

## TAREA DE APRENDIZAJE



El tipo de aprendizaje es supervisado, puesto que entrenamos los modelos con un conjunto de datos ya etiquetados. En este caso, se desea predecir la categoría del Objetivo de Desarrollo Sostenible (ODS) en la que se puede categorizar a una opinión ciudadana.

La tarea de aprendizaje es clasificación, pues cada opinión puede ser catalogada en alguno de los posibles resultados: 1 (para el ODS de Fin de la Pobreza), 3 (para el ODS de Salud y Bienestar) o 4 (ODS de Educación de Calidad).

Se presenta un modelo predictivo que, a partir de una opinión en texto, predice su categoría de ODS. Estos resultados se observan en cuestión de minutos.



## DECISIONES

Los resultados del modelo permiten que las opiniones ciudadanas sean categorizadas en una de los tres tipos de ODS. Esto ayuda a organizaciones, como el Fondo de Población de las Naciones Unidas, responsables de hacer seguimiento de políticas públicas, a identificar tendencias en la opinión ciudadana de manera informada, para poder priorizar problemáticas y tomar decisiones estratégicas frente a ello. Conocer la categoría de ODS más popular actúa como recomendación popular para planificar el diseño políticas, la manera de destinar recursos o de implementar programas, contribuyendo al cumplimiento de su agenda.



## PROPUESTA DE VALOR

El beneficiario final es el Fondo de Población de las Naciones Unidas (UNFPA), así como eventualmente todas las entidades que se encarguen de hacer seguimiento de los Objetivos de Desarrollo Sostenible. Esta organización es un organismo de la ONU.

Los problemas específicos que se abordan incluyen la necesidad de reducir la demanda de recursos significativos y expertos para la clasificación manual de opiniones ciudadanas en ODS, garantizando que esta categorización se realice de manera eficiente. Además, otro reto que se aborda es la clasificación de grandes volúmenes de opiniones, en lugar de analizar solo unas pocas de forma manual.

El riesgo principal del uso del modelo es una clasificación errónea, dado que pueden presentarse falsos positivos y falsos negativos en la predicción. Un falso positivo es que el modelo clasifique una opinión en una categoría de ODS que no le correspondía, mientras que un falso negativo implica que el modelo no identifique la categoría real y por ello la asigne a cualquier ODS. Esto podría llevar a priorizar problemáticas que no corresponden.

Otros riesgos incluyen que exista sesgo al tener más opiniones de un grupo poblacional que de los demás, el manejo que le del modelo al lenguaje sarcástico o el doble







## RECOLECCIÓN DE DATOS – NO SE DEBE DILIGENCIAR



## FUENTES DE DATOS

Los datos utilizados en este proyecto provienen de un conjunto de opiniones ciudadanas en formato de texto, entregadas por los docentes del curso Inteligencia de Negocios como archivo para los registros base de opiniones. Este es una fuente interna, que contiene las opiniones ya recolectadas y clasificadas. Los datos sí se pueden utilizar para realizar el objetivo del análisis, puesto que ya cuentan con la ODS de cada una de las opiniones, lo que permite entrenar los modelos de clasificación.

		sentido, y la confianza ciega del usuario en el modelo, puesto que en realidad se debería hacer seguimiento de su funcionamiento.		
<div>SIMULACIÓN DE IMPACTO</div> <p>Los beneficios de clasificar adecuadamente las opiniones de los ciudadanos usando el modelo comprenden la toma de decisiones informadas para priorizar problemáticas, y por ende optimizar la asignación de recursos, reducir el esfuerzo manual y responder realmente a las necesidades de la sociedad.</p> <p>Por otra parte, si se da una clasificación incorrecta, se puede llegar a priorizar acciones para apoyar un tipo de ODS que no es tan relevante en comparación con los otros. Esto tiene un costo social significativo, pues incluso puede llegar a invisibilizar problemáticas urgentes, perjudicando directamente a los ciudadanos. Además, puede afectar negativamente la reputación de las organizaciones encargadas, ya que se percibiría que sus decisiones no están alineadas con la realidad y se</p>	<div>APRENDIZAJE (USO DEL MODELO)</div> <p>El uso del modelo es por lotes, ya que se entrena y evalúa con los conjuntos de datos que se nos han otorgado. De ese modo, se clasifican varias opiniones ciudadanas a la vez, no clasificaciones individuales.</p> <p>La frecuencia con la que se usa el modelo depende del uso que le dé el usuario, puede hacerse cada vez que se recojan nuevas opiniones de ciudadanos, lo que podría ser cada mes, semana o día, o incluso en fechas específicas en las que el usuario haga seguimiento a las problemáticas sociales.</p>		<div>CONSTRUCCIÓN DE MODELOS</div> <p>Para este proyecto se requieren tres modelos de clasificación supervisada, los cuales deben actualizarse cada vez que se disponga de un nuevo conjunto de opiniones ciudadanas, cuando existan cambios en el contexto de las ODS o al encontrar oportunidades de mejora.</p> <p>El tiempo estimado para generar cada modelo, incluyendo el proceso de ingeniería de características, así como el análisis y evaluación, es de aproximadamente 2 semanas.</p>	<div>INGENIERÍA DE CARACTERÍSTICAS</div> <p>Una característica es una variable de entrada explicativa, o que describe un aspecto de los datos y el modelo emplea para aprender. En nuestro caso, esa es la variable textos, que contiene las opiniones ciudadanas en lenguaje natural. A partir de esta columna se aplicaron las siguientes transformaciones:</p> <p><b>1. Limpieza del texto:</b> se hizo conversión de las opiniones a minúsculas, reemplazo de caracteres numéricos, eliminación de signos de puntuación y stopwords.</p> <p><b>2. Tokenización y lematización:</b> separación de las palabras en cada opinión, y llevamos cada uno de los términos a su forma base.</p> <p><b>3. Vectorización con TF-IDF:</b> transformación de las palabras en valores numéricos, los cuales representan su importancia para describir, o representar la opinión a la que pertenecen.</p> <p>Estas transformaciones permitieron convertir los textos en una matriz numérica de características, la cual se utilizó como entrada en los modelos para ser entrenados, mientras que la columna labels fue la variable objetivo a predecir.</p>

estaría debilitando la confianza en las instituciones. Finalmente, las organizaciones podrían hacer una mala asignación de recursos, generando pérdida o invirtiendo en problemáticas menos prioritarias.

Los criterios de éxito del modelo para su posterior despliegue son un F1 score alto, precisión y recall, además de en lo posible una capacidad de generalización alta, para que el modelo pueda ajustarse a nuevos datos sin hacer overfitting (aprender los datos que ya ha procesado para entrenarse), y baja latencia, que no se demore días en su procesamiento.

Las restricciones de equidad del modelo se relacionan con los sesgos en las opiniones. Todas las opiniones deberían tener la misma probabilidad de ser clasificadas correctamente, sin favorecer a grupos sociales o voces de ciudadanos particulares según sus preferencias o ideologías.

MONITOREO NO SE DEBE  
DILIGENCIAR



## 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**