











<p>TAREA DE APRENDIZAJE</p> <p></p> <p>Es un aprendizaje supervisado en el que se asigna a cada opinión ciudadana una etiqueta correspondiente a un ODS (1, 3 o 4). Los resultados posibles son: clasificación en ODS 1 (Fin de la pobreza), ODS 3 (Salud y bienestar) u ODS 4 (Educación de calidad) y estos se obtienen en el instante de recibir un nuevo texto (tiempo cercano al real).</p>	<p> DECISIONES</p> <p>Los resultados del modelo permiten identificar de manera automática las problemáticas expresadas por los ciudadanos y asociarlas con los Objetivos de Desarrollo Sostenible. De esta manera las organizaciones pueden orientar de forma más precisa las políticas públicas, y en las áreas de los objetivos clave de este proyecto se puede priorizar los recursos en pobreza, salud y educación.</p>	<p> PROPUESTA DE VALOR</p> <p>El proyecto está dirigido para el Fondo de Población de las Naciones Unidas (UNFPA), junto con las entidades gubernamentales que trabajan en torno a los ODS. El problema que buscamos resolver surge de la dificultad que representa analizar de manera manual grandes volúmenes de opiniones ciudadanas, algo que puede resultar bastante costoso y poco eficiente. Además, existe el riesgo de que sesgos en los datos o errores en la clasificación conduzcan a conclusiones equivocadas.</p>	<p> RECOLECCIÓN DE DATOS – NO SE DEBE DILIGENCIAR</p>	<p> FUENTES DE DATOS</p> <p>En este caso, el conjunto de datos forma parte del proyecto OSDG Community Dataset. Este mismo contiene diferentes documentos, que son tenidos en consideración como muestra para el presente análisis. Cabe decir que esta fuente de información empleada, ha sido validada por expertos en la temática, así como por voluntarios. Ahora bien, cabe destacar que estos datos funcionan claramente para el análisis de relación de los textos con los ODSs respectivos. Esto se da gracias a que siempre se encuentran implicadas con el área de investigación, que son los ODS solicitados, que permiten tener patrones para poder llevar a cabo el entrenamiento óptimo del modelo.</p>
<p> SIMULACIÓN DE IMPACTO</p> <p>Primeramente, los valores de costo/beneficio de las decisiones (in)correctas parten de los Falsos Negativos, ya que pueden implicar la oportunidad de identificar o dar más relevancia a un problema urgente. Paralelamente, existe el caso contrario, que es el hacer más énfasis en aquellas problemáticas que quizás no lo requieren, malgastando tiempo y recursos significativos. Por otro lado, sobre las métricas, las más determinantes es el F1-Score, ya que colabora en la revisión de la precisión del modelo en el conjunto de datos, y su relación de equilibrio con el Recall. Adicionalmente, requerimos bastante del segundo en caso de haber empate empleando el F1-Score, ello considerando que indaga en que tan bien el modelo encuentra todos los datos reales o positivos del grupo de instancias. Finalmente, sobre las restricciones de equidad, se puede decir que es incierta. Esto se justifica desde la fuente de los datos, que no exploran intensamente la procedencia de cada texto particular. Al todo hallarse en un mismo idioma, la posibilidad de que sectores sociales o áreas políticas puedan referirse más a una temática que otros, e igualmente ser analizados igualmente, indica que es probable que el modelo pueda inclinarse más hacia un grupo de instancias que a otra.</p>	<p> APRENDIZAJE (USO DEL MODELO)</p> <p>Tomando el caso en el que el modelo ya se encuentre totalmente entrenado, se puede afirmar que se emplearía netamente por lotes. Esto teniendo en cuenta su funcionalidad base, que es la abstracción y análisis de características, a partir de un segmento de datos.</p>		<p> CONSTRUCCIÓN DE MODELOS</p> <p>En este proyecto se implementarán al menos tres algoritmos supervisados, con el fin de comparar su desempeño y seleccionar el más adecuado. El modelo deberá actualizarse cuando se disponga de nuevos datos que estén más actualizados o en el momento en que su rendimiento caiga por debajo del umbral esperado. Se cuenta con aproximadamente dos semanas para la construcción inicial de los modelos, y para realizar los ajustes y la evaluación correspondiente.</p>	<p> INGENIERÍA DE CARACTERÍSTICAS</p> <p>La característica esencial que se usa para este caso es el TF-IDF, que mide la importancia de una palabra en un segmento o en un grupo de textos. En este caso, a los datos originales se les lleva a cabo una limpieza general, relacionados con la unicidad, completitud, consistencia y validez. Adicionalmente, se desarrolló una vectorización a través de tf-idf; precisamente empleando unígramas y bigramas, descartando términos en menos del 1% y en más del 80% de las filas, y normalizando acentos.</p>
	<p>MONITOREO NO SE DEBE DILIGENCIAR</p> <p></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>			