



Universidad de los Andes
Inteligencia de Negocios

Proyecto 1: Analítica de textos para ODS del UNFPA

Etapa 2 – Automatización y uso de modelos de analítica
de textos

Grupo 19

Laura Daniela Arias – 202020621
Nicolás Miguel Murillo – 202213891
Juan Pablo Peña – 202212756

12 de octubre de 2024

TABLA DE CONTENIDO

CONTEXTO	2
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.....	3
RESULTADOS	4
TRABAJO EN EQUIPO.....	4

CONTEXTO

En este proyecto se busca apoyar al Fondo de Poblaciones de las Naciones Unidas (UNFPA). El objetivo principal de la organización es relacionar de forma automática opiniones de los ciudadanos con los Objetivos de Desarrollo Sostenible (ODS) 3, 4 y 5. La tarea se realizará por medio del análisis de información textual proporcionada por este actor en la forma de opiniones de distintos ciudadanos sobre los ODS de interés.

Esta segunda etapa se centra en hacer el modelo accesible al usuario final (UNFPA)

PROCESO DE AUTOMATIZACIÓN DEL PROCESO DE PREPARACIÓN DE DATOS, CONSTRUCCIÓN DEL MODELO, PERSISTENCIA DEL MODELO Y ACCESO POR MEDIO DE API

Justificación de decisiones tomadas:

- Se usó como framework Flask para generar tanto la app web como el API. Esto debido a sus tiempos de ejecución y a la facilidad de comunicación al usar el mismo framework para el front y el back, y el mismo lenguaje de programación que para el modelo (python)

Reentrenamiento

Para la parte de reentrenamiento del modelo hay diferentes estrategias que se pueden aplicar. Tres de estas son *reentrenamiento completo*, *reentrenamiento incremental*, y *reentrenamiento en frío*. Explicamos brevemente qué implica cada definición, sus ventajas y desventajas, y el reentrenamiento elegido para el caso con su justificación.

Reentrenamiento completo:

- En esta definición se reentrena todo el modelo desde cero utilizando el conjunto de datos completo—es decir, tanto los datos antiguos como los nuevos.
- Ventaja: Garantiza que el modelo tenga en cuenta todos los datos disponibles y, por tanto, maximice su capacidad de generalización.

- Desventaja: El proceso es costoso en términos de tiempo y recursos, especialmente con grandes volúmenes de datos como ocurre en este caso.

Reentrenamiento incremental:

- En este caso el modelo se reentrena con los nuevos datos, pero sin olvidar los previos. En otros términos, únicamente se suma la información nueva al modelo.
- Ventaja: Se puede actualizar rápidamente el modelo sin tener que reentrenarlo desde cero, con lo cual se ahorran tiempo y recursos.
- Desventaja: En caso de que los datos lleguen a cambiar mucho en el tiempo, se puede implicar una pérdida de precisión en el modelo.

Reentrenamiento en frío:

- Con este enfoque se reentrena el modelo desde cero utilizando única y exclusivamente el nuevo conjunto de datos.
- Ventaja: Permite que el modelo se adapte rápidamente al nuevo conjunto de datos con sus posibles nuevos patrones y tendencias.
- Desventaja: El modelo pierde “conocimiento” previo, lo cual afecta negativamente el modelo si estos patrones antiguos siguen siendo significativos.

Para el caso de nuestro proyecto, contamos con datos que no variarían su presentación enormemente con el tiempo, por lo cual acumular la información vieja con la nueva es increíblemente útil para refinar el modelo. Debido a esto, teniendo en cuenta las definiciones anteriores, optamos por un reentrenamiento incremental. Con esto, podemos enriquecer el modelo juntando la información previa con la entrante sin consumir tantos recursos como en el caso del reentrenamiento completo.

DESARROLLO DE LA APLICACIÓN Y JUSTIFICACIÓN

Rol que usará la aplicación: Administrador de datos de la UNFPA

Descripción del rol: Este es el rol dentro de la UNFPA que se encarga de desarrollar el proceso de negocio de clasificación de ODS. Las labores relacionadas son utilizar los datos recopilados para clasificarlos en los distintos ODS. Este rol estaría directamente relacionado o subordinado por los administrativos financieros.

Conexión entre la app y proceso de negocio: La app permite clasificar los ODS de manera automática, por lo que, al automatizar el proceso, hace que esta labor sea mucho más rápida ahorrando recursos como tiempo y dinero para la empresa

Importancia de la aplicación para el Rol identificado: Al tener una app que automatiza su trabajo, la persona o personas que desempeñan este rol pueden clasificar una cantidad mayor de datos en mucho menos tiempo y disminuir su carga laboral, o usar el tiempo ahorrado para otras funciones.

RESULTADOS

Link del proyecto en GitHub: <https://github.com/Laudarias/BI-Proy1-GProy19>

Instrucciones de uso (solo se reciben archivos en el formato esperado en .xlsx)

1. Clonar el proyecto
2. Correr el archivo *modelo.py*, posteriormente el *api.py* y luego *app.py*
3. Abrir el navegador en *localhost:5001* el cual permitirá acceder a las funciones de predicción y reentrenamiento.

Resultados generales:

- La aplicación web permite recibir todos los inputs necesarios, para luego llamar al api de *Flask* quien usa el modelo para devolver la predicción y su probabilidad, la cual finalmente se vuelve a mostrar en la web.
- Respecto a las predicciones, se obtiene el sdg asociado a los diferentes textos con la probabilidad de cada uno.
- Para el tema del reentrenamiento, luego de subir el archivo, el modelo hace el reentrenamiento continuo, finalmente muestra las métricas asociadas al modelo.

El video mostrando el funcionamiento del programa se encuentra en padlet.

TRABAJO EN EQUIPO

Líder de proyecto: Laura Arias

Ingeniero de datos: Juan Peña

Ingeniero de software responsable del diseño de la aplicación y resultados: Nicolás Murillo

Ingeniero de software responsable de desarrollar la aplicación final: Nicolás Murillo

Reuniones

- Reunión de lanzamiento y planeación
 - o Fecha: octubre 11.
 - o Se revisó el enunciado y se definieron las tareas a realizar y su distribución
- Reunión seguimiento
 - o Fecha: octubre 12.
 - o Revisión del software producido y planeación de la entrega
- Reunión de finalización:
 - o Fecha: octubre 12.
 - o Finalización del documento, revisión final del software, producción del video y entrega en Bloque Neón y Padlet

Aporte Individual

Laura Daniela Arias Flórez

- Tareas realizadas: Apoyo con código, aporte del modelo de la etapa 1 a utilizar para la etapa 2, realización del documento

- Horas dedicadas: 10 approx.
- Retos enfrentados: Problemas con el funcionamiento del código en máquina personal
- Resolución de los retos: Trabajar mediante llamada apoyando con búsquedas y sugerencias de código al compañero corriendo el programa.
- Puntos: 33.3/100

Juan Pablo Peña Jaime

- Tareas realizadas: Apoyo con el código, especialmente con el primer endpoint (encargado de las predicciones) y el frontend.
- Horas dedicadas: 10 approx.
- Retos enfrentados: Normalmente aparecían varios errores, sobre todo tratando los documentos .xlsx para hacer las predicciones con el modelo.
- Resolución de los retos:
- Puntos: 33.3/100

Nicolás Miguel Murillo Cristancho

- Tareas realizadas: Apoyo con el código, especialmente con el segundo endpoint (encargado del reentrenamiento) y el funcionamiento de la API.
- Horas dedicadas: 10 approx.
- Retos enfrentados: Consumir la API, ya que a veces saltaban errores por el formato o un llamado incorrecto.
- Resolución de los retos:
- Puntos: 33.3/100

Teniendo en cuenta el trabajo realizado para la entrega podemos identificar un par de puntos de mejora. El primero es, muy claramente, realizar las actividades requeridas con mayor antelación.