






<p>TAREA DE APRENDIZAJE</p> <p></p> <p>Tipo de aprendizaje: Supervisado.</p> <p>Predicción: 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>Posibles resultados: Clasificación de cada texto en una de las tres categorías (ODS 1, ODS 3 u ODS 4).</p> <p>Momento de los resultados: El modelo entrega resultados en tiempo real o pocos segundos después del ingreso de la opinión.</p>	<p> DECISIONES</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> PROPUESTA DE VALOR</p> <p>Beneficiario final: Gobiernos locales, UNFPA y otras agencias de cooperación internacional.</p> <p>Empresa/organización: Fondo de Población de las Naciones Unidas (UNFPA).</p> <p>Problemas que se abordan: 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>Riesgos: 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 transparentes o explicables.</p>	<p> RECOLECCIÓN DE DATOS – NO SE DEBE DILIGENCIAR</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> FUENTES DE DATOS</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 Naciones Unidas y ONGs relacionadas con pobreza, salud y educación.</p> <p>API de plataformas de consulta ciudadana y de redes sociales (ejemplo: Twitter API, encuestas digitales).</p>
---	---	--	---	--

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

datos, evitar sesgos lingüísticos y culturales en el análisis de textos.			
	<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)