




<p><b>TAREA DE APRENDIZAJE</b></p> <p></p> <p>Se realiza un aprendizaje supervisado ya que se cuenta con datos etiquetados en el dataset indicando si la noticia es falsa o verdadera. Es necesario predecir si una noticia es falsa o verdadera. Cuando el modelo ya ha sido entrenado, se toma muy pocos segundos para realizar la predicción cuando este se ejecuta sobre un nuevo conjunto de datos, sin embargo, el reentrenamiento del modelo se podría demorar bastantes minutos dependiendo de la cantidad de datos y el algoritmo elegido.</p>	<p> <b>DECISIONES</b></p> <p>Para transformar los resultados del modelo en recomendaciones para el usuario se pueden realizar varias acciones. Como mostrar mensajes de alerta si se encuentra una noticia falsa o si hay sospechas de la veracidad de la noticia se pueden mostrar enlaces confiables relacionados con el tema. Ya que el modelo no es 100% acertado, cuando se encuentre una noticia falsa o con posibilidades de que sea falsa, se enviará a revisión manual para lograr identificar con mayor certeza si la noticia es falsa o verídica.</p>	<p> <b>PROPUESTA DE VALOR</b></p> <p>Los que se benefician de la implementación del modelo son los periodistas, medios de comunicación, plataformas de redes sociales y en general empresas que trabajen con la publicación y uso de la información que se publica. Algunos de los problemas a los que toca enfrentarse son la desinformación y falsas noticias en áreas como la política, ya que se puede manipular la opinión pública e influir en procesos electorales y dañar la imagen de algunas instituciones. Otro problema es la dificultad para verificar las noticias falsas debido a la cantidad de información incorrecta que está circulando por los medios de comunicación. El uso de este modelo puede traer algunos riesgos como los falsos positivos o falsos negativos, es decir noticias que pueden ser clasificadas de manera incorrecta, ya sea como verdaderas o falsas. Otro de los riesgos sería el sesgo en los datos, ya que si se entrena el modelo con datos inclinados hacia un sector de la política se puede favorecer este sector por encima de los demás.</p>	<p> <b>RECOLECCIÓN DE DATOS – NO SE DEBE DILIGENCIAR</b></p> <p>¿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?</p>	<p> <b>FUENTES DE DATOS</b></p> <p>Los datos provienen de documentos csv con toda la información de noticias organizadas por columnas como ID, fecha, titulo y descripción. Esta información fue obtenida a partir de fuentes externas como periódicos, redes sociales y demás medios de comunicación que publican noticias. Los datos de los archivos pueden ser usados para realizar un análisis, sin embargo, primero es necesario realizar una limpieza y un procesamiento de estos.</p>
------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

<div><div>SIMULACIÓN DE IMPACTO</div></div> <p>Algunos de los beneficios dados por las decisiones tomadas son, la automatización del análisis de datos, la clasificación de los textos, la toma de decisiones fundamentada en estadísticas y la escalabilidad. Sin embargo, los costos asociados son los falsos positivos o falsos negativos, los recursos computacionales que se consumen dependiendo del modelo y la dependencia de los datos de entrada, ya que si estos están desbalanceados pueden afectar la toma de decisiones.</p>	<div><div>APRENDIZAJE (USO DEL MODELO)</div></div> <p>Elegimos el modelo Random Forest, el cual principalmente se ejecuta por lotes, ya que requiere múltiples evaluaciones, de manera que procesa grandes volúmenes de datos en intervalos definidos. La frecuencia de actualización depende de la variabilidad de los datos, si se cambian los datos, es recomendable actualizar el modelo cada cierto tiempo.</p>		<div><div>CONSTRUCCIÓN DE MODELOS</div></div> <p>Se realizaron 3 modelos, sin embargo, seleccionamos el modelo de Random Forest, para quedarse solamente con este es necesario monitorear su desempeño. Para actualizar el modelo es necesario esperar a que los datos cambien, aunque es recomendable actualizarlo cada cierto tiempo. Además, hay que reentrenar el modelo si la tendencia de los datos cambia o su desempeño disminuye.</p>	<div><div>INGENIERÍA DE CARACTERÍSTICAS</div></div> <p>El modelo de Random Forest utiliza una combinación de variables clave seleccionadas a partir del análisis de datos, incluyendo características numéricas y categóricas relevantes para la predicción. Se aplican transformaciones como normalización o codificación de variables categóricas, y se pueden generar variables derivadas para mejorar la precisión. Además, se utilizan técnicas de manejo de valores nulos y selección de características para optimizar el desempeño del modelo.</p>
	<div><div>MONITOREO NO SE DEBE DILIGENCIAR</div><div></div></div> <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>			

## 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-nc-sa/4.0/).

Please keep this mention and the link to [ownml.co](https://ownml.co) when sharing.

**OWNML.CO**