|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| TAREA DE APRENDIZAJE    es supervisado, se busca predecir la categoria para un texto dado, los posibles resultados están dados por la columna labels, siendo ODS 1, ODS 3 y ODS 4. Los resultados se observan justo al instante de hacer el modelo y pasar a categorizar los datos (resultado instantáneo) | DECISIONES  Los resultados adquiridos se convierten en una gran fuente de valor para tomar decisiones más basadas en la opinión ciudadana, generar reportes de la distribución de texto y mantener un buen seguimiento de las políticas públicas. | PROPUESTA DE VALOR  El beneficiario final es la UNFPA y todas las empresas en si que usen información ciudadana para temas políticos. Los problemas específicos que se abordan es el tener que analizar manualmente grandes cantidades de información, ya sea por costo o por tiempo. Los posibles riesgos que puede tener un beneficiario al hacer uso del modelo es una clasificación incorrecta de ODS o sesgos dentro de los datos. | RECOLECCIÓN DE DATOS – NO SE DEBE DILIGENCIAR  ¿Cómo se obtiene el conjunto inicial de entidades y resultados (por ejemplo, extractos de bases de datos, extracciones de API, etiquetado manual)? ¿Qué estrategias se aplican para actualizar los datos continuamente, controlando los costos y manteniendo la vigencia? | FUENTES DE DATOS  Las fuentes de datos utilizadas son las fuentes de datos que son proporcionadas ya sea en internet como en bases de datos de la propia UNFPA, todas teniendo en cuenta una relación con los temas políticos. Es necesario usarlos para hacer el objetivo de análisis, ya que son los datos que vamos a estudiar, examinar y plantear para el modelo entrenado. |
| SIMULACIÓN DE IMPACTO  ¿Existen restricciones de equidad?  Las decisiones correctas implican un ahorro inmenso de tiempo y de recursos sin necesidad de hacer un análisis extenso de cada fuente, en contraparte una decisión incorrecta implica malinterpretación del contenido y puede implicar costes en el ambiente. Dicho esto, se tienen en cuenta unos criterios de éxitos, siendo que la precisión del modelo y la métrica ROC AUC sea mayor a 0,8. | APRENDIZAJE (USO DEL MODELO)  Es un modelo usado en tiempo real que permite analizar cualquier información nueva que se presente, la frecuencia de uso de este modelo varia de si se quiere usar cada vez que se recibe un nuevo texto o si se hace con una gran cantidad de textos. |  | CONSTRUCCIÓN DE MODELOS  ¿Cuántos modelos se necesitan? ¿Cuándo deben actualizarse? ¿De cuánto tiempo se dispone para generar el modelo (incluido el proceso de ingeniería de características y el análisis o evaluación del mismo)? | INGENIERÍA DE CARACTERÍSTICAS  ¿Qué variables/características se utilizan en el modelo? ¿Qué agregaciones o transformaciones se aplican a las fuentes de datos originales – incluir las más importantes--?  Las variables utilizadas en el modelo fueron los textos, estos fueron analizados y transformados en un proceso de calidad (lematización) para poder usarlos dentro del modelo clasificatorio. Adicionalmente se realizo una vectorización TF-IDF para convertir los textos lematizados a números y poder analizarlos. |
|  | MONITOREO NO SE DEBE DILIGENCIAR  ¿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? |  |  |  |

# **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://www.ownml.co/) when sharing. | [**OWNML.CO**](https://www.ownml.co/) |

# 