

## Generalidades

<b>Oportunidad/problema Negocio</b>	<p><b>Oportunidad:</b> Crear un modelo analítico automatizado que permita al UNFPA gestionar de manera eficiente grandes cantidades de datos textuales, vinculando automáticamente las opiniones de los ciudadanos con los ODS 3, 4 y 5. Esto no solo mejoraría el uso de los recursos, sino que también le brindaría a la organización información clave y oportuna para apoyar la toma de decisiones estratégicas.</p> <p><b>Problema:</b> El UNFPA necesita una solución analítica automatizada para conectar de forma eficiente y precisa las opiniones de los ciudadanos con los ODS 3, 4 y 5. Actualmente, este proceso es manual, lo que consume mucho tiempo y recursos, y puede llevar a errores humanos. La gran cantidad de datos textuales recopilados complica aún más la situación, ya que su análisis detallado es necesario para extraer información relevante.</p>
<b>Objetivos y criterios de éxito desde el punto de vista del negocio</b>	<p><b>Objetivos del proyecto</b></p> <ul style="list-style-type: none"> <li>• Crear un modelo automatizado de análisis de textos que pueda procesar grandes cantidades de opiniones de ciudadanos y clasificarlas de manera precisa en relación con los ODS 3, 4 y 5 (Salud y bienestar, educación de calidad, igualdad de género).</li> <li>• Mejorar la eficiencia en el uso de recursos al reducir la intervención manual en el análisis de datos textuales, permitiendo que el UNFPA destine más recursos a la ejecución de políticas y programas eficaces.</li> <li>• Generar información valiosa y útil sobre las percepciones y necesidades de la población en cuanto a temas de salud, educación e igualdad de género, para apoyar la toma de decisiones estratégicas y el desarrollo de políticas.</li> <li>• Asegurar la evolución continua del modelo analítico para ajustarse a nuevos datos y tendencias, garantizando que los resultados sigan siendo relevantes y precisos con el tiempo.</li> <li>• Facilitar la transmisión de los resultados a los distintos interesados (stakeholders) del UNFPA, asegurando que los insights obtenidos del análisis sean claros y útiles para la toma de decisiones.</li> </ul> <p><b>Criterios de Éxito</b></p>

	<p>Para cada objetivo, definimos criterios específicos que permitirán evaluar el éxito del proyecto:</p> <ul style="list-style-type: none"> <li>• Para evaluar el éxito del proyecto, se han definido criterios específicos para cada uno de los objetivos. En cuanto a la <b>precisión y eficiencia del modelo analítico</b>, este debe ser capaz de clasificar las opiniones de los ciudadanos con una precisión mínima del 85% para cada uno de los ODS (3, 4 y 5), utilizando métricas como F1-Score, precisión y recall. Además, el modelo debe procesar y clasificar grandes volúmenes de opiniones (por ejemplo, 10,000) en menos de una hora.</li> <li>• En lo que respecta a la <b>reducción del uso de recursos humanos</b>, se espera que el proceso de análisis de opiniones esté automatizado en un 90%, minimizando la intervención manual en comparación con el proceso actual. El modelo también debe incorporar un sistema de reentrenamiento automático, con intervención humana mínima, que permita mejorar su rendimiento de manera continua.</li> <li>• Para asegurar la <b>obtención de insights accionables</b>, los resultados del modelo deben ser relevantes para la formulación de al menos tres nuevas políticas o iniciativas alineadas con los ODS 3, 4 y 5. Además, los resultados deben ser comunicados de manera clara y comprensible para toda la organización, logrando al menos un 90% de comprensión en las pruebas de usuario realizadas con los stakeholders.</li> <li>• En términos de <b>adaptabilidad y mejora continua del modelo</b>, este debe poder adaptarse a nuevos conjuntos de datos sin perder más del 5% de precisión tras el reentrenamiento. También se debe implementar un sistema de monitoreo para evaluar el rendimiento del modelo en tiempo real y hacer los ajustes necesarios según los cambios en los datos o las condiciones.</li> <li>• Finalmente, la <b>satisfacción de los stakeholders</b> se medirá a través de un feedback positivo de al menos el 80% en relación con la utilidad y precisión del modelo para la toma de decisiones. También se desarrollarán reportes mensuales claros que resuman los insights clave y las acciones recomendadas con base en los resultados del modelo.</li> </ul>
<b>Organización y rol dentro de ella que se</b>	<b>Organización:</b> Fondo de poblaciones de las Naciones Unidas (UNFPA)

<b>beneficia con la oportunidad definida</b>	<p>El UNFPA es una organización global que promueve la mejora de la salud reproductiva y sexual, la igualdad de género, la educación y el bienestar general de las poblaciones. En el contexto de este proyecto, el UNFPA se beneficiaría significativamente del uso de un modelo de análisis automatizado que permita clasificar y procesar eficientemente las opiniones ciudadanas sobre los ODS 3 (salud y bienestar), ODS 4 (educación de calidad) y ODS 5 (igualdad de género).</p> <ul style="list-style-type: none"> <li>• <b>Equipo de investigación y análisis de datos:</b> Los analistas de datos, científicos de datos y especialistas en investigación del UNFPA se verán beneficiados por el modelo, ya que actualmente dedican gran parte de su tiempo a procesar manualmente grandes cantidades de datos textuales. El modelo automatizado aumentaría la precisión y eficiencia de su trabajo, liberando tiempo para centrarse en análisis más profundos y en la planificación de estrategias más complejas.</li> <li>• <b>Equipo de políticas y programas:</b> Los especialistas en políticas y planificadores de programas del UNFPA podrán aprovechar los insights generados por el modelo para ajustar y desarrollar políticas que respondan a las preocupaciones reales de la ciudadanía. La información procesada les permitirá identificar áreas prioritarias de intervención y diseñar soluciones más efectivas y basadas en datos actualizados.</li> <li>• <b>Equipo de comunicación y participación ciudadana:</b> Los gerentes de comunicaciones y coordinadores de participación ciudadana se beneficiarán al poder utilizar los resultados del modelo para comunicar de manera más clara y eficiente las prioridades del UNFPA. Esto mejorará la transparencia, fomentará una mayor participación ciudadana y fortalecerá los vínculos entre la organización y las comunidades a las que sirve.</li> <li>• <b>Equipo de monitoreo y evaluación:</b> Los especialistas en monitoreo y evaluación podrán beneficiarse del análisis automatizado al obtener una evaluación rápida y precisa del impacto de las políticas y programas implementados. La correlación entre las opiniones ciudadanas y los resultados esperados en los ODS les permitirá realizar ajustes en tiempo real para optimizar los resultados y mejorar la eficacia de las intervenciones.</li> <li>• <b>Directivos y tomadores de decisiones:</b> Los altos directivos y responsables de la toma de decisiones estratégicas dentro del UNFPA utilizarán la información derivada del modelo para definir estrategias más</li> </ul>
--	---

	<p>informadas y basadas en datos. Esto les permitirá optimizar recursos, alinear sus decisiones con los objetivos estratégicos del UNFPA y maximizar el impacto de las intervenciones.</p>
<p><b>Impacto que puede tener en Colombia este proyecto</b></p>	<ul style="list-style-type: none"> <li>• <b>Mejora en la toma de decisiones basada en datos:</b> La implementación de un modelo analítico que procese automáticamente grandes volúmenes de opiniones ciudadanas permitirá a entidades gubernamentales, ONGs y organizaciones locales en Colombia tomar decisiones más informadas y basadas en datos reales. Esto facilitará una asignación más precisa de recursos y la ejecución más efectiva de políticas públicas. Por ejemplo, al analizar opiniones sobre el acceso a servicios de salud, el gobierno puede guiar la planificación de infraestructura sanitaria en áreas rurales, asegurando que las comunidades más vulnerables reciban atención prioritaria.</li> <li>• <b>Identificación rápida de problemas y necesidades locales:</b> El proyecto permitirá detectar rápidamente las principales preocupaciones de la población en cuanto a salud, educación e igualdad de género. Esta capacidad de respuesta rápida es fundamental para abordar problemas emergentes antes de que se conviertan en crisis. Por ejemplo, si el modelo identifica un aumento en las quejas sobre la calidad de la educación en una región específica, el gobierno puede intervenir rápidamente con programas de mejora escolar.</li> <li>• <b>Fortalecimiento de políticas de igualdad de género:</b> Gracias a la capacidad del modelo para analizar y categorizar opiniones relacionadas con la igualdad de género, se podrán identificar las barreras y desafíos específicos que enfrentan mujeres y niñas en distintas regiones de Colombia. Esto ayudará a diseñar políticas y programas más efectivos para promover la igualdad de género. Por ejemplo, la identificación de problemas de discriminación laboral o violencia de género puede conducir a campañas de concienciación más focalizadas y a la implementación de políticas de protección más estrictas.</li> <li>• <b>Desarrollo de capacidades y transferencia de conocimientos:</b> La implementación del proyecto impulsará el desarrollo de capacidades locales en áreas como análisis de datos, inteligencia artificial y ciencia de datos, beneficiando no solo al UNFPA y a sus colaboradores directos, sino también creando una plataforma para el crecimiento del talento local en estas</li> </ul>

	<p>áreas clave. Por ejemplo, las colaboraciones con universidades y centros de investigación locales permitirán capacitar a nuevos analistas de datos y científicos que podrán liderar futuros proyectos de inteligencia de negocios en Colombia.</p>
<p><b>Enfoque analítico. Descripción de la categoría de análisis (descriptivo, predictivo, etc.), tipo y tarea de aprendizaje e incluya las técnicas y algoritmos que propone utilizar</b></p>	<p><b>1. Descripción de la categoría de análisis</b> El análisis predictivo es el más apropiado para este proyecto, ya que su objetivo es predecir la categoría de un texto en función de su contenido. En este caso, el modelo tiene que clasificar las opiniones ciudadanas en relación con los ODS 3 (salud y bienestar), ODS 4 (educación de calidad), y ODS 5 (igualdad de género). Para lograrlo, se utiliza un enfoque de aprendizaje supervisado, ya que contamos con datos etiquetados que servirán para entrenar el modelo. Esto permitirá que el modelo aprenda a clasificar textos nuevos dentro de las categorías definidas.</p> <p><b>2. Tarea de aprendizaje</b> La principal tarea del modelo es la clasificación multiclase, donde el objetivo es asignar cada opinión de texto a una de las tres categorías: ODS 3, 4 o 5. Este enfoque permitirá identificar automáticamente las preocupaciones de los ciudadanos en estas áreas clave.</p> <p><b>3. Técnicas y algoritmos propuestos</b> Para la construcción del modelo de clasificación de textos, proponemos utilizar una combinación de técnicas de preprocesamiento de texto y algoritmos de aprendizaje automático.</p> <ul style="list-style-type: none"> <li>- <b>Preprocesamiento de texto:</b> <ul style="list-style-type: none"> <li>o <b>Eliminación de stopwords:</b> Quitar palabras comunes que no aportan significado significativo (por ejemplo, "y", "o", "pero").</li> <li>o <b>Stemming/lematización:</b> Reducir las palabras a su forma raíz para normalizar el texto.</li> <li>o <b>Vectorización:</b> Convertir el texto en una representación numérica. Esto puede hacerse utilizando técnicas como TF-IDF.</li> </ul> </li> <li>- <b>Algoritmos de aprendizaje automático:</b> <ul style="list-style-type: none"> <li>o <b>Máquina de soporte vectorial (SVM):</b> Ideal para la clasificación de texto, ya que maneja bien datos de alta dimensionalidad y funciona eficientemente con conjuntos de datos medianos.</li> <li>o <b>Naive Bayes multinomial:</b> Una elección popular para la clasificación de textos,</li> </ul> </li> </ul>

	<p>especialmente cuando se trabaja con datos textuales que representan conteos de palabras o frecuencias. Es eficaz y rápido para entrenar y predecir.</p> <ul style="list-style-type: none"> <li>○ <b>XGBoost:</b> se basa en la idea de <b>boosting</b>, que combina varios modelos débiles (generalmente árboles de decisión) en un modelo fuerte. La clave está en corregir los errores cometidos por modelos anteriores, con el objetivo de mejorar progresivamente las predicciones. XGBoost utiliza una técnica de optimización de gradiente para ajustar los errores y reducir el sesgo en las predicciones, lo que lo hace muy efectivo para tareas de clasificación complejas.</li> </ul> <p>- <b>Evaluación de modelos:</b></p> <ul style="list-style-type: none"> <li>○ <b>Métricas de evaluación:</b> Utilizar métricas como F1-Score, precisión, recall y la matriz de confusión para evaluar la efectividad de los modelos.</li> <li>○ <b>Validación cruzada:</b> Implementar una validación cruzada para asegurar que el modelo no está sobreajustado y puede generalizar bien con datos no vistos.</li> </ul>
--	--

## Mapa de actores relacionado con el producto de datos creado (Modelo SVM sobre ODS 3, 4 y 5)

Rol dentro de la organización	Tipo de actor	Beneficio	Riesgo
Coordinación de proyectos sociales	Usuario-cliente	El modelo facilita la identificación de áreas clave para implementar acciones en torno a los ODS, permitiendo a la coordinación tomar decisiones rápidas y basadas en datos. Esto ayuda a priorizar temas de	Si el modelo no clasifica correctamente las opiniones, podrían priorizarse problemas menos urgentes, lo que conllevaría a una mala asignación de recursos y tiempo,

		salud, educación y género que tienen un mayor impacto.	afectando la efectividad de políticas y programas.
<b>Dirección de responsabilidad social</b>	Financiero	Gracias al modelo, esta área puede tomar decisiones más precisas sobre la asignación de presupuestos, dirigiendo los recursos hacia los ODS que son más relevantes para la ciudadanía y maximizando el impacto de los proyectos financiados.	Si el modelo falla, los recursos podrían malgastarse en áreas incorrectas, disminuyendo el impacto social de los proyectos y generando desconfianza en el uso de IA para la toma de decisiones.
<b>Departamento de tecnologías de la información</b>	Proveedor	Este departamento garantiza la correcta implementación y seguridad del sistema que soporta el modelo, asegurando un procesamiento seguro y eficiente de los datos de opiniones ciudadanas.	Un mal manejo de los datos podría generar problemas de privacidad, dañando la reputación de la organización. Además, si el sistema falla, el análisis de las opiniones ciudadanas se vería afectado.
<b>Analistas de políticas públicas</b>	Proveedor	Los analistas pueden utilizar los resultados del modelo para conocer las principales preocupaciones de la ciudadanía respecto a los ODS, permitiendo crear políticas públicas más ajustadas a las necesidades reales. El modelo también	Si el modelo produce resultados inexactos o sesgados, las políticas propuestas podrían estar mal fundamentadas, afectando a las comunidades. La falta de confianza en el

		ahorra tiempo en el análisis manual de grandes volúmenes de datos.	modelo podría hacer que los analistas no confíen en las herramientas automáticas.
<b>Ciudadanos (participantes de encuestas)</b>	Beneficiario	Los ciudadanos pueden recibir respuestas más rápidas y personalizadas sobre sus inquietudes en temas de salud, educación e igualdad de género, influyendo directamente en las políticas públicas.	Si el modelo no clasifica bien las opiniones, es posible que algunos problemas importantes no sean atendidos, lo que podría generar una percepción de desatención o desigualdad en la asignación de recursos.
<b>Organizaciones no gubernamentales (ONG's)</b>	Usuario-cliente	Las ONGs pueden enfocar mejor sus esfuerzos y acciones basándose en los datos clasificados por el modelo, alineándose con las áreas de mayor necesidad.	Predicciones incorrectas podrían llevar a que las ONGs destinen recursos a problemas que no son tan urgentes, disminuyendo la efectividad de sus acciones.
<b>Ministerio de salud/educación/igualdad</b>	Usuario-cliente	Estas instituciones públicas pueden planificar mejor sus programas y asignar recursos a áreas donde la ciudadanía muestra más preocupación, mejorando la eficiencia en la implementación de políticas relacionadas con los ODS.	Si el modelo proporciona resultados incorrectos, los recursos públicos podrían malgastarse en áreas no prioritarias, lo que afectaría la confianza pública en estas instituciones y en sus políticas.



# Trabajo en equipo

**Líder de proyecto:** Andrés Felipe Pérez Martínez

**Ingeniero de datos:** John Jairo Serrato Suarez

**Ingeniera de software responsable del diseño de la aplicación y resultados:** Juana María Morales Ramos

**Ingeniero de software responsable de desarrollar la aplicación final:** Andrés Felipe Pérez Martínez

## Tabla de reuniones:

Tipo de reunión	Descripción	Horas dedicadas
Reunión de lanzamiento y planeación	Definir roles, lluvia de ideas y planeación de tiempos.	2 horas
Reunión de trabajo sobre la preparación de datos.	Se realizará la primera parte del trabajo, siendo esta el perfilamiento y la preparación de datos.	3 horas
Reunión para definir los modelos.	Cada integrante presenta posibles modelos y entre todos se escogen los 3 mejores y se define quien lo realizará.	30 minutos
Reunión de seguimiento.	Avances del proyecto, revisión de entregables y designación de tareas faltantes.	1 hora
Reunión de grabación.	Grabación del video de resultados y revisión de la completitud de los entregables.	2 horas
Reunión final.	Consolidación del proyecto, revisión de resultados y retroalimentación.	1 hora

**Tabla de tareas:**

<b>Tarea</b>	<b>Descripción</b>	<b>Integrante</b>	<b>Horas dedicadas</b>
Preparación de los datos	Arreglar los datos para usar en cada modelo.	Andrés, Juana	3 horas
Implementar modelo	Cada miembro desarrollará su modelo y presentara los resultados.	Todos	4 horas
Elegir mejor modelo	Según los resultados de cada modelo, se elegirá el mejor para continuar el desarrollo y se justificará el por qué.	Todos	30 minutos
Realizar la clasificación de los datos nuevos.	Según el mejor modelo, se clasificaran los nuevos datos.	Andrés (Líder)	2 horas
Entregas mapa de actores y generalidades.	Incluir estos ítems a el documento.	Juana	2 horas
Organización de tareas, git y seguimiento.	Incluir estos ítems del proyecto.	John	1 hora
Grabación del video.	Se grabará un video de máximo 5 minutos para la presentación de resultados.	Todos	1 hora

**Tabla Puntaje:**

<b>Integrante</b>	<b>Descripción</b>	<b>Puntuación</b>
Andrés Pérez	Alto esfuerzo en el desarrollo del código.	35
Juana Morales	Esfuerzo en el código y en el desarrollo de los entregables.	35
John Serrato	Buen trabajo, pero menor carga.	30