

# Reto Análisis del contexto y la normatividad

Renata Vargas

September 6, 2023

La base de datos que seleccionaré es la base de datos de cáncer de mama de Wisconsin (Breast Cancer Wisconsin, BCW). Comúnmente, esta base de datos se utiliza para entrenar modelos de aprendizaje que como objetivo tienen distinguir entre tumores benignos y malignos en diversos escaneos que realizaron

El conjunto de datos BCW contiene como variables un set de mediciones que sacaron de imágenes digitalizadas de biopsias de tejido mamario. Algunas de ellas son el tamaño y la forma de los núcleos celulares, la textura del tejido y otros atributos.

## 1 Normatividad

**Investiga cual es la normativa asociada al tipo de datos que utilizaste.**

Al entrar a la página del parámetro de la base de datos brindada directamente por Scikit Learn [https://scikit-learn.org/stable/modules/generated/sklearn.datasets.load\\_breast\\_cancer.html](https://scikit-learn.org/stable/modules/generated/sklearn.datasets.load_breast_cancer.html), hay una opción donde te redirige a la página oficial de la Universidad de Wisconsin. Dentro de la página principal de información, está la sección de Licencias que redacta el siguiente mensaje: **This dataset is licensed under a Creative Commons Attribution 4.0 International (CC BY 4.0) license. This allows for the sharing and adaptation of the datasets for any purpose, provided that the appropriate credit is given.**

La Licencia a la que hace referencia es la que dictamina lo siguiente:

*Las licencias públicas de Creative Commons proporcionan un conjunto estándar de términos y condiciones que los creadores y otros titulares de derechos pueden utilizar para compartir obras originales de su autoría y cualquier otro material sujeto a derechos de autor y a otros derechos que se especifican en la licencia pública de más abajo. Las consideraciones siguientes son solo con fines informativos, no son exhaustivas y no forman parte de nuestras licencias.*

Adicional a la Licencia previa, investigué un poco más sobre las posibles regulaciones que podrían aplicar al uso de esta base de datos en específico.

- **Ley de Portabilidad y Responsabilidad del Seguro de Salud (HIPAA):** Esta ley en los Estados Unidos establece estándares para la privacidad y seguridad de la información de salud. Sin embargo, la aplicación de HIPAA gen-

eralmente se aplica a organizaciones de atención médica y entidades relacionadas, en lugar de conjuntos de datos utilizados en investigación o educación. [1]

- **Regulaciones de protección de datos de la Unión Europea (UE):** Si está utilizando datos de pacientes de la UE, es posible que deba cumplir con el Reglamento General de Protección de Datos (GDPR), que establece estándares estrictos para la protección de datos personales, incluidos los datos de salud. [2]
- **Confidencialidad de datos médicos:** Independientemente de las regulaciones específicas, es fundamental garantizar la confidencialidad de los datos médicos en todos los aspectos de su trabajo, incluso al utilizar conjuntos de datos de investigación. [3]

## 2 Fuentes

**Explica porque medio utilizaste los datos y como te estas asegurando de no violar la normativa establecida. Respalda esto con los documentos oficiales en los que te estás basando y cítalos.**

Estoy utilizando la base de datos a través de una fuente registrada y oficial que es Scikit learn. Además de eso, la misma página cita las fuente de la base de datos. Otro factor que cumple con la lista de las normas es que todo el uso de esta base de datos es con fines educativos proporcionados por la institución. De igual manera, la base de datos no tiene ningún tipo de información que pueda dar indicios de a quien le pertenece el registro; es decir, la base de datos no contiene datos sensibles ni privada. En la página de la Universidad de Wisconsin encontramos la siguiente información: ,

*Consideraciones para licenciantes: Nuestras licencias públicas están destinadas a ser utilizadas por aquellos autorizados a dar permiso al público para utilizar material en formas que de otra manera estarían restringidas por los derechos de autor y por otros derechos. Nuestras licencias son irrevocables. Los licenciantes deben leer y entender los términos y las condiciones de la licencia que elijan antes de aplicarla. Los licenciantes deben también contar con todos los derechos necesarios antes de aplicar nuestras licencias, de modo que el público pueda reutilizar el material según lo esperado. Los licenciantes deben identificar con claridad cualquier material que no esté sujeto a la licencia. Esto incluye materiales que se encuentren distribuidos con otra licencia de CC o materiales utilizados en virtud de alguna excepción o limitación al derecho de autor.*

## 3 Modelo

**Una vez analizado el dataset ahora hay que analizar la herramienta, explica cómo tu solución cumple con la normativa establecida en la misma industria y no incurre en ningún sesgo ético.**

Como mencioné previamente, el uso de esta base de datos así como el modelo de predicción o clasificación tiene únicamente fines académicos y de investigación. Gracias al anonimato de los registros de esta base de datos, cumplimos con la no-discriminación ni sesgos que pueden perjudicar el modelo. De igual manera, los modelos que realizamos no siguen ningún tipo de plagio. Cumpliendo con las normas impuestas por la institución, la distribuidora de la base y reglas éticas

básicas de la tecnología.

## 4 Faltas Éticas

**Analiza bajo qué escenarios se podría incurrir en una falta ética si tu herramienta es mal usada, contempla casos como malicia o negligencia.**

El peor caso que se podría desarrollar en caso de mal uso de la herramienta, en este caso específico, es un diagnóstico erróneo. El modelo tiene un porcentaje de error propenso a dar un mal diagnóstico y como consecuencia poner en riesgo la vida de una paciente. Otro caso sería en donde tuviera información suficiente para poder sesgar los datos de tal manera que tengan una tendencia discriminatoria hacia mujeres que cumplan con ciertas características. En caso de que la base de datos SI tuviera información sensible (información privada de las pacientes que puedan exponer sus identidades), y yo decidiera usar la base de datos, también sería una falta ética por el hecho de que conozco que esta es información sensible y probablemente no consensuada a ser utilizada.

Es por eso que es muy importante que las personas que tengamos acceso a este tipo de herramientas e información, sepamos como utilizarla de manera correcta y no corrosiva hacia las demás personas.

## References

- [1] Health insurance portability and accountability act (HIPAA). 1996.
- [2] General data protection regulation (GDPR). 2018.
- [3] Confidentiality of medical records. various.