











Adaptación de OWNML MACHINE LEARNING CANVAS

<div>TAREA DE APRENDIZAJE </div> <div>Es aprendizaje supervisado, más específicamente un modelo de clasificación con datos etiquetados. Los posibles resultados son resultados binarios siendo estos: 0 si es una noticia falsa y 1 si es una noticia verdadera. Los resultados se obtienen en minutos.</div>	<div>DECISIONS </div> <div>Los datos tienen que estar en español y ser noticias con un título y descripción, estos dos se deben concatenar y deben pasar por un proceso de token y "lemming". Luego estos tokens deben ser vectorizados para que ya los modelos de aprendizaje puedan retornar resultados.</div>	<div>PROPUESTA DE VALOR </div> <div><p>Los beneficiarios serían la sociedad en general, más en específico hispanohablante. Otro subgrupo serían las instituciones gubernamentales. El problema por resolver es la polarización y manipulación de la opinión pública a nivel internacional y nacional debido a la desinformación y noticias falsas.</p><p>Un posible riesgo para los usuarios es una excesiva seguridad sobre el modelo de clasificación y que se tomen como correctas todas sus respuestas, teniendo en cuenta que el modelo puede no estar preparado para la variabilidad de la desinformación y pueda tener sesgos no detectados.</p></div>	<div>RECOLECCIÓN DE DATOS </div> <div>Se ordenó no diligenciar este apartado.</div>	<div>FUENTES DE DATOS </div> <div><p>Un grupo de académicos se encargaron de la recolección y creación de estos datos extraídos de periódicos en línea en español. Los datos originales fueron extraídos de los periódicos 'Público,' 'La Marea,' y 'El Común.' y tienen licencia ATTRIBUTION 4.0 INTERNATIONAL. Los datos son útiles para cumplir el objetivo ya que son periódicos reales que podemos tomar de referencia y los datos fueron extraídos de forma que el procesamiento de texto es fácil y estandarizado.</p></div>
<div>SIMULACIÓN DE IMPACTO </div> <div>El costo de las decisiones incorrectas incluye la perpetuación de la desinformación y sus consecuencias sociales y el beneficio de las decisiones correctas es el deceso de la desinformación y fortalecimiento de la confianza en el periodismo en línea. Existen restricciones de equidad ya que el modelo no debe presentar sesgos y debe mantenerse objetivo para no apoyar o atacar injustamente a grupos y su postura.</div>	<div>APRENDIZAJE (USO DEL MODELO) </div> <div>El uso del modelo es por lotes, pueden ingresar cualquier cantidad de estos y se debe ejecutar cada vez que se quiera identificar si una noticia es real o falsa.</div>		<div>CONSTRUCCIÓN DE MODELOS </div> <div>Se solicitaron 3 modelos con algoritmos distintos, para posteriormente ser comparados y elegir el mejor de estos como modelo definitivo. El modelo no aspira a actualizarse por el momento. Y el tiempo de desarrollo del modelo es de 15 días, siendo la liberación de datos el 7 de febrero y la entrega el 22 de febrero.</div>	<div>INGENIERÍA DE CARACTERÍSTICAS </div> <div>Las variables de la fuente datos usadas para el modelo son la etiqueta, el título y la descripción de la noticia, que se concatenaron en otra columna y posteriormente pasan por un proceso de tokenización y un proceso de lematización. Por último, estos tokens se vectorizaron para generar los distintos modelos.</div>
	<div>MONITORING </div> <div>Se ordenó no diligenciar este apartado.</div>			



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 when sharing.

[OWNML.CO](https://ownml.co)