# OWNML MACHINE LEARNING CANVAS

# Nicolas Arango - 202220342 Mateo Rincon- 202221402 Amalia Carbonell - 202122079

| Tipo de Aprendizaje ¿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. El tipo de aprendizaje sería aprendizaje supervisado y busca predecir si una noticia es falsa o verdadera. Los posibles resultados serían noticia falsa (1) y noticia verdadera (0). Los resultados de esta tarea se esperarían en el momento en el que se ingresen las noticias, no predice cuáles serán falsas en un futuro. | DecisionesHow are predictions turned into actionable recommendations or decisions for the end-user? (Mention parameters of the process / application for this.) El modelo convierte predicciones matemáticas en probabilidades de que una noticia sea falsa o verdadera. Esta información llega al usuario final, quien con base en el “output” del modelo puede tomar una decisión informada sobre si leer o no una noticia o al menos puede estar advertido y ser más crítico a la hora de leer noticias. | Propuesta de Valor¿Quién es el beneficiario final?¿De qué empresa es? ¿Qué problemasespecíficos se abordan? ¿Qué riesgo puedetener para ese beneficiario el uso de estemodelo? **Beneficiario Final**: Lectores de noticias, Medios de comunicación, redes sociales, agencias gubernamentales.  **Empresa**: No se tiene una empresa específica asociada con el proyecto.  **Problema Abordado**: La propagación de noticias falsas que afecta la opinión pública y la estabilidad política. Las noticias falsas polarizan las opiniones políticas de los lectores. Falta de confianza en las cadenas de noticias al no poder creer la validez de su contenido.  **Riesgos**: Falsos positivos y falsos negativos. | Recolección de datos How is the initial set of entities and outcomes sourced (e.g., database extracts, API pulls, manual labeling)? What strategies are in place to update data continuously while controlling cost and maintaining freshness? | Fuentes de datos ¿Qué fuentes de datos se utilizan?(Menciona 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? **Bases de datos utilizadas:**  Internas :   * fake\_new\_spanish.csv * fake\_news\_test.csv   Externas:  Si se pueden utilizar para realizar el objetivo del análisis pero se debería probar al final con noticias falsas de otra base de datos para garantizar que los resultados no sean por aprendizaje de los datos. |
| --- | --- | --- | --- | --- |
| Simulación de Impacto ¿Cuáles son los valores decosto/beneficio de lasdecisiones (in) correctas? ¿Cuáles son los criterios de éxito del modelo para su posterior despliegue? ¿Existen restricciones de equidad? Beneficios de las decisiones correctas:   * Acceso a información confiable y real para los ciudadanos * Confianza en un medio con noticias reales.   Costo de las decisiones incorrectas:   * Daño a la reputación de los medios ya sea por esparcir noticias falsas o por clasificar noticia verdadera como falsa. * Propagación de desinformación y noticias falsas | Aprendizaje (uso del modelo) Are predictions made in batch or in real time? How frequently? How much time is available for this (including featurization and decisions)? Which computational resources are used? Las predicciones se realizan en batch. Esto sucede por varios motivos. Uno de ellos es la capacidad computacional que tenemos a disposición. Hay bloques de código que pueden durar entre 4 y 5 horas con las máquinas personales que tenemos. Por esto no se pueden hacer predicciones en el momento que salga una nueva noticia. Estas predicciones se pueden generar diariamente durante horas de la noche, cuando no hay menor cantidad de gente leyendo noticias. Sin embargo, ya que la información corre muy rápido, es necesario mantener el modelo al día lo más eficientemente posible. |  | Construcción de Modelos How many models are needed in production? When should they be updated? How much time is available for this (including featurization and analysis)? Which computation resources are used? Para la producción en sí solo es necesario un modelo. Sin embargo, esto puede llevar a decisiones muy vagas y poco acertadas. Es ideal tener al menos 3 modelos diferentes para ir comprobando cuál es el que mejor se ajusta a las necesidades del problema. Estos modelos deben ser actualizados lo más frecuentemente posible para tener más y más datos de entrenamiento y de esta manera lograr predecir las fake news de mejor manera. Sin embargo, un PC no es suficiente para llevar a cabo todas estas tareas diariamente ya que son operaciones con gran costo computacional y temporal. | Ingeniería de características What representations are used for entities at prediction time? What aggregations or transformations are applied to raw data sources? Variable : Para la predicción del modelo, se utiliza la variable Label en la cual se asigna el valor de 0 en caso de que la noticia sea falsa y 1 en caso de que sea verdadera.  Características: Se utilizaron las palabras claves encontradas tanto en la descripción como en el título de la noticia.  Transformaciones: Se realizaron un par de transformaciones sobre los datos para garantizar que estuvieran limpios y sin datos atípicos:   1. Se eliminaron las stop words o palabras irrelevantes. 2. Se eliminaron las filas o noticias duplicadas 3. Se eliminó la columna de fecha pues no era relevante en la detección de noticias falsas. 4. Normalización del texto 5. Vectorización del texto con TD IDF, se hace para tratar con el texto de manera numérica. |
|  | MonitoreoWhich metrics and KPIs are used to track the ML solution’s impact once deployed, both for end-users and for the business? How often should they be reviewed? |  |  |  |

|  | 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/) |
| --- | --- | --- |