**Universidad de los Andes**

Facultad de Ingeniería

Departamento de Ingeniería de Sistemas

Inteligencia de negocios

2023-20

Icono

Descripción generada automáticamente con confianza media

**Proyecto 1 – Etapa 2**

Daniel Fernando Gomez Barrera - 201728920

Lindsay Vanessa Pinto Morato – 202023138

Yei Hong Zhang Cárdenas – 201823976

29 de octubre de 2023

Bogotá D.C.

Contenido

[1 Descripción del proceso e implementación. 2](#_Toc149325979)

[2 Desarrollo de la aplicación y justificación 2](#_Toc149325980)

[2.1 Descripción del usuario que utilizará la aplicación 2](#_Toc149325981)

[2.2 Apoyo del grupo de estadística 3](#_Toc149325982)

[2.3 Desarrollo de la aplicación 3](#_Toc149325983)

[Realización de la página web mediante el API 3](#_Toc149325984)

[3 Resultados 3](#_Toc149325985)

[4 Trabajo en equipo 3](#_Toc149325986)

[4.1 Reuniones 3](#_Toc149325987)

[4.2 Puntos que mejorar para la siguiente entrega 4](#_Toc149325988)

# Descripción del proceso e implementación.

## Construcción del pipeline

Primero, se importan diversas librerías para el análisis de datos y el procesamiento de texto, incluyendo herramientas como NumPy, Pandas y Scikit-learn. A continuación, se carga un conjunto de datos desde un archivo Excel llamado "cat\_6716.xlsx" usando Pandas, lo que genera un DataFrame denominado df. El código luego aborda el preprocesamiento del texto, que es esencial para la correcta clasificación. Se definen funciones como change\_non\_ascii y remove\_accents para tratar caracteres no ASCII y quitar acentos, respectivamente. También se llevan a cabo acciones como convertir a minúsculas, eliminar puntuaciones y stopwords, además de resolver caracteres especiales. Estas funciones son esenciales para normalizar el texto y facilitar la tarea de clasificación.

Una vez que el texto se ha procesado adecuadamente, se procede con la tokenización y el stemming. Se utiliza el stemmer Snowball específico para el español para reducir las palabras a sus formas base, lo que simplifica el proceso de clasificación. Posteriormente, se emplea el vectorizador TF-IDF para transformar el texto preprocesado en una representación numérica que pueda ser utilizada por el modelo. Se configura un modelo de clasificación logística que será entrenado en los datos transformados. Para facilitar el flujo de trabajo, se crea un pipeline que combina todas estas etapas de procesamiento y clasificación. Este pipeline permite encadenar operaciones y ejecutarlas de manera más eficiente.

Finalmente, el modelo entrenado se guarda en un archivo binario llamado "modeloP1E2Regresion.pkl" para su posterior uso. Es importante destacar que este código es parte de un proceso más amplio y su eficacia dependerá en gran medida de la calidad y adecuación de los datos de entrada, así como de la elección del modelo y los parámetros.

# Desarrollo de la aplicación y justificación

## Descripción del usuario que utilizará la aplicación

|  |  |  |  |
| --- | --- | --- | --- |
| **Rol dentro de la**  **empresa** | **Tipo de actor** | **Beneficio** | **Riesgo** |
| Equipo de Evaluación de Proyectos | Empleados de CAF | Evaluar y seleccionar proyectos que estén alineados con los ODS de manera más efectiva y objetiva. | Evaluar y seleccionar proyectos que estén alineados con los ODS de manera más efectiva y objetiva. |

**Descripción del usuario de la organización:**

El usuario principal de la aplicación será un miembro del Equipo de Evaluación de Proyectos del Banco de Desarrollo de América Latina y el Caribe. Este equipo está compuesto por profesionales especializados en la evaluación y selección de proyectos que buscan apoyar el desarrollo sostenible en la región. Además, tienen la ventaja de que tienen la capacidad de analizar proyectos en términos de su alineación con los ODS.

**Conexión entre la aplicación y el proceso de negocio:**

La aplicación desempeña un papel crítico en el proceso de evaluación y selección de proyectos. Cuando el Equipo de Evaluación de Proyectos recibe propuestas de financiamiento, utilizan la aplicación para ingresar los textos relacionados con cada proyecto. La aplicación utiliza técnicas de procesamiento de lenguaje natural (NLP) para analizar estos textos y determinar a qué ODS se alinea cada proyecto. Este análisis proporciona a los evaluadores una estimación objetiva y cuantitativa de la contribución de cada proyecto a los ODS. Esto es crucial para asegurar que los recursos del banco se asignen de manera efectiva y que los proyectos seleccionados tengan un impacto positivo y medible en el desarrollo sostenible de la región.

**Importancia para el rol:**

La existencia de esta aplicación es importante para el rol del Equipo de Evaluación de Proyectos pues les proporciona una herramienta eficiente para evaluar y seleccionar proyectos de manera efectiva y objetiva. Les permite tomar decisiones basadas en datos concretos y asegura que los recursos del banco se utilicen para respaldar proyectos que están alineados de manera precisa con los ODS. Además, la aplicación agiliza el proceso de evaluación, lo que significa que el equipo puede manejar un mayor volumen de propuestas sin comprometer la calidad de la evaluación. Esto es esencial en un entorno en el que hay una creciente demanda de financiamiento para proyectos de desarrollo sostenible en la región.

## Apoyo del grupo de estadística

La persona del grupo de estadística nos apoyó con el diseño de la aplicación y la elección de gráficas presentadas en la aplicación se llama Isabella Gómez. El proceso fue primero explicar cómo era el funcionamiento de nuestro modelo, específicamente la regresión logística y cómo funcionan su asignación de peso según el token.

En la presentación de la aplicación final, le mostramos a Isabella cómo funcionaba la aplicación y las dos funcionalidades, y después le dijimos que nos diera una retroalimentación desde el punto de vista como si fuera el usuario final. La retroalimentación que nos dio fue que le parecía buena la elección de gráficas y que estuvo bien que mostráramos de donde salen los datos estadísticos y las tablas.

## Desarrollo de la aplicación

## Realización de la página web mediante el API

Para el caso del desarrollo de la pagína web y del API esta se conformó por dos partes principalmente, el back y el front. Para el caso del front, este fue desarrollado en .html, por el cual incluimos cuadros de texto en los que el usuario podía incluir el texto que se quiere predecir.

Continuando con el desarrollo de la aplicación, se usó Django como tecnología para el despliegue de la aplicación. La aplicación tiene dos funcionalidades, la primera es la clasificación de textos según el ODS, para usarla el usuario tiene que ingresar los textos separados por un salto de línea. El modelo clasifica todos los textos y le asigna el ODS al que predice, para probar la aplicación se usaron los textos otorgados en el proyecto.

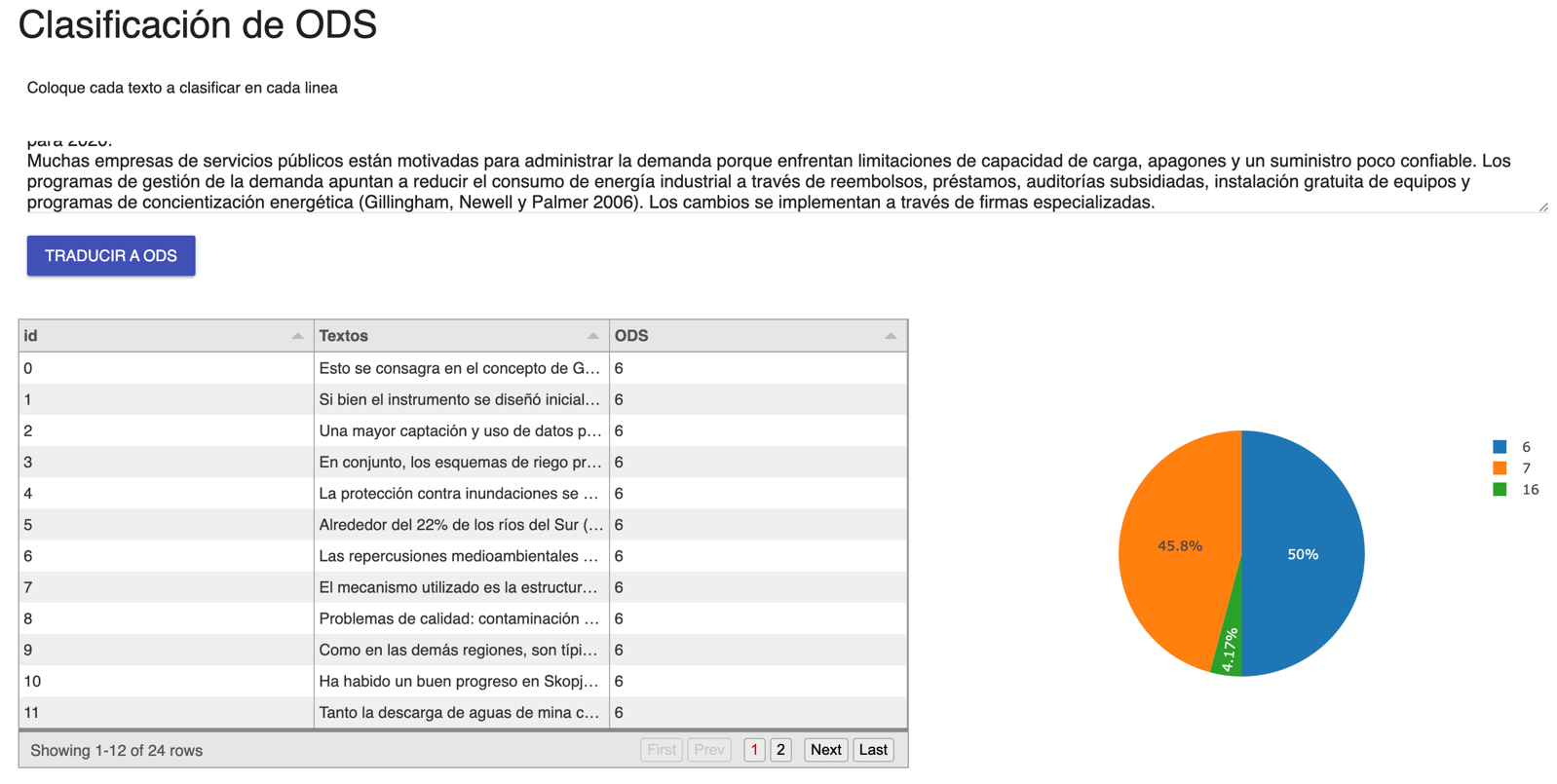


Ilustración Funcionalidad de clasificación

Como se puede observar en la ilustración 1, el usuario podrá insertar todos los textos que requieran clasificarse y al dar clic en *traducir a ODS* el modelo indica el texto que se clasificó y a que ODS pertenece, además muestra una gráfica donde se puede observar el porcentaje de todos los textos que se clasificó.

La segunda funcionalidad es observar las características de un solo texto, este análisis es posible ya que usamos el modelo de regresión lineal lo que nos permite ver la tokenización del documento, el peso que le otorga a cada ODS según el token y así mismo la probabilidad a la que puede pertenecer según el peso y el token.

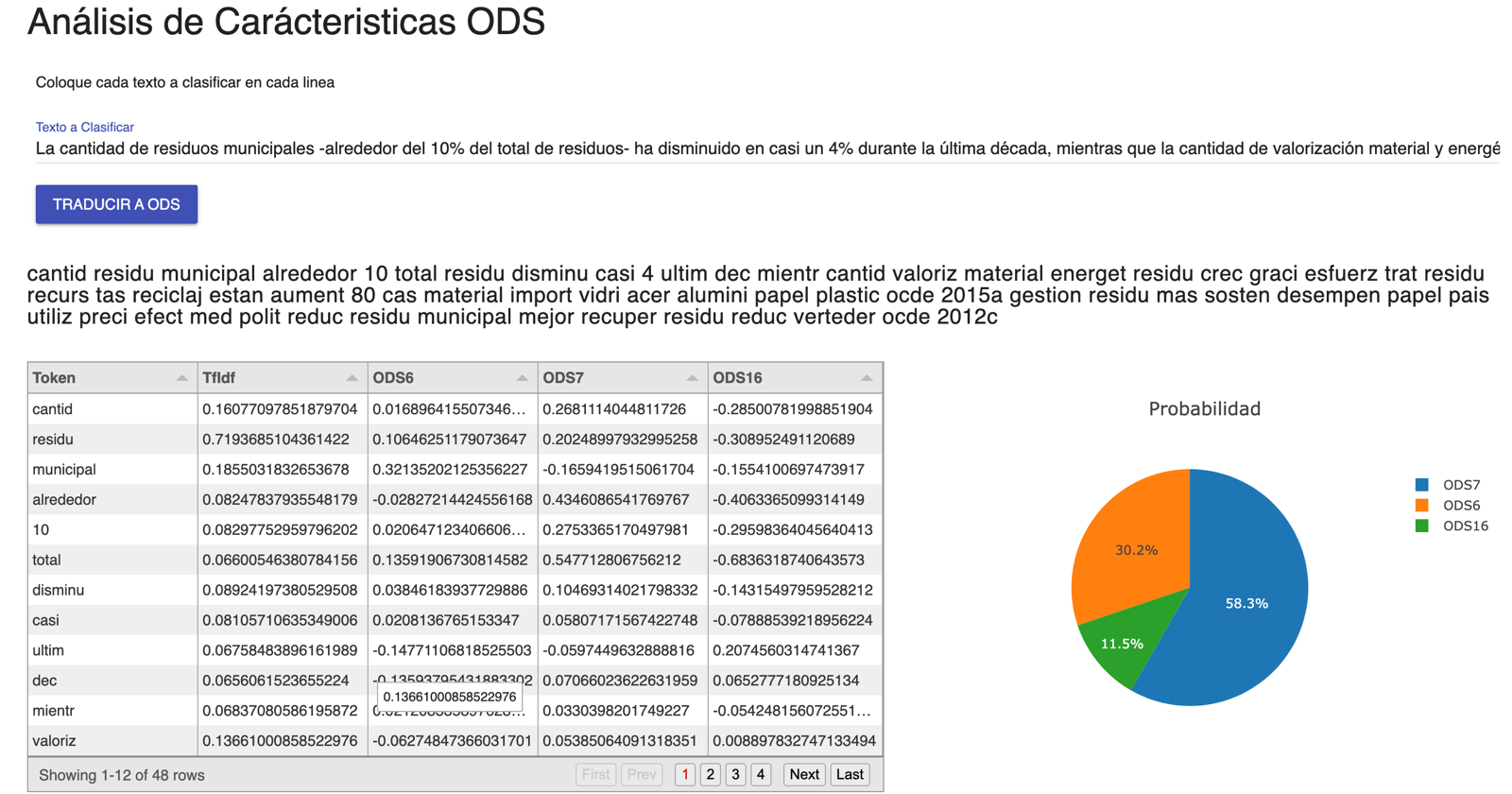


Ilustración Funcionalidad de características

# Resultados

Esta última entrega se tratarán tres objetivos de desarrollos sostenibles; agua limpia y saneamiento, energía asequible no contaminante y finalmente paz justicia instituciones sólidas, estos son los ODS 6, 7 y 16 planteados por las naciones unidas. Como se demostró en la etapa 1 se desarrolló un modelo de clasificación que permitiera relacionar automáticamente textos con los ODS y en esta segunda etapa es desarrollar una aplicación que facilitara la interacción de un usuario final con los resultados permitiéndole interpretar y analizar la información.

El criterio de éxito es que el modelo tuviera una métrica superior al 90% y que la aplicación fuera intuitiva y eficiente en la presentación de resultados. Para retomar la etapa 1, el modelo elegido fue la regresión logística que le otorga un peso a cada token o palabra demostrando que tan importante o influyente es en el texto para clasificarla en alguno de los objetivos.

La aplicación final como se mencionó antes fue desarrollada con Django y tiene dos funcionalidades, el clasificador de OBS y el análisis de características. Para el despliegue de la aplicación se recomienda usar crear un entorno virtual y activarlo, antes de iniciar la aplicación debe instalar todos los paquetes de Python requeridos.

# Trabajo en equipo

## Reuniones

|  |  |  |
| --- | --- | --- |
|  | **Fecha** | **Temas tratados** |
| **Primera reunión** | 17 de octubre de 2023 | En esta reunión se abordaron los temas relacionados con las dudas que surgen en el desarrollo de la aplicación, así como las posibles tareas a llevar a cabo. |
| **Segunda reunión** | 25 de octubre de 2023 | Se realiza seguimiento al trabajo realizado hasta la fecha haciendo énfasis en las dudas que se tienen y los roles a partir del momento. Se coordinan temas de entrega y fechas para apoyo con estadística. |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Estudiante** | **Rol o Roles** | **Trabajo realizado** | **Retos y cómo se resolvieron** | **Puntos asignados** | **Número de horas** |
| Daniel Gomez | Líder de proyecto | Seguimiento y coordinación del proyecto.  Creación de la aplicación y despliegue del modelo.  Despliegue de Api para producción. | Reto: Creación de la aplicación con tecnologías nuevas.  Solución: Aplicar conocimientos de Django para crear la aplicación. | 36 | 12 Horas |
| Lindsay Pinto | Ingeniero de datos  Ingeniero de software responsable del diseño de la aplicación y resultados | Diseño y creación de la interfaz de usuario  Creación del pipeline dentro del proyecto para exportación de archivo pkl.  Inicio de la realización del API | Reto: Ajustar lo obtenido en la etapa 1 en el pipeline de exportación del proyecto  Solución: Buscar documentación y apoyo con mis compañeros | 34 | 12 Horas |
| Yei Hong Zhang | Ingeniero de software responsable de desarrollar la aplicación final | Coordinación con el equipo de estadística  Presentación de aplicación al equipo de estadística. | Reto: Despliegue y comprobación de la Api  Solución: Usar postman para el desarrollo, integración y pruebas del desplieje. | 30 | 10 Horas |

## Puntos que mejorar para la siguiente entrega

Una de las áreas de mejora para la próxima entrega implica dedicar más tiempo a revisar tanto los elementos entregados como el proyecto en su conjunto. Es importante programar una reunión para aclarar y complementar la comprensión de cada miembro del equipo sobre su contribución. Aunque seguiremos apoyándonos mutuamente, es esencial mejorar la comunicación para poder abordar las inquietudes de nuestros compañeros de forma más ágil.