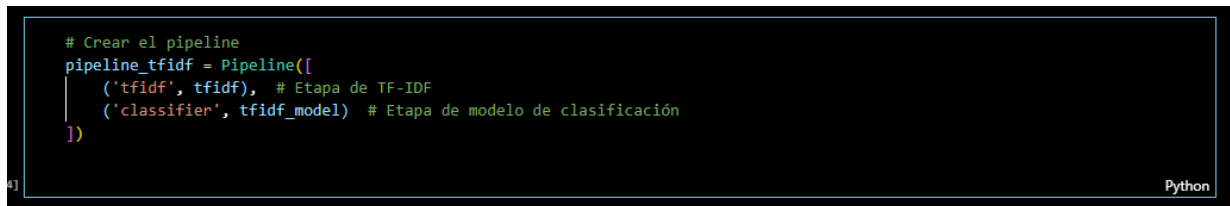


Figura 1. Creación del pipeline.

Una vez se construye el pipeline, se procede a exportarlo o persistirlo en un archivo que posteriormente será cargado y utilizado para conectar el modelo a la API.

```
import joblib

filename_model = "tfidf_model.joblib"
filename_transform = "tfidf_transform.pkl"

# Guarda el modelo en un archivo
joblib.dump(pipeline_tfidf, filename_model)
joblib.dump(tfidf, filename_transform)
```

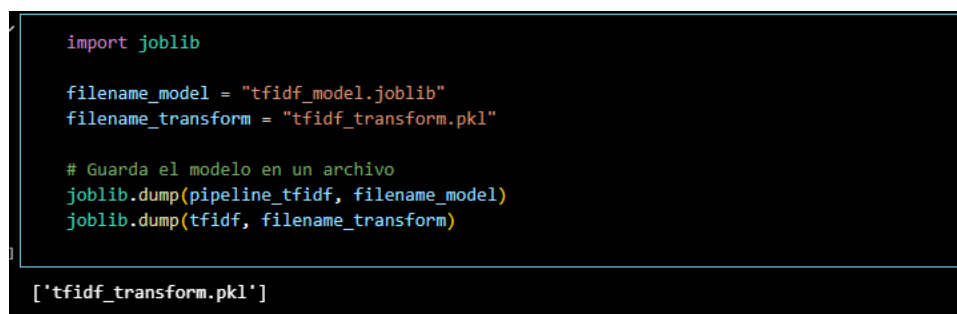


Figura 2. Persistencia del modelo.

Finalmente, se hizo uso del framework FastAPI para poder conectar el modelo a la API, y así poder usar esta para la predicción de nueva información textual mediante la aplicación web.

2. Desarrollo de la aplicación y justificación (40%)

El usuario principal de la aplicación será la entidad UNFPA, más conocido como el Fondo de Poblaciones de las Naciones Unidas. Esta es una entidad importante involucrada en el seguimiento y la evaluación de las políticas públicas y su impacto a nivel social. El rol principal dentro de la UNFPA que hará uso del modelo desarrollado será el analista de datos, esto se debe principalmente a que son los encargados de recopilar y analizar la información textual recolectada, la cual se encuentra relacionada con los Objetivos de desarrollo sostenible. Este es uno de los roles más importantes dentro de la organización, al desempeñar un papel crucial en

la interpretación, entendimiento y análisis de textos que pretende apoyar los esfuerzos realizados por la UNFPA en la consecución de los ODS.

Ahora bien, la aplicación tendría una conexión directa con el proceso de seguimiento y evaluación realizado por la UNFPA sobre las políticas públicas y su impacto a nivel social, que como se ha mencionado antes, es una actividad esencial para el cumplimiento de los ODS. La aplicación se presenta como una herramienta que permite automatizar y agilizar la clasificación de textos de acuerdo a un ODS correspondiente. Lo anterior permite el ahorro de tiempo, recursos y personal, pues reduce la necesidad de tener una persona experta que realice una clasificación manual de cada documento. Una vez esta información es clasificada, puede ser usada en el análisis de las políticas públicas y la toma de decisiones informadas que promuevan y mejoren la calidad de vida de las personas.

Como se mencionó anteriormente, la existencia de la aplicación implica una alta importancia para los analistas de datos de la UNFPA, pues les permite agilizar el proceso de clasificación de información de acuerdo a algún ODS (3,4,5). La importancia de esta aplicación se presenta en los siguientes apartados:

- **Precisión:** al realizar la clasificación mediante un modelo de clasificación robusto y eficiente, la precisión de la clasificación puede mejorar en comparación con un proceso de clasificación manual, lo que podría garantizar que cada documento esta alineado de una manera mucho más precisa con los ODS.
- **Eficiencia:** la aplicación permite automatizar la tarea de la clasificación de información textual, lo que permite ahorrar tiempo, dinero y recursos al comparar dicha actividad con un proceso manual.
- **Escalabilidad:** la aplicación podría manejar grandes volúmenes de datos de manera más eficiente, presentándose como una oportunidad importante, considerando la diversidad y cantidad de información que podría ser recopilada por UNFPA.
- **Análisis informado:** el resultado de las predicciones puede ser usado en el análisis de políticas públicas y la toma de decisiones, contribuyendo directamente al logro de los ODS y al mejoramiento de la calidad de vida de las personas implicadas en estas políticas.

De acuerdo con lo anterior, la aplicación se presenta como una oportunidad importante y como una herramienta esencial para la UNFPA, pues permite automatizar un proceso tan importante y que requiere mayor esfuerzo como lo es la clasificación de información textual. Lo anterior apoya completamente los esfuerzos realizados por la ONU para la consecución de los Objetivos de Desarrollo sostenible y la evaluación de proyectos y políticas públicas que garanticen el bienestar de las personas.

3. Trabajo en equipo (10%)

Miembro	Roles	Tareas	Tiempos
---------	-------	--------	---------

Juan Sebastián Sánchez	Líder de proyecto Ingeniero de software responsable de desarrollar la aplicación final	Desarrollo de la aplicación	3h
		Informe	1h
Miguel Cárdenas Cárdenas	Ingeniero de software responsable del diseño de la aplicación y resultados	Video de resultados	1h 30min
		informe	1h
Santiago Paeres Gonzales	Ingeniero de datos	Automatización de la construcción y persistencia del modelo	4h

Tabla 1. Distribución de trabajo

En la Tabla 1 se muestran las tareas desarrolladas por cada integrante, los tiempos que se dedicó a cada una de ellas y los roles asignados a cada miembro del equipo. Durante el desarrollo del presente proyecto el equipo de trabajo tuvo que enfrentar algunos retos. El mayor reto para la presente entrega fue garantizar que la aplicación del Frontend pudiera acceder a los servicios disponibles en la API sin contratiempos. Para superar este obstáculo, fue necesario implementar constantes correcciones relacionadas con las versiones de las librerías, el formato de los JSON intercambiados, las rutas y métodos de la API, etc.

Con respecto a la distribución de los puntos, si bien todos los miembros del equipo colaboraron de manera activa en el desarrollo de esta segunda entrega, consideramos que el método más justo para repartir los puntos consiste en tomar como base las tareas realizadas cada uno. En concreto, la idea sería hacer un balance general entre las tareas hechas junto con los tiempos de realización de cada una para distribuir los 100 puntos disponibles. También sería importante considerar el peso que tiene cada tarea dentro de la calificación al momento de hacer dicha ponderación.

3. Referencias

Green Globe Sostenibilidad y Proyectos Ambientales. (2018, 5 septiembre). *Qué son los objetivos de Desarrollo Sostenible ODS y la agenda 2030*. <https://www.greenglobe.es/los-objetivos-desarrollo-sostenible-ods-la-agenda-2030/>

Caribe, C. E. P. A. L. Y. E. (s. f.). *Objetivos de Desarrollo Sostenible (ODS)*.
CEPAL. <https://www.cepal.org/es/temas/agenda-2030-desarrollo-sostenible/objetivos-desarrollo-sostenible-ods>

Scikit-learn. (S.F.). Text feature extraction. https://scikit-learn.org/stable/modules/feature_extraction.html#text-feature-extraction