# Etapa 1 - Entendimiento del negocio y enfoque analítico

Paula Daza Diaz 202111276

Juan Camilo Reyes - 201922989

Sofia Torres Ramírez – 202014872

Es esencial llevar a cabo un proceso metódico para comprender plenamente el negocio y definir el enfoque analítico para el proyecto asignado. Se llevarán a cabo las siguientes tareas para cumplir con esto:

1. **Comprender el contexto**

Se inicia por comprender la relevancia y el contexto de los Objetivos de Desarrollo Sostenible (ODS) y su importancia en el ámbito global. Los ODS son un conjunto de metas globales de la ONU para abordar desafíos sociales, económicos y ambientales.

* Definición de ODS

Los Objetivos de Desarrollo Sostenible (ODS) son un conjunto de 17 objetivos globales adoptados por la Asamblea General de las Naciones Unidas en septiembre de 2015. Estos objetivos se han establecido como un llamado universal a la acción para poner fin a la pobreza, proteger el planeta y garantizar que todas las personas gocen de paz y prosperidad para el año 2030. Representan un compromiso global para abordar desafíos interconectados en áreas sociales, económicas y ambientales.

* + Definición de ODS Específicos en el Proyecto (ODS 3, 4 y 5)

El proyecto se enfoca en tres ODS específicos:

**Objetivo 3: Garantizar una vida sana y promover el bienestar para todos en todas las edades.** Algunas metas de este objetivo son:

3.1 Para 2030, reducir la tasa mundial de mortalidad materna a menos de 70 por cada 100.000 nacidos vivos.

3.2 Para 2030, poner fin a las muertes evitables de recién nacidos y de niños menores de 5 años, logrando que todos los países intenten reducir la mortalidad neonatal al menos hasta 12 por cada 1.000 nacidos vivos, y la mortalidad de niños menores de 5 años al menos hasta 25 por cada 1.000 nacidos vivos.

3.3 Para 2030, poner fin a las epidemias del SIDA, la tuberculosis, la malaria y las enfermedades tropicales desatendidas y combatir la hepatitis, las enfermedades transmitidas por el agua y otras enfermedades transmisibles.

3.4 Para 2030, reducir en un tercio la mortalidad prematura por enfermedades no transmisibles mediante la prevención y el tratamiento y promover la salud mental y el bienestar.

3.5 Fortalecer la prevención y el tratamiento del abuso de sustancias adictivas, incluido el uso indebido de estupefacientes y el consumo nocivo de alcohol.

**Objetivo 4: Garantizar una educación inclusiva, equitativa y de calidad y promover oportunidades de aprendizaje durante toda la vida para todos.** Algunas metas de este objetivo son:

4.1 De aquí a 2030, asegurar que todas las niñas y todos los niños terminen la enseñanza primaria y secundaria, que ha de ser gratuita, equitativa y de calidad y producir resultados de aprendizaje pertinentes y efectivos.

4.2 De aquí a 2030, asegurar que todas las niñas y todos los niños tengan acceso a servicios de atención y desarrollo en la primera infancia y educación preescolar de calidad, a fin de que estén preparados para la enseñanza primaria.

4.3 De aquí a 2030, asegurar el acceso igualitario de todos los hombres y las mujeres a una formación técnica, profesional y superior de calidad, incluida la enseñanza universitaria.

4.4 De aquí a 2030, aumentar considerablemente el número de jóvenes y adultos que tienen las competencias necesarias, en particular técnicas y profesionales, para acceder al empleo, el trabajo decente y el emprendimiento.

**Objetivo 5: Lograr la igualdad entre los géneros y empoderar a todas las mujeres y las niñas.** Algunas metas de este objetivo son:

5.1 Poner fin a todas las formas de discriminación contra todas las mujeres y las niñas en todo el mundo.

5.2 Eliminar todas las formas de violencia contra todas las mujeres y las niñas en los ámbitos público y privado, incluidas la trata y la explotación sexual y otros tipos de explotación.

5.3 Eliminar todas las prácticas nocivas, como el matrimonio infantil, precoz y forzado y la mutilación genital femenina.

* Impacto en Colombia

El cumplimiento de estos ODS en Colombia puede tener un impacto significativo en la vida de la población colombiana y en el desarrollo sostenible del país. Aquí hay ejemplos de cómo cada uno de estos ODS puede beneficiar a Colombia:

* **ODS 3 - Salud y Bienestar:** Lograr una vida más saludable reducirá la mortalidad materna y la mortalidad infantil, mejorando la calidad de vida de las madres y los niños colombianos. Además, poner fin a epidemias como el SIDA, la tuberculosis y la malaria reducirá la carga de enfermedades en el país.
* **ODS 4 - Educación de Calidad:** Garantizar una educación inclusiva y de calidad proporciona a los colombianos igualdad de oportunidades para el aprendizaje, lo que puede impulsar el desarrollo educativo y económico del país.
* **ODS 5 - Igualdad de Género:** Al eliminar la discriminación de género y la violencia contra las mujeres y las niñas, Colombia puede avanzar hacia una sociedad más justa y equitativa. Esto permite que más mujeres tengan acceso a oportunidades educativas y laborales, lo que contribuye al crecimiento económico y al empoderamiento de las mujeres.

1. **Definición de Objetivos del Negocio**

Para llevar a cabo este proyecto de manera efectiva, es fundamental establecer objetivos claros y medibles que reflejen las necesidades y metas del proyecto. Aquí se detallan los objetivos del negocio específicos:

* Automatización de la Clasificación de Textos:
* **Objetivo:** Desarrollar un modelo de clasificación de textos basado en técnicas de aprendizaje automático que permita asignar automáticamente un texto a uno de los Objetivos de Desarrollo Sostenible (ODS) específicos.
* **Importancia:** La automatización de la clasificación de textos agiliza el proceso de identificar los temas relacionados con los ODS en grandes volúmenes de datos, lo que ahorra tiempo y recursos.
* Mejora de la Eficiencia en la Evaluación de Políticas Públicas:
* **Objetivo:** Facilitar la identificación y evaluación de políticas públicas relacionadas con los ODS mediante el análisis automatizado de opiniones y comentarios de la población local.
* **Importancia:** La mejora de la eficiencia en la evaluación de políticas públicas permite a UNFPA y otras entidades identificar desafíos y oportunidades más rápidamente, lo que a su vez acelera el proceso de toma de decisiones.
* Identificación de Problemas y Soluciones de Manera Más Rápida y Precisa:
* **Objetivo:** Proporcionar una herramienta que permita identificar problemas y soluciones en relación con los ODS en un contexto territorial de manera más rápida y precisa.
* **Importancia:** La capacidad de identificar problemas y soluciones de manera eficiente ayuda a enfocar los esfuerzos en áreas críticas y a lograr un impacto más significativo en el desarrollo sostenible.
* Criterios de Éxito:

Es esencial definir criterios de éxito claros para evaluar el progreso y el impacto del proyecto. Los criterios de éxito deben ser medibles y orientados a los objetivos del negocio.

* **Precisión de la Clasificación de Textos:** Medir la precisión del modelo de clasificación de textos en asignar correctamente los textos a los ODS específicos. Un alto nivel de precisión es fundamental para la utilidad de la herramienta.
* **Eficiencia en el Procesamiento de Datos:** Evaluar la eficiencia del proceso de procesamiento de datos, incluyendo la velocidad de clasificación de textos y la capacidad de manejar grandes volúmenes de datos de manera oportuna.
* **Tiempo de Respuesta en la Identificación de Problemas y Soluciones:** Medir cuánto tiempo se ahorra en la identificación de problemas y soluciones en comparación con enfoques anteriores.
* **Impacto en la Toma de Decisiones:** Evaluar cómo el proyecto contribuye a la toma de decisiones más informadas y efectivas en la implementación de políticas relacionadas con los ODS.

1. **Enfoque Analítico:**

En este proyecto, se utilizarán técnicas de aprendizaje automático para la clasificación de textos relacionados con los Objetivos de Desarrollo Sostenible (ODS). Los algoritmos seleccionados son los siguientes:

* K-Nearest Neighbors (KNN): KNN es un algoritmo de clasificación basado en la proximidad entre los datos. Funciona asignando una etiqueta a un punto de datos basándose en las etiquetas de sus vecinos más cercanos en el espacio de características. KNN puede ser útil para la clasificación de texto si se utilizan representaciones numéricas de los textos, como vectores TF-IDF.
* Árboles de Decisión: Los árboles de decisión son modelos que utilizan estructuras de árbol para tomar decisiones basadas en las características de entrada. Son efectivos para la clasificación de texto cuando se utilizan características derivadas de la frecuencia de palabras o TF-IDF.
* Random Forest: Random Forest es una técnica que combina múltiples árboles de decisión para mejorar la precisión y reducir el sobreajuste. Puede ser especialmente útil en la clasificación de texto para manejar la alta dimensionalidad de los datos.
  1. **Preprocesamiento de Datos:**

El preprocesamiento de datos es crucial para el procesamiento de texto. Aquí se describen algunas de las tareas de preprocesamiento que se llevarán a cabo:

* Tokenización: Se dividirán los textos en palabras o tokens, lo que permitirá analizar las palabras individualmente.
* Eliminación de Stop Words: Se eliminarán las palabras comunes (stop words) que no aportan información significativa a la clasificación.
* Lematización o Stemming: Se reducirán las palabras a su forma base para eliminar las variaciones de palabras, lo que ayuda a agrupar palabras similares.
* Vectorización de Texto: Los textos se convertirán en vectores numéricos utilizando técnicas como TF-IDF (Term Frequency-Inverse Document Frequency) para representar la importancia de las palabras en los textos.
  1. **Entrenamiento del Modelo:**

El proceso de entrenamiento del modelo implicará los siguientes pasos:

* *Recopilación de Datos Etiquetados:* Se recopilarán textos relacionados con los ODS que estén etiquetados con la categoría correspondiente (ODS 3, 4 o 5). Estos datos se utilizarán para entrenar y evaluar el modelo.
* Selección de Características: Se seleccionarán las características más relevantes, como vectores TF-IDF, para representar los textos en el modelo.
  1. **Evaluación y Ajuste:**

La evaluación del modelo y el ajuste son procesos continuos para garantizar un rendimiento óptimo. Se llevarán a cabo las siguientes acciones:

* División de Datos: Se dividirán los datos en conjuntos de entrenamiento y prueba para evaluar el rendimiento del modelo.
* Evaluación del Modelo: Se utilizarán métricas de evaluación, como precisión, recall, F1-score y matriz de confusión, para evaluar la capacidad del modelo para clasificar los textos correctamente.
* Ajuste del Modelo: Se realizarán ajustes en los hiperparámetros de los algoritmos, como el número de vecinos en KNN o la profundidad de los árboles en Árboles de Decisión y Random Forest, para mejorar el rendimiento del modelo.
* Validación Cruzada: Se puede aplicar validación cruzada para evaluar la capacidad de generalización del modelo en diferentes conjuntos de datos.

El ciclo de evaluación y ajuste se repetirá hasta que se alcance un nivel de rendimiento satisfactorio y los criterios de éxito del proyecto se cumplan. El objetivo final es desarrollar un modelo preciso y eficiente que pueda automatizar la clasificación de textos relacionados con los ODS de manera efectiva.

A continuación, se proporciona la información solicitada en una tabla consolidada:

|  |  |  |  |
| --- | --- | --- | --- |
| **Oportunidad/Problema del negocio** | **Enfoque analítico** | **Organización y rol que se beneficia** | **Contacto con experto externo del proyecto** |
| Automatización de la Clasificación de Textos relacionados con los ODS | Se propone utilizar técnicas de aprendizaje automático, incluyendo K-Nearest Neighbors (KNN), Árboles de Decisión y Random Forest. Se aplicará preprocesamiento de datos, que incluye tokenización, eliminación de stop words, lematización, y vectorización de texto. El modelo se entrenará con datos etiquetados y se evaluará utilizando métricas como precisión, recall y F1-score. | Fondo de Poblaciones de las Naciones Unidas (UNFPA) |  |
| Mejora de la Eficiencia en la Evaluación de Políticas Públicas relacionadas con los ODS | Se utilizarán técnicas de procesamiento de lenguaje natural y análisis de sentimientos para automatizar la evaluación de opiniones y comentarios de la población local. Esto permitirá identificar problemas y soluciones más rápido y con mayor precisión. | UNFPA y otras entidades públicas involucradas en la evaluación de políticas públicas relacionadas con ODS |  |
| Identificación de Problemas y Soluciones relacionados con los ODS de manera eficiente | Se aplicará un modelo de clasificación de textos para identificar problemas y soluciones en un contexto territorial. Esto acelerará el proceso de identificación de desafíos y oportunidades. | UNFPA y otras entidades interesadas en el desarrollo sostenible a nivel territorial. |  |