






<p><b>TAREA DE APRENDIZAJE</b></p> <p></p> <p>¿Cuál es el tipo de aprendizaje? ¿Si es aprendizaje supervisado, indicar qué se predice? ¿Cuáles son los posibles resultados de la tarea de aprendizaje? ¿Cuándo se observan los resultados de esta tarea? Por ejemplo, si es un modelo predictivo indicar si el resultado se obtiene unas horas, días, semanas, o meses antes.</p> <p>Para esta tarea nos interesa realizar una tarea de clasificación es decir de aprendizaje supervisado. Por medio del método de clasificación queremos poder predecir si una noticia es falsa o no de acuerdo con sus características. Para</p>	<p> <b>DECISIONES</b></p> <p>¿Cómo se convierten los resultados del modelo en recomendaciones o decisiones procesables para el usuario final?</p> <p>Para esta tarea nos interesa realizar una tarea de clasificación es decir de aprendizaje supervisado. Por medio del método de clasificación queremos poder predecir si una noticia es falsa o no de acuerdo con sus características. Para este contexto nos centraremos en clasificar cada noticia como 0 o 1, donde 0 indica que la noticia es falsa y 1 que la noticia es verdadera. Para poder observar los resultados de esta tarea primero nos toca utilizar un conjunto de noticias y lo ideal es obtener la clasificación para todas las noticias en un tiempo no superior a 10 minutos</p>	<p> <b>PROPUESTA DE VALOR</b></p> <p>¿Quién es el beneficiario final? ¿De qué empresa es? ¿Qué problemas específicos se abordan? ¿Qué riesgo puede tener para ese beneficiario el uso de este modelo?</p> <p>El beneficiario final son todas las personas que leen noticias y que no pueden saber a ciencia cierta si las noticias que están leyendo son verdaderas o falsas. Podríamos decir entonces que el público en general se podría beneficiar de esta actividad ya que les permitiría el poder separar las noticias que son verdaderas de aquellas que no lo son. Adicionalmente a eso, el poder separar las noticias falsas de las verdaderas permite que las personas puedan tener certeza de que la información que están leyendo es verídica y confiable. No obstante, uno de los posibles problemas es que este procedimiento no es</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>¿Qué fuentes de datos se utilizan? (Mencione tablas de bases de datos internas y externas o métodos API). ¿De dónde se toman los datos? ¿Se pueden utilizar para realizar el objetivo del análisis?</p> <p>Los datos que se utilizan para poder realizar el modelo son noticias de diferentes periódicos como lo son "Público", "La Marea" y "El Común". Estos datos vienen compuestos por un título, descripción y fecha de publicación de la noticia, adicionalmente a eso incluyen un índice que permite conocer si la noticia es falsa o no. De modo que, por medio de este índice, el título de la noticia y su descripción se podrá realizar la construcción del modelo para poder categorizar futuras noticias.</p>
---	--	---	--	--

<p>este contexto nos centraremos en clasificar cada noticia como 0 o 1, donde 0 indica que la noticia es falsa y 1 que la noticia es verdadera. Para poder observar los resultados de esta tarea primero nos toca utilizar un conjunto de noticias y lo ideal es obtener la clasificación para todas las noticias en un tiempo no superior a 60 minutos</p>		<p>una herramienta permanente, ya que solo se encarga de predecir el tipo de una noticia basándose en datos históricos y que siempre puede haber un margen de error, por mínimo que sea, además este modelo es susceptible a modificaciones, lo que harían que la veracidad de este se vea reducida.</p>		
<p> <b>SIMULACIÓN DE IMPACTO</b></p> <p>¿Cuáles son los valores de costo/beneficio de las decisiones (in)correctas? ¿Cuáles son los criterios de éxito del modelo para su posterior despliegue? ¿Existen restricciones de equidad?</p> <p>La relación costo/beneficio de la toma de decisiones</p>	<p> <b>APRENDIZAJE (USO DEL MODELO)</b></p> <p>¿El uso del modelo es por lotes o en tiempo real? ¿Con qué frecuencia se usa?</p> <p>Lo ideal sería poder realizar la ejecución del modelo en tiempo real para poder saber si una noticia es falsa o verdadera apenas se publica, obviamente esto requiere una gran infraestructura computacional que se encuentre disponible lo más que se pueda. La importancia de que se conozca el tipo de noticia apenas salga es para</p>		<p> <b>CONSTRUCCIÓN DE MODELOS</b></p> <p>¿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)?</p> <p>Para la construcción del modelo se deben</p>	<p> <b>INGENIERÍA DE CARACTERÍSTICAS</b></p> <p>¿Qué variables/características se utilizan en el modelo? ¿Qué agregaciones o transformaciones se aplican a las fuentes de datos originales – incluir las más importantes--?</p> <p>En cuanto a la ingeniería, para el desarrollo de este tendremos que enfocarnos en buscar un patrón de palabras que se encuentren en las noticias falsas, palabras que busquen llamar la atención del lector son las más comunes. Por su parte, para tratar los datos tenemos que eliminar palabras que no aportan un valor agregado al texto, palabras como los artículos, conectores, prefijos y sufijos solamente sirven para rellenar y no nos proporcionan información valiosa para poder identificar y entender el contenido del texto en su totalidad. También es importante evitar los tiempos verbales y mantener los verbos de ser posible en su forma de infinitivo para poder asimilar varios verbos en varios documentos, es decir que podamos clasificar las noticias teniendo en</p>

<p>erróneas y correctas en la elaboración del modelo es que estas pueden afectar la exactitud con la cual una noticia es declarada como falsa. Si tomamos una decisión radical de excluir algunos elementos particulares de las noticias podríamos terminar sesgando el modelo y algunas noticias falsas podrían terminar pasando como verdaderas. El principal criterio de éxito del modelo a producir es que queremos que el porcentaje de aceptación de noticias falsas sea lo más cercano a cero, es decir que el porcentaje de detección de noticias verdaderas sea lo más alto posible. En cuanto a sus restricciones de equidad lo ideal es que no se puede descartar una noticia por tener palabras que comúnmente se</p>	<p>poder de ese modo evitar un aumento en la desinformación por parte de los diferentes medios.</p>	<p>emplear tres algoritmos de clasificación diferentes para que de ese modo se puedan clasificar las noticias por medio de tres procesamientos diferentes y que de ese modo la clasificación no dependa únicamente de lo que dicta un modelo, para de ese modo evitar un posible sesgo que se puede dar durante la clasificación. Por su parte, estos modelos deberían actualizarse cada mes utilizando las noticias de meses pasados para poder entrenarlo y que de ese modo se mantenga actualizado en cuanto a las tendencias y diferentes formas de desinformación. Por último, en cuanto al tiempo de construcción, lo ideal es poder tener el modelo en funcionamiento lo más rápido posible para evitar la propagación de la desinformación.</p>	<p>cuenta que el verbo jugar que se puede encontrar como jugando o jugará puedan ser procesados por el modelo como el propio verbo jugar.</p>
---	---	---	---

<p>encuentran en noticias falsas, es decir que por ejemplo no podemos descartar todas las noticias que hablan sobre figuras políticas y declararlas como falsas porque eso sesgaría el modelo y no nos estaría dando una clasificación aproximada.</p>				
	<p><b>MONITOREO NO SE DEBE DILIGENCIAR</b></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>			

## 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://ownml.co) when sharing.

**OWNML.CO**

