

En este caso, para el análisis de contexto y normatividad, elegí el primer portafolio del módulo 2, el cual fue sobre aprendizaje de máquina, en mi caso mediante el uso del dataset de diabetes, accediendo a este por medio de la librería de scikit-learn de Python.

Con respecto a la normativa asociada al tipo de datos, en este caso el acceso al dataset se realizó por medio de scikit-learn, y este dataset se encuentra ahí disponible de manera gratuita y sin muchas restricciones, ya que es un conjunto de datos abierto y es utilizado exclusivamente para fines educativos y de investigación. Este tipo de datos generalmente no está obligado a requerir o seguir regulaciones estrictas como el Reglamento General de Protección de Datos (GDPR), o también en este caso dado el tema del dataset, la Ley de Portabilidad y Responsabilidad de Seguros de Salud (HIPAA), ya que no contiene información personal identificable.

Sin embargo, es recomendable que siempre se trate de evitar usar datos que pudieran comprometer la seguridad o privacidad, o que puedan ser utilizados fuera de su propósito y contextos originales o que sean usados para fines maliciosos. Si fuera un dataset con información sensible o que no fuera de uso abierto, algunas normativas relevantes serían el GDPR (Reglamento General de Protección de Datos) ya que se debe de trabajar en base a este en caso de que se trabaje con datos personales en el dataset, y de igual forma está el HIPAA (Health Insurance Portability and Accountability Act) en el caso de datos de salud en EE.UU.

En este caso, no es obligatorio tomar estas normativas en cuenta, ya que el dataset de diabetes se utilizó exclusivamente para propósitos educativos, el cual es uno de sus fines originales, y además el acceso al dataset se hizo mediante scikit-learn, la cual además de ser de licencia abierta, es una biblioteca ampliamente utilizada en machine learning, ya que se trata de datos abiertos a todo público y que no contienen información sensible, lo cual reduce la carga normativa con respecto a la protección de

datos personales. Dado que estos datos no tienen elementos identificables, no violan la normativa de protección de datos.

Para respaldar que no se viola ninguna normativa de datos al usar el dataset de diabetes de scikit-learn, se puede citar el hecho de que en este no se incluyen datos personales identificables, lo que lo exime del cumplimiento del Reglamento General de Protección de Datos (GDPR), dado que este solo se aplica cuando se trabaja con datos que permitan identificar a una persona. Según el artículo 2 del GDPR, “los principios de protección de datos no deben aplicarse a la información anónima, [...] el presente reglamento no afecta al tratamiento de dicha información anónima, inclusive con fines estadísticos o de investigación” [1]. Dado que el dataset de diabetes es precisamente anónimo y es utilizado para fines educativos y de investigación, no entra en conflicto con esta normativa.

Para asegurar que la herramienta cumple con las normativas de la industria, se deben seguir prácticas responsables, como tratar de que los modelos con los que se trabajen no generen sesgos que puedan ser injustos, como por ejemplo basarse de manera excesiva en características como la edad, género, nacionalidad o raza, ya que esto podría derivar en otros problemas como exclusión o discriminación.

Con respecto a consideraciones éticas, se podrían implementar auditorías en busca de sesgos, para que de esta manera se supervisen los modelos de machine learning para detectar posibles sesgos y tratar de prevenirlos, evitarlos o eliminarlos en caso de que ya existan. Por ejemplo, en el caso del dataset de diabetes, tratar de que el modelo no sobrevalore características discriminatorias para sus predicciones. De igual manera, es recomendable comentar el código y mantener una documentación clara de los datos utilizados y del proceso de entrenamiento del modelo para poder reproducir su uso en caso de ser necesario.

La cuestión ética es muy importante, ya que, si se incumple esta parte, puede darse a que haya algunos escenarios en los que el mal uso de datos y herramientas de machine learning o similares (ya sea por malicia o negligencia), podría tener algunas consecuencias éticas negativas, como por ejemplo:

- **Discriminación:** Si los datos y el machine learning se usan en un contexto real para tomar decisiones sobre tratamientos médicos y el modelo se entrena con datos sesgados, podría discriminar a ciertos grupos (por ejemplo, por raza, género, edad, etc).
- **Negligencia en la interpretación de resultados:** Un mal uso del modelo por parte de personal que no esté bien capacitado puede llevar a decisiones inadecuadas por ejemplo en el ámbito médico, como tratamientos erróneos, falsos positivos, falsos negativos, etc.
- **Manipulación intencional de los resultados:** Un mal uso de forma maliciosa podría ser por ejemplo la modificación intencional del modelo para que favorezca o desfavorezca ciertos resultados, generando daño a pacientes, violando la ética y posiblemente dando lugar a que haya discriminación.

Es por esto que se debe evitar o al menos limitar lo más que se pueda el mal uso (tanto accidental como intencional) de esta tecnología y los datos sensibles; y se debe de procurar concientizar y capacitar correctamente a los usuarios en la ética del uso y la protección de datos y modelos de machine learning.

[1] Unión Europea. (2016). *Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo de 27 de abril de 2016 relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos (Reglamento general de protección de datos)*. Diario Oficial de la Unión Europea.
<https://eur-lex.europa.eu/legal-content/ES/TXT/?uri=CELEX%3A32016R0679>