






<p>TAREA DE APRENDIZAJE</p> <p></p> <p>Este proyecto utiliza aprendizaje supervisado, ya que el dataset contiene noticias etiquetadas como falsas (1) o verdaderas (0), lo que permite al modelo aprender patrones a partir de estos ejemplos. El objetivo es predecir si una noticia es falsa o verdadera en función de su título y descripción. Los posibles resultados son una clasificación binaria donde 0 indica una noticia verdadera y 1 una noticia falsa, pudiendo incluir una probabilidad asociada. Durante el entrenamiento, los resultados se observan en minutos u horas según el tamaño del dataset, mientras que</p>	<p> DECISIONES</p> <p>Los resultados del modelo se convierten en recomendaciones procesables al clasificar cada noticia como falsa (1) o verdadera (0) y presentar esta información de manera clara para el usuario final. Esto puede lograrse a través de un sistema de alertas o etiquetas que indiquen el nivel de confiabilidad de una noticia. Además, se pueden generar explicaciones sobre por qué una noticia fue clasificada como falsa, resaltando palabras clave o patrones detectados. Para medios de comunicación o plataformas de verificación, el modelo puede integrarse en herramientas de monitoreo que ayuden a filtrar contenido sospechoso antes de su publicación o difusión. En el caso de gobiernos o investigadores, los resultados pueden alimentar informes analíticos que ayuden a identificar tendencias de desinformación y a tomar decisiones estratégicas para combatirlas.</p>	<p> PROPUESTA DE VALOR</p> <p>El beneficiario final de este modelo puede ser gobiernos, medios de comunicación, plataformas de verificación de noticias, investigadores o el público en general, dependiendo de su aplicación. Si es un proyecto académico, su propósito sería mejorar la detección de noticias falsas para estudios sobre desinformación. Los problemas que aborda incluyen la propagación de noticias falsas, la manipulación de la opinión pública, la polarización social y la pérdida de confianza en instituciones. Sin embargo, su uso conlleva riesgos como falsos positivos y negativos, sesgos en los datos que pueden favorecer ciertos discursos, la posible censura de información legítima y un impacto en la libertad de</p>	<p> RECOLECCIÓN DE DATOS – NO SE DEBE DILIGENCIAR</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> FUENTES DE DATOS</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 provienen de un conjunto recolectado por académicos para entrenar modelos de detección de noticias falsas, el cual incluye artículos de Público, La Marea y El Común.</p> <p>Son adecuados para el análisis ya que incluyen noticias etiquetadas como verdaderas o falsas, para entrenar un modelo de clasificación binaria.</p>
---	---	---	--	---

<p>en producción el modelo puede realizar predicciones en tiempo real o en cuestión de segundos al analizar nuevas noticias.</p>		<p>expresión si no se implementa con criterios éticos adecuados.</p>		
<p> SIMULACIÓN DE IMPACTO</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> APRENDIZAJE (USO DEL MODELO)</p> <p>El modelo se usará por lotes, lo que significa que procesará conjuntos de noticias en intervalos definidos en lugar de analizar cada noticia en tiempo real. La frecuencia de uso dependerá de la disponibilidad de nuevos datos y la necesidad de actualizar los análisis, pudiendo ejecutarse diariamente, semanalmente o mensualmente según el caso.</p>		<p> CONSTRUCCIÓN DE MODELOS</p> <p>Para este proyecto se necesitan tres modelos de aprendizaje supervisado: Decision Tree, KNN y Random Forest, los cuales serán comparados para determinar cuál ofrece mejor desempeño en la detección de noticias falsas. Estos modelos deben actualizarse</p>	<p> INGENIERÍA DE CARACTERÍSTICAS</p> <p>Las principales variables utilizadas en el modelo son el título de la noticia, el contenido o descripción y la etiqueta de veracidad (donde 0 indica noticia verdadera y 1 indica noticia falsa). Estas variables permiten entrenar un modelo de clasificación binaria que diferencie entre noticias falsas y verdaderas a partir de su contenido textual. Para mejorar la calidad de los datos antes de entrenar el modelo, se aplican varias transformaciones. Primero, se realiza una limpieza de texto, eliminando caracteres especiales, números y signos de puntuación que no aportan valor al análisis. Luego, se aplica una normalización, convirtiendo todo el texto a minúsculas para evitar diferencias en el procesamiento de palabras idénticas con diferentes formatos. También se eliminan stopwords, es decir, palabras comunes que no aportan significado relevante (como "de", "la", "en"). Además, se utiliza lematización o stemming para reducir las palabras a su forma base y mejorar la representación semántica del texto (por ejemplo, "corriendo" se transforma en "correr"). Finalmente, el texto se convierte en una representación numérica mediante vectorización, utilizando técnicas como TF-IDF o word embeddings, lo que permite que los modelos de aprendizaje automático (Decision Tree, KNN y</p>

		<p>periódicamente, idealmente cada tres a seis meses, para adaptarse a nuevas tendencias de desinformación y cambios en los datos. El tiempo disponible para generar el modelo, incluyendo la ingeniería de características, entrenamiento, análisis y evaluación en general puede tomar entre uno y tres meses para garantizar un modelo optimizado y listo para pruebas en producción.</p>	<p>Random Forest) puedan analizar el contenido de manera estructurada y mejorar su capacidad de clasificación.</p>
	<div><div><div><div><div><div>MONITOREO NO SE DEBE DILIGENCIAR</div><div></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></div></div></div></div>		

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 when sharing.

OWNML.CO