










<p>TAREA DE APRENDIZAJE</p> <p></p> <p>El tipo de aprendizaje es supervisado, específicamente un problema de clasificación de texto. El objetivo del modelo es predecir a qué categoría de Objetivo de Desarrollo Sostenible (ODS) corresponde una opinión pública, pudiendo ser ODS 1: Fin de la pobreza, ODS 3: Salud y bienestar u ODS 4: Educación de calidad. Los resultados del modelo corresponden a la categoría asignada junto con su probabilidad asociada, y se generan de manera inmediata en el momento en que la opinión es procesada.</p>	<p> DECISIONES</p> <p>Los resultados del modelo permiten identificar cuántas opiniones de la ciudadanía se asocian a cada ODS, lo que facilita conocer cuáles son los temas de mayor interés o preocupación en una comunidad. De esta forma, el UNFPA puede priorizar acciones, programas o inversiones en el ODS con mayor relevancia en las opiniones, logrando un acercamiento más preciso a las necesidades de la población.</p>	<p> PROPUESTA DE VALOR</p> <p>El beneficiario final es el Fondo de Población de las Naciones Unidas (UNFPA) y las entidades públicas que participan en procesos de planificación territorial. Estas instituciones enfrentan la dificultad de analizar miles de opiniones ciudadanas, una tarea que resulta costosa en términos de tiempo y recursos. Contar con un modelo automatizado facilita este análisis, permitiendo obtener información oportuna y confiable. Sin embargo, resultados erróneos en la clasificación podrían derivar en decisiones equivocadas, afectando la priorización de políticas públicas y la distribución de recursos.</p>	<p> RECOLECCIÓN DE DATOS – NO SE DEBE DILIGENCIAR</p>	<p> FUENTES DE DATOS</p> <p>Los datos forman parte del proyecto “OSDG Community Dataset” en su versión de 2023, estos se encuentran en formato .xlsx, el cual contiene un conjunto de opiniones de ciudadanos en lenguaje natural. Cada opinión tiene una etiqueta correspondiente al ODS al que pertenece (1, 3 o 4). Estos datos permiten realizar el objetivo del análisis, ya que están etiquetados y organizados, lo que permite entrenar y evaluar los modelos de clasificación.</p>
<p> SIMULACIÓN DE IMPACTO</p> <p>Las decisiones correctas del modelo generan un alto impacto al clasificar adecuadamente las opiniones con su ODS correspondiente, de esta forma la organización puede identificar con mayor precisión las problemáticas e invertir en estas áreas. Sin embargo, una clasificación errónea puede generar costos asociados a la toma de decisiones equivocadas en políticas públicas. El criterio principal de éxito es alcanzar un F1-score superior a 0.75. Además, es fundamental asegurar la equidad del modelo, evitando sesgos y garantizando que todas las opiniones sean consideradas de manera justa.</p>	<p> APRENDIZAJE (USO DEL MODELO)</p> <p>El uso del modelo es en tiempo real, cada opinión nueva ingresada puede clasificarse de manera inmediata. La frecuencia de uso se aplica de forma continua a medida que llegan los datos.</p>		<p> CONSTRUCCIÓN DE MODELOS</p> <p>Se necesita un modelo de clasificación multiclase, el cual tenga como objetivo que al recibir una opinión pública, este pueda asignarle una de las posibles categorías: ODS 1, ODS 3 u ODS 4. Sin embargo, en la etapa de pruebas buscamos probar con varios tipos de algoritmos recomendados en clase, tales como Bayes, regresión logística, k-vecino o Random Forest, para que por medio de la comparación entre sus rendimientos, al final se seleccione el modelo que mejor cumpla con el F1-score. Este modelo debe de actualizarse de</p>	<p> INGENIERÍA DE CARACTERÍSTICAS</p> <p>Para que el modelo de machine learning pueda procesar las opiniones ciudadanas, es necesario transformar el texto original, que es la variable de entrada principal, en un conjunto de características numéricas. Este proceso comienza con una etapa de preprocesamiento de los datos, en la cual se normaliza el texto convirtiéndolo a minúsculas y se realiza una limpieza para eliminar signos de puntuación, números y caracteres especiales que no aportan valor semántico. A continuación, cada opinión se divide en palabras individuales mediante un proceso de tokenización. Luego, se eliminan las palabras comunes o <i>stop words</i> (como "el", "la", "de") y se aplica un proceso de <i>stemming</i> para reducir cada palabra a su raíz, agrupando así términos con significados similares. Una vez que el texto está limpio y estandarizado, se aplica la técnica de vectorización TF-IDF (<i>Term Frequency-Inverse Document Frequency</i>). Este método convierte cada opinión en un vector numérico que refleja la importancia de cada palabra en relación con el conjunto total de documentos, permitiendo así que el</p>

		<p>forma periódica para que pueda incorporar nuevas opiniones, y logre adaptarse más viable a las prioridades de la población. También debe de mejorarse para poder tener un mejor F1-score y que por medio del reentrenamiento este sea más confiable y también que tenga en cuenta los cambios del contexto (nuevos ODS o definiciones que el modelo necesite conocer).</p> <p>En cuanto al tiempo propuesto para realizar el modelo, se cuenta con tiempo hasta el jueves 11 para poder lograr solucionar dudas del proyecto.</p>	<p>modelo identifique los términos más relevantes para clasificar correctamente las opiniones en los ODS 1, 3 o 4.</p>
	<p>MONITOREO NO SE DEBE DILIGENCIAR</p> <p>¿Qué métricas y KPI se utilizan para hacer un seguimiento del impacto de la solución de ML una vez desplegada, tanto para los usuarios finales como para la empresa? ¿Con qué frecuencia deben revisarse?</p>		

Adaptación de OWNML MACHINE LEARNING CANVAS



Version 1.2. Created by Louis Dorard, Ph.D. Licensed under a [Creative Commons Attribution-ShareAlike 4.0 International License](https://creativecommons.org/licenses/by-sa/4.0/). Please keep this mention and the link to ownml.co when sharing.

[OWNML.CO](https://ownml.co)

