










<p><b>TAREA DE APRENDIZAJE</b></p> <p></p> <p><b>Tipo de aprendizaje:</b> Supervisado.</p> <p><b>Predicción:</b> El modelo predice a qué Objetivo de Desarrollo Sostenible (ODS) pertenece una opinión ciudadana: ODS 1 (Fin de la pobreza), ODS 3 (Salud y bienestar) u ODS 4 (Educación de calidad).</p> <p><b>Posibles resultados:</b> Clasificación de cada texto en una de las tres categorías (ODS 1, ODS 3 u ODS 4).</p> <p><b>Momento de los resultados:</b> El modelo entrega resultados en tiempo real con las opiniones de un momento dado.</p>	<p> <b>DECISIONES</b></p> <p>Los resultados permiten identificar cuáles temas (pobreza, salud, educación) son más relevantes en la ciudadanía.</p> <p>Los tomadores de decisiones (UNFPA, gobiernos locales, entidades públicas) pueden priorizar políticas públicas y recursos en función de las opiniones clasificadas.</p> <p>Los reportes agregados sirven para generar visualizaciones y reportes estratégicos de apoyo a la Agenda 2030.</p>	<p> <b>PROPUESTA DE VALOR</b></p> <p><b>Beneficiario final:</b> Gobiernos locales, UNFPA y otras agencias de cooperación internacional.</p> <p><b>Empresa/organización:</b> Fondo de Población de las Naciones Unidas (UNFPA).</p> <p><b>Problemas que se abordan:</b> Alto costo y tiempo en el análisis manual de textos, dificultad en vincular sistemáticamente la participación ciudadana con los ODS, necesidad de decisiones basadas en evidencia.</p> <p><b>Riesgos:</b> sesgos en los datos pueden llevar a clasificaciones erróneas, y el mal uso de las clasificaciones podría desinformar a los responsables de políticas públicas. Además, el riesgo de pérdida de confianza ciudadana si los resultados no son explicables.</p>	<p> <b>RECOLECCIÓN DE DATOS – NO SE DEBE DILIGENCIAR</b></p> <p>¿Cómo se obtiene el conjunto inicial de entidades y resultados (por ejemplo, extractos de bases de datos, extracciones de API, etiquetado manual)?</p> <p>¿Qué estrategias se aplican para actualizar los datos continuamente, controlando los costos y manteniendo la vigencia?</p>	<p> <b>FUENTES DE DATOS</b></p> <p>Opiniones recolectadas en plataformas de participación ciudadana (encuestas, foros, redes sociales).</p> <p>Bases de datos internas de UNFPA y de gobiernos locales.</p> <p>Datos abiertos de la ONU relacionadas con pobreza, salud y educación.</p>
---	---	--	---	---

<div data-bbox="183 223 416 351">  <div>SIMULACIÓN DE IMPACTO</div> </div> <div data-bbox="183 351 416 925"> <p><b>Costo de decisiones correctas:</b> optimización en la asignación de recursos públicos y políticas más alineadas a las necesidades ciudadanas.</p> <p><b>Costo de decisiones incorrectas:</b> desviación de recursos hacia áreas que no reflejan las prioridades reales de la población, pérdida de credibilidad en el uso de IA y en las entidades promotoras.</p> </div> <div data-bbox="183 925 416 1149"> <p><b>Criterios de éxito:</b> precisión del modelo en clasificación superior al 80%, capacidad de explicar decisiones del modelo (transparencia)</p> </div> <div data-bbox="183 1149 416 1331"> <p><b>Restricciones de equidad:</b> garantizar que los grupos vulnerables estén representados en los</p> </div>	<div data-bbox="416 223 757 351">  <div>APRENDIZAJE (USO DEL MODELO)</div> </div> <div data-bbox="416 351 757 606"> <p><b>Modo de uso:</b> Principalmente en tiempo real (clasificación instantánea de nuevas opiniones).</p> <p><b>Frecuencia de uso:</b> Continua, cada vez que se ingresen nuevas opiniones al modelo.</p> </div>		<div data-bbox="1043 223 1276 351">  <div>CONSTRUCCIÓN DE MODELOS</div> </div> <div data-bbox="1043 351 1276 973"> <p><b>Número de modelos:</b> Realizamos 3 modelos para la clasificación multiclase (ODS 1, 3 y 4) para ir comprobando cuál es el que mejor se ajusta a las necesidades del problema. También hicimos un cuarto modelo de embeddings (BERT).</p> <p><b>Actualización:</b> El modelo debe reentrenarse periódicamente (ej. cada semestre o cuando se acumule un volumen significativo de nuevos datos).</p> </div> <div data-bbox="1043 973 1276 1331"> <p><b>Tiempo disponible para generar el modelo:</b> se tiene un entrenamiento inicial de semanas y los reentrenamientos posteriores serán de días, gracias al pipeline ya definido.</p> </div>	<div data-bbox="1276 223 2080 351">  <div>INGENIERÍA DE CARACTERÍSTICAS</div> </div> <div data-bbox="1276 351 2080 414"> <p><b>Variables/características principales:</b> texto procesado (tokenización, embeddings), frecuencia de palabras clave asociadas a pobreza, salud, educación.</p> </div> <div data-bbox="1276 414 2080 494"> <p><b>Transformaciones aplicadas:</b> Se realizaron un par de transformaciones sobre los datos para garantizar que estuvieran limpios y sin datos atípicos:</p> </div> <div data-bbox="1276 494 2080 1053"> <ul style="list-style-type: none"> <li>• Limpieza de los textos <ol style="list-style-type: none"> <li>1. Convierte todo a minúsculas.</li> <li>2. Elimina caracteres no ASCII.</li> <li>3. Sustituye números por su representación en palabras.</li> <li>4. Quita signos de puntuación.</li> <li>5. Elimina stopwords en español (palabras muy frecuentes sin valor semántico, como “de”, “la”, “que”).</li> </ol> </li> <li>• Tokenización y reconstrucción <ol style="list-style-type: none"> <li>1. Divide el texto en palabras (tokens).</li> <li>2. Aplica la limpieza a cada token.</li> <li>3. Une de nuevo el resultado en un texto plano limpio.</li> </ol> </li> <li>• Vectorización con TF-IDF <ol style="list-style-type: none"> <li>1. Transforma cada documento en un vector numérico.</li> <li>2. El valor de cada posición refleja la importancia de una palabra o bigrama dentro del documento y en el conjunto completo de textos.</li> </ol> </li> </ul> </div>
--	---	--	--	---

datos, evitar sesgos culturales en el análisis de textos.			
	<p><b>MONITOREO NO SE DEBE DILIGENCIAR</b></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](https://ownml.co) when sharing.

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