

<div> <b>TAREA DE APRENDIZAJE</b>  </div> <p>Aprendizaje supervisado. Se predice la clasificación de opiniones ciudadanas en texto en uno de tres ODS: ODS 1 (Fin de la pobreza), ODS 3 (Salud y bienestar), o ODS 4 (Educación de calidad).</p> <p>Los posibles resultados son:</p> <ul style="list-style-type: none"> <li>• Clasificación en una de las 3 clases de ODS (1, 3, 4)</li> <li>• Probabilidades de pertenencia a cada clase</li> <li>• Identificación de características textuales más relevantes por clase (palabras clave con mayor peso en el modelo)</li> <li>• Detección de patrones lingüísticos asociados a cada ODS</li> </ul> <p>Los resultados se observan inmediatamente después del entrenamiento mediante validación cruzada estratificada y se evalúan usando métricas F1-weighted.</p>	<div> <b>DECISIONES</b>  </div> <p>Las predicciones del modelo se transforman en decisiones procesables mediante:</p> <ul style="list-style-type: none"> <li>• Clasificación automática: Cada opinión ciudadana se asigna automáticamente al ODS más relevante (1, 3, o 4)</li> <li>• Priorización de recursos: Las organizaciones pueden enfocar recursos humanos y financieros en las problemáticas más frecuentes por ODS</li> <li>• Análisis de tendencias: Seguimiento temporal de cómo evolucionan las preocupaciones ciudadanas por ODS</li> <li>• Asignación de equipos: Derivación automática de opiniones a equipos especializados según el ODS identificado</li> </ul>	<div> <b>PROPUESTA DE VALOR</b>  </div> <p>El beneficiario final es: la UNFPA (Fondo de Población de las Naciones Unidas), entidades gubernamentales, planificadores de políticas públicas y organizaciones de desarrollo sostenible. Se abordan los siguientes problemas específicos:</p> <ul style="list-style-type: none"> <li>• Análisis manual intensivo: Clasificación manual de miles de opiniones ciudadanas requiere expertos especializados y tiempo considerable</li> <li>• Escalabilidad limitada: Dificultad para procesar grandes volúmenes de información textual de participación ciudadana</li> <li>• Vinculación ODS-ciudadanía: Falta de herramientas automatizadas para relacionar opiniones ciudadanas con objetivos específicos de la Agenda 2030</li> <li>• Recursos humanos escasos: Necesidad de expertos capaces de relacionar contenidos textuales con ODS de manera consistente y precisa</li> </ul> <p>Los riesgos que puede existir para el beneficiario el uso de este modelo son</p> <ul style="list-style-type: none"> <li>• Clasificaciones incorrectas: Asignación errónea de recursos públicos hacia problemáticas no prioritarias</li> <li>• Sesgo de representación: El modelo podría no capturar adecuadamente la diversidad lingüística regional o de grupos poblacionales específicos</li> <li>• Dependencia tecnológica: Riesgo de pérdida de capacidades de análisis manual si el modelo falla</li> <li>• Sobre confianza en automatización: Posible reducción de análisis cualitativo profundo por parte de expertos humanos</li> </ul>	<div> <b>RECOLECCION DE DATOS</b>  </div>	<div> <b>FUENTES DE DATOS</b>  </div> <p>Los datos de entrenamiento vienen del archivo Excel "Datos_proyecto.xlsx" con textos etiquetados manualmente por expertos, textos en español de opiniones ciudadanas sobre problemáticas territoriales y etiquetas correspondientes a ODS 1, 3, y 4 asignadas por especialistas. Los datos de prueba con el archivo Excel "Datos de prueba_proyecto.xlsx" con textos sin etiquetar para evaluación final.</p> <p>Los datos provienen de:</p> <ul style="list-style-type: none"> <li>• Procesos de participación ciudadana territorial documentados por UNFPA</li> <li>• Consultas y opiniones ciudadanas recolectadas en procesos de planificación participativa</li> <li>• Textos pre-etiquetados por expertos en ODS para entrenamiento supervisado</li> </ul> <p>Si son completamente adecuados porque:</p> <ul style="list-style-type: none"> <li>• Contienen opiniones ciudadanas reales sobre problemáticas territoriales</li> <li>• Están pre-clasificadas por expertos en los ODS objetivo (1, 3, 4)</li> <li>• Representan el lenguaje natural usado por ciudadanos para expresar problemáticas</li> <li>• Tienen diversidad en longitud de texto y complejidad temática</li> </ul>
--	--	--	--	--

SIMULACIÓN DE IMPACTO	
<p>Clasificaciones correctas (Beneficios):</p> <ul style="list-style-type: none"> <li>Alto impacto: Orientación eficiente de recursos hacia problemáticas ciudadanas reales</li> <li>Ahorro de tiempo: Reducción de tiempo de análisis manual.</li> <li>Mejor participación: Mayor capacidad de procesamiento permite incluir más voces ciudadanas en toma de decisiones.</li> </ul> <p>Clasificaciones incorrectas (Costos):</p> <ul style="list-style-type: none"> <li>Falsos positivos: Costo medio, recursos asignados a problemáticas menos prioritarias</li> <li>Falsos negativos: Costo alto, problemáticas críticas ciudadanas no identificadas con corrección, especialmente en salud (ODS 3)</li> <li>Sesgo sistemático: Costo muy alto, grupos poblacionales específicos podrían quedar subrepresentados.</li> </ul> <p>¿Cuáles son los criterios de éxito del modelo para su posterior despliegue?</p> <p>Basado en los resultados obtenidos:</p> <ul style="list-style-type: none"> <li>F1-score mínimo</li> <li>Consistencia cross-validation</li> <li>Balance entre clases: Precisión y Recall balanceados para las 3 clases de ODS.</li> <li>Tiempo de respuesta: segundos para clasificación individual</li> <li>Interpretabilidad: Capacidad de identificar las 3 palabras más importantes por clase.</li> </ul> <p>¿Existen restricciones de equidad?</p> <p>Sí, identificadas en el análisis</p> <ul style="list-style-type: none"> <li>Distribución de clases: Se observó desbalance en las etiquetas.</li> <li>Diversidad lingüística: Evitar sesgo hacia variantes específicas del español.</li> <li>Representación geográfica: Asegurar que el modelo funcione bien para diferentes regiones territoriales</li> <li>Longitud de texto: El modelo debe funcionar tanto para textos cortos como largos (se observó variabilidad significativa)</li> </ul>	

APRENDIZAJE (USO DEL MODELO)	
<p>¿El uso del modelo es por lotes o en tiempo real? Implementación híbrida:</p> <ul style="list-style-type: none"> <li>Tiempo real: <ul style="list-style-type: none"> <li>Clasificación inmediata de opiniones individuales en aplicación web/móvil</li> <li>Uso del pipeline completo.</li> <li>Respuesta inmediata para usuarios durante procesos participativos activos</li> </ul> </li> <li>Por lotes: <ul style="list-style-type: none"> <li>Procesamiento masivo de archivos.</li> <li>Análisis de grandes volúmenes de opiniones históricas.</li> </ul> </li> </ul> <p>¿Con qué frecuencia se usa?</p> <ul style="list-style-type: none"> <li>Tiempo real: Continuo durante jornadas de participación ciudadana.</li> <li>Por lotes: Procesamiento periódico de nuevas opiniones acumuladas</li> <li>Reentrenamiento: Periódicamente con nuevas opiniones validadas por expertos</li> </ul>	



CONSTRUCCIÓN DE MODELOS	
<p>¿Cuántos modelos se necesitan? 3 modelos evaluados (selección del mejor):</p> <ul style="list-style-type: none"> <li>Regresión Logística</li> <li>Naive Bayes</li> <li>K-Nearest Neighbors</li> </ul> <p>¿Cuándo deben actualizarse?</p> <ul style="list-style-type: none"> <li>Reentrenamiento completo: Periódicamente con nuevos datos etiquetados por expertos</li> <li>Validación de performance: Periódicamente usando métricas F1-weighted</li> <li>Actualización inmediata: Si F1-score cae por debajo de un rango predefinido.</li> <li>Ajuste de hiperparámetros: Periódicamente para optimizar el funcionamiento del modelo.</li> </ul> <p>¿De cuánto tiempo se dispone para generar el modelo? Si bien el objetivo es cumplir con la agenda de 2030, se requiere una iteración inicial de inmediato para poder empezar a reconocer las falencias y llegar a un modelo apropiado más rápidamente.</p>	

INGENIERÍA DE CARACTERÍSTICAS	
<p>¿Qué variables/características se utilizan en el modelo?</p> <ul style="list-style-type: none"> <li>Texto limpio: Eliminación de caracteres no alfabéticos</li> <li>Normalización: Conversión a minúsculas y eliminación de espacios extra.</li> <li>Tokenización: División por espacios en blanco.</li> <li>Stemming: Aplicación de stemming para reducir palabras a raíz.</li> <li>Stopwords: Eliminación de palabras vacías en español.</li> </ul> <p>¿Qué agregaciones o transformaciones se aplican a las fuentes de datos originales?</p> <ul style="list-style-type: none"> <li>Limpieza de ruido: Remoción de números, símbolos especiales, y puntuación</li> <li>Normalización textual: Conversión a minúsculas consistente</li> <li>Reducción morfológica: Stemming con PorterStemmer para español</li> <li>Filtrado de stopwords: Lista personalizada de palabras vacías españolas</li> <li>Vectorización semántica: Transformación TF-IDF que captura importancia relativa de términos</li> <li>Reducción dimensional: Limitación a 500 características más informativas</li> <li>Balanceo de clases: En Regresión Logística para compensar desbalance.</li> </ul> <p>Características interpretables extraídas:</p> <ul style="list-style-type: none"> <li>Top 3 palabras por ODS: Identificación automática de términos más predictivos.</li> <li>Coefficientes de regresión: Pesos que indican importancia de cada término por clase.</li> <li>Indicadores negativos: Términos que contraindican pertenencia a cada ODS</li> </ul>	

MONITORIO

